Rabindranath Bera
Subir Kumar Sarkar
Om Prakash Singh
Hemanta Saikia   *Editors*

Advances in Communication, Devices and Networking

Proceedings of ICCDN 2018

Springer

Lecture Notes in Electrical Engineering

Volume 537

Rabindranath Bera · Subir Kumar Sarkar ·
Om Prakash Singh · Hemanta Saikia
Editors

Advances in Communication, Devices and Networking

Proceedings of ICCDN 2018

Springer

Editors
Rabindranath Bera
Department of Electronic
and Communication Engineering
Sikkim Manipal Institute of Technology
Rangpo, Sikkim, India

Om Prakash Singh
Department of Electronic
and Communication Engineering
Sikkim Manipal University
Gangtok, Sikkim, India

Subir Kumar Sarkar
Department of Electrical
and Telecommunication Engineering
Jadavpur University
Kolkata, West Bengal, India

Hemanta Saikia
Department of Electronic
and Communication Engineering
Sikkim Manipal University
Gangtok, Sikkim, India

Organizing Committee

Chief Patron

Lt. Gen.(Dr.) Venkatesh, VC, SMU

Patron

Dr. K. S. Sherpa, Registrar, SMU
Dr. A. Sharma, Director, SMIT

General Chair

Dr. Rabindranath Bera, ECE, SMIT

International Advisory Committee

Dr. Hiroshi Iwai, Tokyo Institute of Technology, Yokohama, Japan
Dr. Durgamadhab Misra, New Jersey Institute of Technology (NJIT), NJ, USA
Dr. Valentina Emilia Balas, Romania
Dr. EfeFrancis Orunmwense, CRART, South Africa
Dr. Subhas Chandra Mukhopadhyay, Macquarie University, Australia
Dr. Rajeev Kumar Kanth, University of Turku, Finland
Dr. Rupesh Kumar, Technicolor, Rennes, France
Dr. BabuSena Paul, University of Johannesburg, South Africa
Dr. Sanjeevi Kumar Padmanaban, University of Johannesburg, South Africa

National Advisory Committee

Dr. Sanjay Dahal, SMIT
Dr. Rabindranath Bera, SMIT
Dr. Ajeya Jha, SMIT
Dr. K. S. Sherpa, SMU
Dr. Kalpana Sharma, SMIT
Dr. B. B. Pradhan, SMIT
Dr. H. K. D. Sarma, SMIT
Dr. Sangeeta Jha, SMIT
Dr. Vinod Kumar Sayal, SMIT
Dr. Gobinda Chandra Mishra, SMIT
Dr. D. R. Poddar, Jadavpur University, Kolkata
Dr. J. P. Banerjee, IRPEL, Calcutta University
Dr. Sayan Chatterjee, Jadavpur University
Dr. Kiran Shankar Hazra, Scientist, Institute of Nanoscience and Technology
Dr. Rohit Sinha, IIT Guwahati
Dr. Rajib Kumar Panigrahi, IIT Roorkee
Dr. Ratnajit Bhattacharjee, IIT Guwahati
Dr. Shaik Rafi Ahamed, IIT Guwahati
Mr. Chinmay Kumar Nayak, SC-E, DRDO, Chandipur
Dr. Singam Jayanthu, NIT Rourkela
Dr. Kanak Chandra Sarma, Gauhati University
Dr. B. K. Rai, IIT Guwahati
Dr. Subir Kumar Sarkar, Jadavpur University, Kolkata
Dr. Rowdra Ghatak, NIT Durgapur
Dr. V. Sarala, SC-F, DRDO, Hyderabad
Dr. Monojit Mitra, IIEST, Shibpur
Mr. Arijit Mazumdar, SC-E, SAMEER, Kolkata
Dr. Bansibadan Maji, NIT, Durgapur
Dr. Samarjit Ghosh, Thapar University, Punjab
Dr. Debdatta Kandar, NEHU, Shillong, Meghalaya, India
Dr. Rajeeb Dey, NIT, Silchar
Dr. S. B. Gadgil, OS and SC-H, DRDO
Dr. Roy P. Paily, IIT Guwahati
Dr. Abhilasha Mishra, Maharashtra Institute of Technology
Dr. Sumana Kumari, University Polytechnic, BIT Mesra
Dr. Nagendra Pratap Singh, IIT-BHU
Dr. Sanjay Kumar Ghosh, Bose Institute, Kolkata
Dr. C. K. Sarkar, Jadavpur University
Dr. Sujit Kumar Biswas, Jadavpur University
Dr. P. C. Pradhan, SMIT
Dr. S. R. M. Prasanna, IIT Guwahati
Dr. P. K. Bora, IIT Guwahati

Dr. Kuntal Deka, IIIT Guwahati
Dr. Rahul Shrestha, IIT Mandi
Dr. Jolly Dhar, SC-F, ISRO-SAC
Dr. Ajit Pal, IIT Kharagpur
Dr. Debadatta Pati, NIT Nagaland

Technical Programme Committee

Dr. Rabindranath Bera, SMIT
Dr. Subir Kumar Sarkar, Jadavpur University, Kolkata
Dr. Rupesh Kumar, Rennes, France
Dr. Babu Sena Paul, University of Johannesburg, South Africa
Dr. Rajeev Kumar Kanth, University of Turku
Dr. Ashik Paul, IRPEL, Calcutta University
Dr. Angsuman Sarkar, Kalyani Government Engineering College, Kalyani
Dr. Himadri Sekhar Dutta, Kalyani Government Engineering College, Kalyani
Dr. Prolay Saha, Jadavpur University, Kolkata, India
Dr. Sanatan Chattopadhyay, University of Calcutta
Dr. Kaustavl, Physics Department, Jadavpur University
Dr. Valentina Emilia Balas, Romania
Dr. Nagendra Pratap Singh, Banaras Hindu University
Dr. Rajeeb Dey, National Institute of Technology, Silchar
Dr. Roy P. Paily, IIT Guwahati
Dr. Abhilasha Mishra, Maharashtra Institute of Technology
Dr. Sumana Kumari, University Polytechnic, BIT Mesra
Dr. Debdatta Kandar, NEHU
Dr. P. K. Banerjee, Jadavpur University, Kolkata
Dr. Sukla Bose, Kalyani Government Engineering College, Kalyani
Dr. Saurabh Das, Indian Statistical Institute, Kolkata
Dr. Barnali Dey, SMIT
Dr. L. Joyprakash Singh, NEHU
Mr. S. K. Dutta, NEHU
Dr. Sam Darshi, IIT Ropar
Mr. Pranab Kishore Dutta, NERIST
Dr. Brijesh Kumbhani, IIT Ropar
Mr. Madhusudhan Mishra, NERIST
Dr. Murli Manohar, IIIT Manipur
Dr. Anjan Kundu, IRPEL, Calcutta University
Dr. Dipanjan Bhattacharjee, SMIT
Dr. Md. Ruhul Islam, SMIT
Md. Nasir Ansari, SMIT
Mr. Amit Kumar Singh, SMIT
Dr. Mousumi Gupta, SMIT

Dr. Samarjeet Borah, SMIT
Dr. Utpal Deka, SMIT
Dr. Bibhu Prasad Swain, SMIT
Dr. Somenath Chatterjee, SMIT
Dr. Sourav Dhar, SMIT
Dr. Tanushree Bose Roy, SMIT
Mr. Akash Kumar Bhoi, SMIT
Dr. Amrita Biswas, SMIT
Dr. Swastika Chakraborty, SMIT
Dr. H. Saikia, SMIT
Mr. Debjyoti Ghosh, SMIT
Mr. Om Prakash Singh, SMIT
Ms. Soumyasree Bera, SMIT
Mr. Arun Kumar Singh, SMIT
Mr. Samarendra Nath Sur, SMIT
Mr. Amit Agarwal, SMIT
Dr. Swanirbhar Majumder, NERIST
Dr. Nagesh Ch, IIIT Manipur
Dr. Satyabrata Das, VSSUT, Burla
Dr. Rajib Jana, NIT AP
Dr. J. P. Singh, NIT Patna
Dr. Jitendra Mohan, JIIT Noida
Mrs. Deepika Hazarika, Tezpur University
Dr. Vijay Kumar Nath, Tezpur University
Dr. Aheibam Dinamani Singh, NERIST
Dr. Dipanjan Bhattacharjee, SMIT
Mr. Bikash Sharma, SMIT
Dr. Swarup Sarkar, SMIT

Convener

Mr. Samarendra Nath Sur

Co-conveners

Ms. Soumyasree Bera
Mr. Arun Kumar Singh
Mr. Amit Agarwal

Preface

The second International Conference in Communication, Devices and Networking (ICCDN 2018) was organized by the Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Majitar, Sikkim, during June 2–3, 2018.

Smart cities are the wave of the future, providing one of the most fascinating and innovative applications for the Internet of things (IoT) and offering untold benefits for governments and citizens around service provision, quality, and safety of life. This motivates the organizer to have "SMART CITY" as the theme of the conference.

The aim of the conference was to provide a platform for researchers, engineers, academicians, and industry professionals to present their recent research works and to explore future trends in various areas of engineering. The conference also brought together both novice and experienced scientists and developers, to explore newer scopes, collect new ideas, establish new cooperation between research groups, and exchange ideas, information, techniques, and applications in the fields of electronics, communication, devices, and networking.

The ICCDN 2018 committees rigorously invited submissions for many months from researchers, scientists, engineers, students, and practitioners related to the relevant themes and tracks of the conference. The call for papers of the conference was divided into six tracks as mentioned—Track 1: Electronics and Nano-technology; Track 2: Energy and Power; Track 3: Microwave; Track 4: Wireless Communication and Digital Signal Processing; Track 5: Control and Instrumentation; Track 6: Data Communication and Networking.

All the submissions underwent a strenuous peer-review process which is comprised of expert reviewers. The papers were then reviewed based on their contributions, technical content, originality, and clarity. The entire process, which includes the submission, review, and acceptance processes, was done electronically. A total of 156 papers have been received, out of which 103 papers have been accepted. All these efforts undertaken by the organizing committees led to a high-quality technical conference program, which featured high-impact presentations from guest speakers

and from paper presenters. All attendees appreciate and expand their expertise in the latest developments in their relevant fields.

We would like to thank the patrons, general chairs, the members of the technical program committees, advisory committees, and reviewers for their excellent and tireless work. We also want to thank Springer for the support and authors and sponsor for sponsoring this very successful conference.

Contents

About the Editors

Rabindranath Bera is Professor and Head of the Department of Electronics and Communication Engineering (ECE) at SMIT and has been driving the ECE Department for 13 years. He is also engaged in R&D activities including sponsored and in-house projects. He completed his B.Tech., M.Tech., and Ph.D. at the Institute of Radio Physics and Electronics, University of Calcutta, Kolkata. His areas of interest are remote sensing (active and passive), communication, 4G mobile communication, integrated sensing and communication, adaptive signal processing and cognitive radio and radar systems, radio astronomy, and process control instrumentation. He won the URSI "International Young Scientists Award" in 2005 and attended the IEEE 5G Summit in Kolkata (March 3–4, 2017). He has 3 books, 85 journal papers, and 80 conference papers published to his credit.

Subir Kumar Sarkar completed his B.Tech., M.Tech., and Ph.D. (Tech) at the Institute of Radio Physics and Electronics, University of Calcutta, and his postdoc at Virginia Commonwealth University (VCU), USA. He has worked 10 years in the industry as Executive Engineer for the Oil and Natural Gas Corporation (ONGC), and 25 years at universities (8 years at BESU and 17 years at JU) in various capacities. He was Head of the Department of Electronics and Telecommunication Engineering, Jadavpur University, from 2011 to 2013, and has coordinated IC Design and Fabrication Centre, Jadavpur University, from 2016 to present. He has authored five engineering textbooks published by CRC Press, USA; Artech House, USA; Pan Stanford, USA; and S. Chand & Company Pvt. Ltd., India. He has been engaged in 15 R&D projects sponsored by various funding agencies of Government of India (completed/ongoing) and has published more than 600 technical research papers in archived international/national journals and peer-reviewed conference proceedings. His research areas include nanodevices and low-power VLSI circuits, computer networks, digital watermarking, and RFID. He is Senior Member of

IEEE, IEEE EDS Distinguished Lecturer, Life Fellow of IEI (India) and IETE, Life Member of ISTE, and Life Member of Indian Association for the Cultivation of Science (IACS). He has successfully organized two IEEE-sponsored international conferences as Convener (2004) and as General Chair (2012).

Om Prakash Singh is Associate Professor in the Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, and Coordinator of the DSP/VLSI Laboratory. He received his B.Tech. in electronics from SRTMU, Nanded, India, and his M.Tech. in communication and signal processing from IIT Bombay. He has more than 10 years of teaching experience in adaptive signal processing, VLSI, advanced communication, etc., and is currently pursuing research in speech processing and image processing.

Hemanta Saikia is Associate Professor in the Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, a constituent college of Sikkim Manipal University. He received his B.E. in instrumentation engineering from Jorhat Engineering College under Dibrugarh University, India; M.Tech. in electronics design and technology from Tezpur University, India; and Ph.D. in technology from Gauhati University, India. His research interests are pattern recognition, instrumentation, and digital systems. He has been in the field of education and research for more than 14 years. He is Member of IEEE, Corporate Member of the Institution of Engineers (India), and Life Member of Systems Society of India. He has published eight research papers in various international journals of repute.

A Simple VIS–NIR LED-Based Model for Estimation of Plant Chlorophyll

Wesheü Mero and Subra Mukherjee

Abstract Estimation of chlorophyll content often involves destructive analysis and lengthy procedures. Several sensors systems have been developed that measure and monitor plant health status; however, most of these methods are destructive or they involve posh devices. We intended for a cost-effective and noninvasive method for the estimation of chlorophyll content of tulsi (*Ocimum tenuiflorum*) leaves. The proposed sensor system works for diffused transmittance method using VIS–NIR sources kept in the near target area of different conditions of the leaves. Leaves can be sensed under different sensing conditions and their outcome is analyzed. The transmittance values obtained are measured using a detector. Based on the transmittance values, Simple Ratio Vegetation Index (SRVI) is computed and the result is correlated to the chlorophyll content of the leaves. Also, the proposed sensor system has been compared with three different standard laboratory methods and the results obtained are satisfactory. The performance of the proposed sensor system is statistically analyzed using indicators such as RMSE and R^2. The RMSE obtained is 0.004 and R^2 is found to be greater than 0.9 in almost all the cases. The proposed methodology aims to produce a nondestructive, reliable, convenient, and cost-effective way of estimating chlorophyll content of leaves and thereby aid the agri-tech industry. Further, this work has the potential of providing information for use in remote sensing and energy balance studies.

Keywords Chlorophyll · Transmittance · Spectroscopy · Optical sensor · Plant nutrients · Vegetation index

W. Mero (✉) · S. Mukherjee
Department of Electronics & Communication Engineering, Assam Don Bosco University, Guwahati, India
e-mail: ashelemerowk@gmail.com

S. Mukherjee
e-mail: subra.mukherjee@dbuniversity.ac.in

1 Introduction

Generally, chlorophyll is a green pigment comprising of tetrapyrrole ring and a central magnesium ion. It is found mostly in plants and algae. A long hydrophobic phytol chain is present in the structure of chlorophyll [1]. Chlorophyll is the most important pigment and is necessary for the conversion of light energy to the stored chemical energy. Chlorophyll absorbs light generally in the red (650–700 nm) and the blue-violet (400–500 nm) regions of the visible spectrum. Green region (~550 nm) is reflected giving chlorophyll its specific color. Chlorophyll aids the body in a unique and distinctive way by detoxifying harmful toxins and fights infection from the body [2]. Two important chlorophyll present are chlorophyll a and b, where chlorophyll a holds a green-blue color, and chlorophyll b holds a green-yellow color [3]. In tumor therapy, chlorophyll or chlorophyll results can be employed such as a photodynamic agent [4]. It can also be further studied, amended, and produced in disciplines like chemistry and physics for different uses comprising of electronic, photophysics, optoelectronic, and electrochemistry [5]. By the color of leaf, one can specify the proportion and amount of chlorophyll in leaves, which are related to the health status of the particular plant based on different conditions. Hence, chlorophyll content can be used as a key to detecting nutrient status. Various kinds of color chart and chlorophyll meters have been developed to estimate leaf color or chlorophyll content which are helpful in agro-industry but may lead to certain limitations [6]. Spectral signature characteristics are closely related to leaf chlorophyll content and depend on the interaction that occurs when light emits plant tissue, where some of the signals will be absorbed, transmitted, and reflected from the leaf surface [7]. Readings from the chlorophyll meter are basically a measure of the greenness of leaf and in general shows the linear relationship with extractable chlorophyll concentrations for a wide variety of crops. However, environmental conditions may affect the relationship between plant species [8]. The usability of Soil Plant Analysis Development (SPAD) meter transforms the indices to interpretable data. Before the readings are measured for a particular species, calibration is required as the sensor changes the signal into indices that could cause variation [9]. However, spectral signatures are useful in sensing plant stress only if one single factor is involved. It may be challenging when different stress factors affect a plant at the same time. This is due to the fact that some stressors may affect the physiology of the plant species in a similar way, as in the case of nutrient deficiency and varies in pigment, moisture, and canopy of plant species [10]. In paper [11–14], estimation of chlorophyll content was mostly performed by chlorophyll meters and digital cameras using reflectance measurement, in different types of leaves. Extraction and different types of vegetative indices were also obtained with the measurement data. Different techniques were employed but show the nonlinear relationship with the chlorophyll content.

In this paper, optical properties for detection, monitoring, and estimation of chlorophyll content are considered. In a nutshell, the transmittance measurement has been currently employed in designing of the sensor system. The system uses transmittance measurement in such a way that the light transmitted by the sample, i.e., based on

different conditions of leaves, is collected by the detector and gives the correlate output voltage. Several VIS–IR sources were used. But further, it was preceded with IR and red sources, because it gives a better correlation with the chlorophyll content. So, we intended to produce a simple cost-effective sensor system for detection and monitoring of factor(s) related to plant health status.

2 Materials and Methods

2.1 Single-Leaf Method

Leaves were collected from different conditions of tulsi (*Ocimum tenuiflorum*) leaves. Transmittance measurement was performed. Several experiments were carried out and the transmittance readings were obtained for different conditions of tulsi (*Ocimum tenuiflorum*) leaves. Analysis of the experiments was performed in a dark room at room temperature. The two sources red (650 nm) and IR (930 nm) were used during the experiment. So, for transmittance response, it was converted into Simple Ratio Vegetative Index (SRVI) and statistical analysis was carried out.

2.2 Spectrophotometer Method

Leaves of tulsi (*Ocimum tenuiflorum*) plant under different conditions, i.e., healthy, early stress, slightly stress, and fully stress, were collected. Separate sample solutions were prepared and the absorbance and transmittance values were obtained using a spectrophotometer. The values obtained were converted into Chlorophyll Concentration Index (CCI) and chlorophyll concentration for the validation of further analysis.

2.3 Design of the System

Generally, the design of the sensor system shows the use of devices like LEDs for measuring chlorophyll content. The properties of diffuse transmittance were used in order to check the transmitted light from the samples. The system consists of the sensing part which has two sources—IR (930 nm) and red (650 nm)—and detector along with the sample to be examined. It has a control unit for estimating and monitoring the chlorophyll of leaves based on different conditions, and finally display unit, where the result will be displayed based on the decision. Figure 1 shows the block diagram and Fig. 2 shows the flowchart of the proposed methodology.
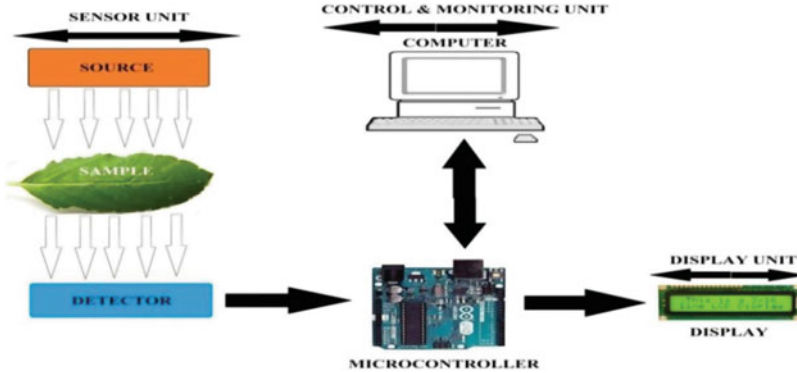
Fig. 1 Block diagram of the proposed sensor system
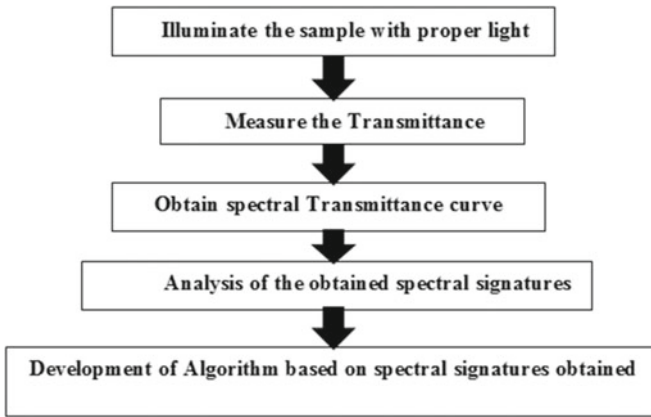
Fig. 2 Flowchart for the proposed methodology

3 Results and Discussion

Greens are highly useful for the maintenance of health and prevention of diseases and also considered important sources of protective food [2]. To study this, we considered an important parameter known as **Simple Ratio Vegetation Index (SRVI)** which measures the greenness of leaf.

Figure 3 shows the comparison done for SRVI obtained for different conditions of tulsi leaves. Several experiments were carried out to analyze the relation of red and infrared wavelengths to the chlorophyll content. Different conditions, i.e., healthy, early stress, slightly stress, and fully stress of tulsi (*Ocimum tenuiflorum*) leaves, were used for estimating the chlorophyll content and it was observed that SRVI value varies for different conditions of tulsi (*Ocimum tenuiflorum*) leaves.

Simple Ratio Vegetation Index (SRVI) is a simple indicator employed to check the greenness of a target area. It is given by
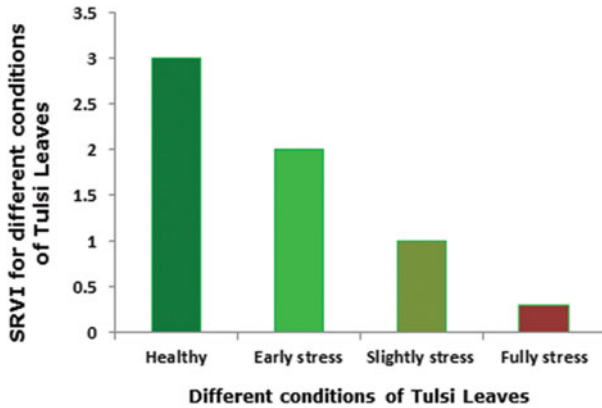
Fig. 3 SRVI for different conditions of tulsi (*Ocimum tenuiflorum*) leaves

$$SRVI = \frac{\text{Transmitted or reflected light from the leaf in IR region}}{\text{Transmitted or reflected light from the leaf in Red region}}$$

3.1 Comparison of SRVI for Different Methods

Comparing reported data collected from tulsi (*Ocimum tenuiflorum*) plant from different conditions, i.e., healthy, early stress, slightly stress, and fully stress, different methods were employed, i.e., SRVI of single-leaf method, SRVI from the Spectrophotometric method (SPH), chlorophyll concentration (Theoretical), and Chlorophyll Concentration Index (CCI) based on spectrophotometric method. These four methods showed the strong relationship and were correlated with different conditions of the plant (Fig. 4).

In CCI of tulsi (*Ocimum tenuiflorum*) leaves, the values correlated linearly to SRVI with R^2 of 0.90. The chlorophyll concentration affects the plants of various species including the photosynthetic rate. So from several analyses obtained, the concentration relates well with SRVI.

The SRVI (SPH) was also linearly correlated with SRVI and has an R^2 of 0.96. These further confirm the need of developing a definite relationship from the obtained results. Higher R^2 obtained furthermore check the strength of the sensor system for use in other plant species (Fig. 5).

In Fig. 6, the chlorophyll concentration (SPH) was linearly correlated with SRVI and has an R^2 of 0.94. Table 1 also shows the performance of the proposed sensor system when compared with different methods, i.e., Spectrophotometer, Theoretical chlorophyll concentration, and Chlorophyll Concentration Index (CCI). The error performance was best observed in spectrophotometer for the different conditions of
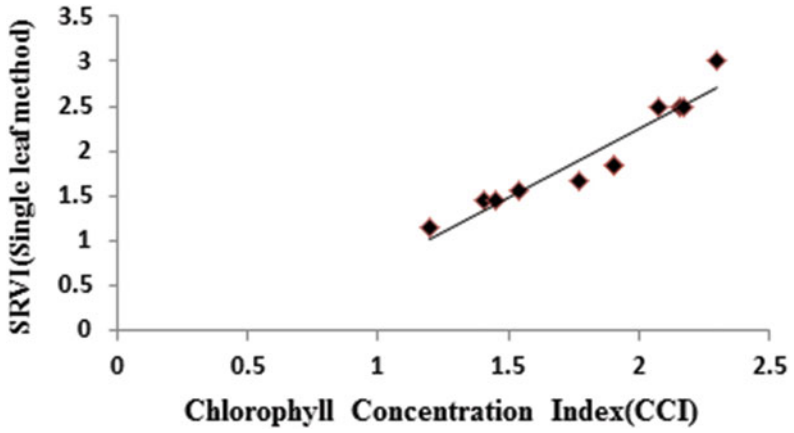
Fig. 4 Relationship between SRVI of single-leaf method with Chlorophyll Concentration Index (CCI)
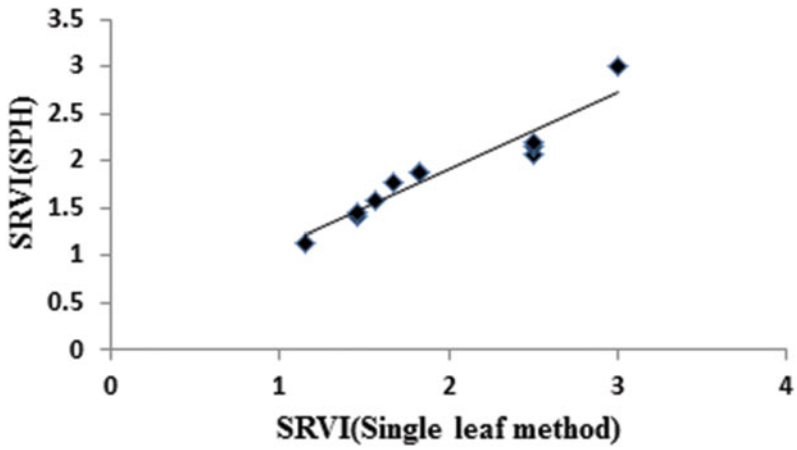
Fig. 5 Relationship between SRVI of single-leaf method with SRVI (SPH)

tulsi (*Ocimum tenuiflorum*) plant obtaining R^2 value of 0.997 and RMSE value of 0.004.
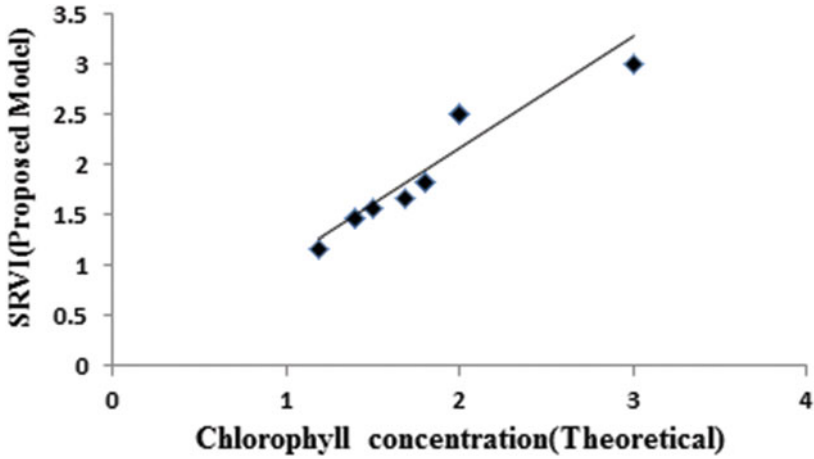
Fig. 6 Relationship between SRVI of single-leaf method with chlorophyll concentration (theoretical)

Table 1 Performance analysis of different methods

Conditions of leaves	SRVI of proposed sensor system	SRVI of spectrophotometer	SSE	SST	R^2	RMSE
Healthy	2.99	3	0.0122	4.0074	0.99	0.045
Early stress	1.4	l.5	0.0001	0.008	0.99	0.004
Slightly stress	0.877	0.9	0.012	4.0074	0.99	0.0447
Fully stress	0.2	0.1	0.012	4.0074	0.99	0.0447
	SRVI of theoretical chlorophyll concentration					
Healthy	3.89	4	0.05	1.479	0.94	0.0912
Early stress	12	1.3	0.0025	0.003	0.96	0.02
Slightly stress	0.78	0.37	0.002	0.008	0.97	0.0183
Fully stress	0.288	0.3	0.002	0.008	0.97	0.0183
	SRVI of chlorophyll concentration index (CCI)					
Healthy	1.887	2	0.025	1.479	0.98	0.0645
Early stress	1.25	1.267	0.001	0.005	0.98	0.0129
Slightly stress	0.49	0.5	0.018	0.007	0.94	0.0547
Fully stress	0.373	0.4	0.018	0.007	0.94	0.0547

4 Conclusion

A simple method based on VIS–NIR LED for estimation of chlorophyll content/nitrogen content in tulsi (*Ocimum tenuiflorum*) leaves has been reported. Estimation of chlorophyll content of leaves has been employed from the red and IR wavelength region. An indicator called SRVI was calculated for the different conditions of tulsi (*Ocimum tenuiflorum*) leaves. The SRVI value varies based on the status of the leaves. The SRVI is greater than 1 for high chlorophyll content indicating healthy leaf and less than 1 for leaves having low chlorophyll content. Different standard laboratory methods were also performed and compared to validate the performance of the proposed sensor system and it showed satisfactory results. RMSE and R^2 were also obtained for healthy, early stress, slightly stress, and fully stress leaves. The success rate from the proposed sensor system was found to be 93%. Further studies will continue with estimation of nitrogen content in different types of leaves. Also, other features of the leaves can be explored and evaluated for their improved representation of nitrogen content in leaves. All different plant species have different concentrations of chlorophyll. So for a specific plant, the content also varies due to different regions. Based on different conditions of the leaves, the content of the chlorophyll may differ.

References

1. Aminot A, Rey F (2000) Standard procedure for the determination of chlorophyll a by spectroscopic methods. International council for the exploration of the sea. ISSN 0903-260
2. Rajalakshmi K, Banu N (2013) Extraction and estimation of chlorophyll from medicinal plants. Int J Sci Res (IJSR) 4(11). ISSN (Online): 2319-7064, Index Copernicus Value (2013): 6.14, Impact Factor (2014): 5.611
3. Aron D (1949) Copper enzymes isolated chloroplasts, polyphenoloxidase in *Beta vulgaris*. Plant Physiol 24:1–15
4. Brandis AS, Salomon Y, Scherz A (2006) Chlorophyll sensitizers in photodynamic therapy. Adv Photosynth Respir 25:461–483
5. Nurhayati N, Suendo V (2011) Isolation of chlorophyll a from spinach leaves and modification of center ion with Zn^{2+}: study on its optical stability. J Mat Sains 16(Nomor 2)
6. Kawashima S, Nakatani M (1997) An algorithm for estimating chlorophyll content in leaves using a video camera. National Institute of Agro-Environmental Sciences, Tsukuba, and Kyushu National Agriculture Experiment Station, Miyakonojo, Annals of Botany Company, 9 Sept 1997
7. Liew OW, Jenny Chong PC, Li B, Asundi AK.(2008) Signature optical cues: emerging technologies for monitoring plant health. Sensors
8. Wood CW, Reeves DW, Himelrick DG (1993) Relationships between chlorophyll meter readings and leaf chlorophyll concentration, N status, and crop yield: a review. Proc Agron Soc N Z 23:1–9
9. Marquard, RD, Tipton JL (1987) Relationship between extractable chlorophyll and an in situ method to estimate leaf greenness. HortScience 22:1327–1327
10. Li X, Zhao X, Fang Y, Jiang X, Duong T, Fan C, Huang C-C, Kain SR (1998) Generation of destabilized green fluorescent protein as a transcription reporter. J Biol Chem Am Soc Biochem Mol Biol Inc 273(52):34970–34975

11. Mark S, Gitelson AA, Rundquist D (2008) Non-destructive estimation of leaf chlorophyll content in grapes. Pap Nat Resour 282
12. Liang Y, Urano D, Liao K-L, Hedrick TL, Gao Y, Jones AM (2017) A non-destructive method to estimate the chlorophyll content of Arabidopsis seedlings
13. Pal K, Singh RD, Prasad R (2012) Non-destructive estimation of chlorophyll and nitrogen content in leaf of Rosa damascena
14. Lichtenthaler HK, Gitelson A, Lang M (1996) Non-destructive determination of chlorophyll content of leaves of a green and aurea mutant of tobacoo by reflectance measurements

A Novel High-Performance Leakage-Tolerant Keeper Domino Circuit for Wide Fan-In Gates

Catherine Langpoklakpam, Sapna R. Ghimiray and Pranab Kishore Dutta

Abstract   A circuit with less power consumption, the minimum amount of leakage, and least possible delay are the primary aim of the circuit designer. In this paper, a circuit exhibiting similar qualities is proposed. The proposed circuit works on the principle to diminish contention between the pull-down network (PDN) and the keeper, according to the leakage flowing across PDN; moreover, the diode-footed configuration aids in reducing circuit leakage current. 180 nm CMOS technology is used for simulation with 1.8 V supply voltage at 500 M clock frequency. The proposed circuit has reduced power consumption by 29% with improved energy efficiency compared to the standard footless domino. The result is verified with 180 nm CMOS technology TCAD EDA tools.

Keywords   Dynamic circuit · Domino logic · PDN · Energy efficiency · Leakage current

1   Introduction

In the modern digital circuits system, dynamic circuits play a vital role. The dynamic circuit properties such as minimum power consumption, fewer transistor numbers, and high performance are the major benefits of dynamic over static CMOS [1, 2]. Dynamic logic is usually used in high-geared wide fan-in circuit, namely, tag comparator, L1 cache, MUX–DEMUX, microprocessors, etc. [3]. With the escalating number of wireless and portable devices, a prompt improvement of VLSI circuit becomes a mandatory subject. In order to get this improvement, high-speed micro-

C. Langpoklakpam (✉) · S. R. Ghimiray · P. K. Dutta
North Eastern Regional Institute of Science and Technology, Itanagar, Arunachal Pradesh, India
e-mail: langpoklakpam.catherin@gmail.com

S. R. Ghimiray
e-mail: sapna2203.sanu@gmail.com

P. K. Dutta
e-mail: pkdutta.07@gmail.com

processors with a reasonable power consumption are required. But to obtain these requirements, the transistors dimension along with supply voltage must be reduced. To mitigate the loss of speed owing to reduction in supply voltage, the transistor threshold voltage should be reduced. Reduction in threshold voltage increases leakage current, and the impunity toward noise is less than that of the static circuit which is the intrinsic property of dynamic circuit [4]. The dynamic circuit operates depending on the input clock pulse which implies that the output is effective only for a certain duration of clock pulse [1]. There are two phases, a pre-charging phase, and an evaluation phase. *Pre-charge phase* (clock = 0), during this phase dynamic node is charged to VDD which implies that the node will remain high regardless of the PDN input data. *Evaluation phase* (clock = 1), during this phase, dynamic node voltage discharges or remained at its previous state depending on input data of PDN. The main concerns of the circuit occur in evaluation phase especially when entire input data of PDN are low [1]. Ideally, during this phase, dynamic node voltage should remain high but reduced noise margin and increased leakage [4] result in an incorrect evaluation because of the deterioration of dynamic node voltage level. The leakage current increases with the increased fan-in, and to improve the performance of scaled-down transistor network circuit, the threshold voltage of transistor should be reduced. But this reduction leads to the increment of sub-threshold leakage which tends to abate circuit robustness toward the noise. Therefore, performance and robustness adjustment become a major agitation of domino logic.

Several domino design techniques are proposed to improve the performance circuit impunity toward noise. Changing the keeper circuitry of keeper transistor and modifying circuit design of footer transistor configuration are some of the design techniques [5]. Standard footless domino circuit [6] helps in combating undesirable discharging but utilization of strong keeper creates contention problem between PDN and keeper results in performance degradation. Domino circuit based on leakage current across PDN [7] reduces the contention between PDN and keeper at the expenses of an extra PMOS and a joint current mirror in every dynamic node. A robust domino circuit which compares the leakage and current across PDN transistors to reduce contention [4] was proposed to improve the performance. The leakage has been reduced but the performance of the circuit has not improved effectively and also the required area is increased. Stacking effect of domino logic with clock and input-dependent transistors (DOIND) [8] were utilized for the reduction of leakage and delay to provide better domino circuit. But this circuit has to depend on DOIND NMOS transistor which operates depending on input combinations of PDN. To overcome all of this drawback, our circuit is proposed which will have better power consumption with improved efficiency.

Section 1 of this paper gives a brief introduction about domino logic followed by basic of power consumption in Sect. 2. The proposed circuit is explained in Sect. 3. Section 4 explains the results of the proposed circuit and finally, Sect. 5 concludes the paper.

2 Power Consumption

The crucial objective of VLSI circuit deviser is to design high-density energy-efficient circuits, particularly for handy devices. The more the power consumption, the more will be casing and cooling costs which questions the circuit reliability. The roots of dynamic circuit power consumption are as follows [9]:

(1) Switching power consumption,
(2) Short-circuit power consumption, and
(3) Static power consumption.

Therefore, the overall average power consumption of the dynamic circuit is [4]

$$P_{\text{AVG}} = P_{\text{Switching}} + P_{\text{Short-circuit}} + P_{\text{Leakage}} \tag{1}$$

where $P_{\text{Switching}}$, $P_{\text{Short-circuit}}$ and P_{Leakage} is power consumption due to switching, short-circuit current and leakage current, respectively. Leakage power escalates with decreasing in technology dimension and also the robustness and impunity toward leakage shrink with an increase in temperature.

The power consumed during charging along with discharging of load capacitance is given by [5]

$$P_{\text{Switching}} = \alpha_{0-1} C_{\text{L}} \text{VDD} \cdot V_{\text{Swing}} f \tag{2}$$

where α_{0-1} is the switching action, which gives the likelihood of an output (0 to 1) alteration in each clock cycle, C_{L} is the load capacitance, f is the clock rate, and V_{Swing} is the utmost output voltage fluctuation of the logic circuit.

The power consumed during transition, i.e., when both PMOSs and NMOSs are active at same time, is given below [10]:

$$P_{\text{Short-circuit}} = \text{VDD}\, I_{\text{SC}} \tag{3}$$

$$I_{\text{SC}} = \frac{\left\{ \mu_o C_{\text{OX}} \left(\frac{W}{L}\right)_{\text{NMOS/PMOS}} (\text{VDD} - 2V_{\text{TH}}) 3\tau N f_{\text{CLK}} \right\}}{\text{VDD}} \tag{4}$$

where μ_o is the unbiased mobility of transistor, C_{OX} is the oxide capacitance of transistor, $\left(\frac{W}{L}\right)_{\text{NMOS/PMOS}}$ is the aspect ratio of NMOS or PMOS, VDD is the supply voltage, V_{TH} is the threshold voltage of transistor, τ is the rise or fall time, N is the total numbers of transistor switching activities, and f_{CLK} is the clock frequency.

The power consumed by leakage current, i.e., the current that flows through the circuit when all transistors are in off state or inactive state, is [11]

$$P_{\text{Leakage}} = \text{VDD}\, I_{\text{Off}} \tag{5}$$

$$I_{\text{off}} = I_{\text{GT}} + I_{\text{GIDL}} + I_{\text{REV}} + I_{\text{SUB}} \tag{6}$$

where I_{GT} is the gate tunneling leakage current, I_{GIDL} is the gate-induced drain leakage, I_{REV} is the reverse bias junction leakage current, and I_{SUB} is the sub-threshold leakage current.

3 Leakage-Controlled Keeper

The proposed circuit is demonstrated in Fig. 1. It has a diode-footed configuration to reduce leakage. To minimize contention between PMOS keeper and pull-down network (PDN), the operation of PMOS keeper depends on leakage across PDN. To yield stacking effect, an NMOS transistor (M_1) is imported in-between dynamic output node (Y_0) and PDN. The circuit works in two phases, a pre-charge phase and an evaluation phase.

In *pre-charge phase* (clock = 0), the pre-charging transistor (M_{pre}) is active and M_1 is inactive. Hence, the dynamic output node (Y_0) is charged to a logic high value
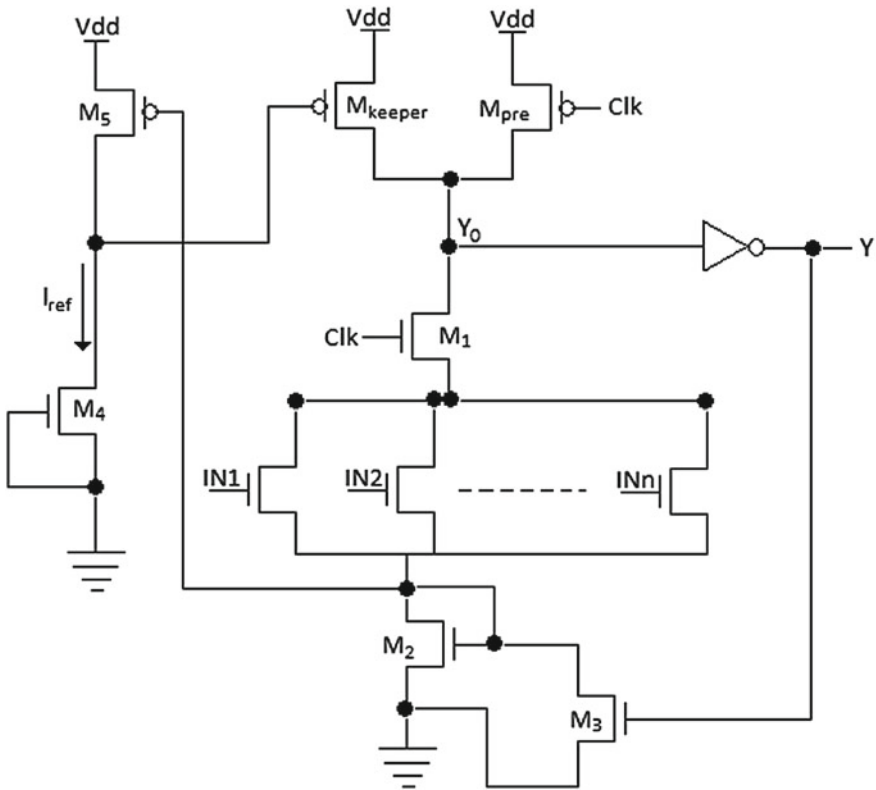
Fig. 1 Leakage controlled keeper

and because of inverter in-between output node (Y) and Y_0, the output (Y) is low, regardless of input data combinations of PDN in Fig. 1.

In *evaluation phase* (clock = 1), the pre-charging transistor (M_{pre}) is inactive and M_1 is active. The logic state of Y_0 in this phase relies on PDN input data. The keeper (M_{keeper}) is turned ON according to leakage magnitude across PDN and reference current (I_{ref}) across M_5 and M_4. The size of M_4 is directly proportionate to the fan-in of the circuit, i.e., for 8 fan-in circuit, size of M_4 is 8 times the size of PDN transistor. Since M_4 is inactive, the only current flowing in this branch is leakage. Hence, I_{ref} is the exhaustive leakage which flows when all transistors of PDN are inactive. So, leakage across dynamic circuit will be lower than I_{ref}. If I_{ref} is greater than PDN current, then M_{keeper} is active or else it is turned OFF.

In the pre-charge phase, M_1 is inactive so the connection between Y_0 and PDN is removed. Leakage current is the only current across PDN in this phase. Therefore, in pre-charge phase, the keeper is in an active mode which further helps dynamic node (Y_0) to remain in the logic high state. M_3 is used to lower the node voltage between PDN and the diode-connected transistor (M_2) which helps in the reduction of delay caused by the stacking effect of M_2.

Decreasing keeper dimension and increasing pre-charging transistor dimension for the proper functioning of the network escalates speed and power consumption. Therefore, adjustment between power consumption and delay is required. So proper sizing of both pre-charging transistor (M_{pre}) and keeper transistor (M_{keeper}) is required to obtain acceptable power consumption with a reasonable delay.

4   Simulation and Result

The simulation of the proposed circuit has done using 180 nm technology. The supply voltage is kept at 1.8 v and 500 MHz for the clock. All simulations have been carried out in the same environment. The transient analysis of the 8, 16, 32, and 64 bits of proposed OR gate circuit is demonstrated in Fig. 2. During the pre-charging phase, irrespective of the input logic combination, e.g., $IN_1, IN_2, …, IN_8$ for 8 inputs, Y_0 is pre-charged to a logic high value and output (Y) discharges with some delay caused by the inverter in-between Y_0 and Y. During the evaluation, when all input data to PDN transistors are low, the output (Y) discharges to ground potential, whereas output (Y) is high if one of the inputs is high. Hence, it is observed that the circuit operates as basic dynamic OR gate. As the number of fan-in increases, the delay of charging the output (Y) to a logic high value increases. Therefore, in the transient response of the proposed circuit, 8 inputs has the least delay while 64 inputs have the highest delay among 8, 16, 32, and 64 inputs as shown in Fig. 2. The performance comparisons between proposed design and standard footless along with various keeper-based domino designs for 8, 16, 32, and 64 inputs are given in Table 1 and its graphical representation of 8 inputs are shown in Fig. 3.

From the simulated result in Table 1, the proposed circuit has a lesser delay and power consumption. In the proposed circuit, the dynamic node discharges completely
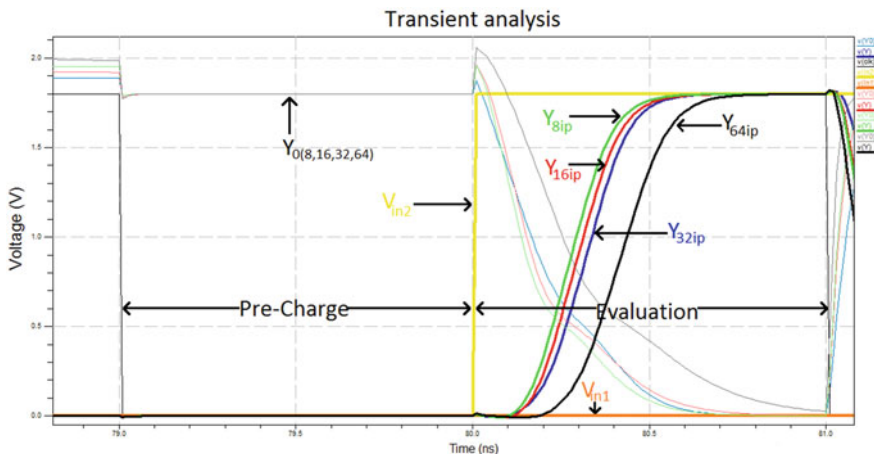
Fig. 2 Transient analysis of leakage-controlled keeper for 8, 16, 32, and 64 inputs

by turning the keeper in inactive. Moreover, stacking effect of the diode-connected transistor (M_2) and the presence of M_1 transistor leads to reduction dynamic, leakage short-circuit power consumption of the circuit. This is evident in the result which shows 20–29% when compared with that of Std-footless. The presence of M_3 and reduced transistor count leads to the reduction of delay by 25–38% when compared with Std-footless, CKCCD, and LCR, and due to the presence of stacking effect of M_2, the delay of the proposed circuit is increased by 1.1–8.9% from DIOND. Power–delay product (PDP) is used to compute energy consumed per switching event. PDP gives the energy efficiency of the circuit. The proposed circuit PDP is reduced by 18–50% when compared to Std-footless, CKCCD, LCR, and DIOND. The leakage current or the standby current in Table 1 is the current across PDN
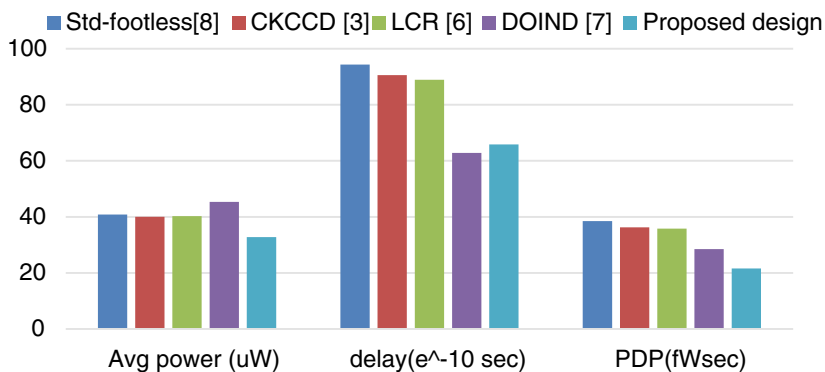
Fig. 3 Graphical comparison of average power consumption, delay, and power–delay product of proposed and existing domino logic circuits for 8 inputs

Table 1 Comparison of the simulated result of various performance parameters for wide fan-in OR gates

I/P	Parameters	Std-footless [8]	CKCCD [3]	LCR [6]	DOIND [7]	Proposed design
8	AvgPower (µW)	40.83	40.06	40.27	45.37	32.79
	Power$_{norm}$	1	0.98	0.99	1.11	0.80
	Delay (ps)	943.1	905.5	889.3	628.4	658.2
	PDP (fW s)	38.5	36.27	35.82	28.51	21.58
	EDP (e^{-24} W s^2)	36.30	32.84	31.85	17.92	14.20
	Leakage current (pA)	454.65	72.75	453.63	0.51	5.55
16	AvgPower (µW)	61.04	59.68	60.97	66.92	44.94
	Power$_{norm}$	1	0.98	0.99	1.09	0.73
	Delay (ps)	1019.4	1022.8	971.4	669.5	678.6
	PDP (fW s)	62.22	61.04	59.23	44.80	30.5
	EDP (e^{-24} W s^2)	63.42	62.43	57.5	29.99	20.69
	Leakage current (pA)	454.64	52.58	454.61	0.27	5.12
32	AvgPower (µW)	70.85	68.05	69.81	76.55	50.81
	Power$_{norm}$	1	0.96	0.98	1.08	0.71
	Delay (ps)	1787.2	1819.6	1760.2	1061.7	1205.4
	PDP (fW s)	126.63	123.84	122.89	81.27	61.24
	EDP (e^{-24} W s^2)	226.3	225.3	216.3	86.29	73.81
	Leakage current (pA)	454.62	37.97	454.58	0.14	3.45
64	AvgPower (µW)	93.04	90.14	92.17	97.72	72.93
	Power$_{norm}$	1	0.96	0.99	1.05	0.78
	Delay (ps)	2079.7	2182.1	1976.3	1207.9	1326.4
	PDP (fW s)	193.49	196.70	182.16	118.04	96.73
	EDP (e^{-24} W s^2)	402.4	429.2	360	142.6	128.3
	Leakage current (pA)	454.60	27.48	454.56	0.12	2.11

transistors when the entire proposed circuit PDN transistors are in an inactive mode. The leakage of the proposed circuit is diminished by 98–99% when compared to standard/conventional Footless Domino Logic and by 90–92% than that of CKCCD from the simulated values in Table 1 and the graphical comparison of leakage current of all domino logic circuits along with the proposed circuit is shown in Fig. 4.
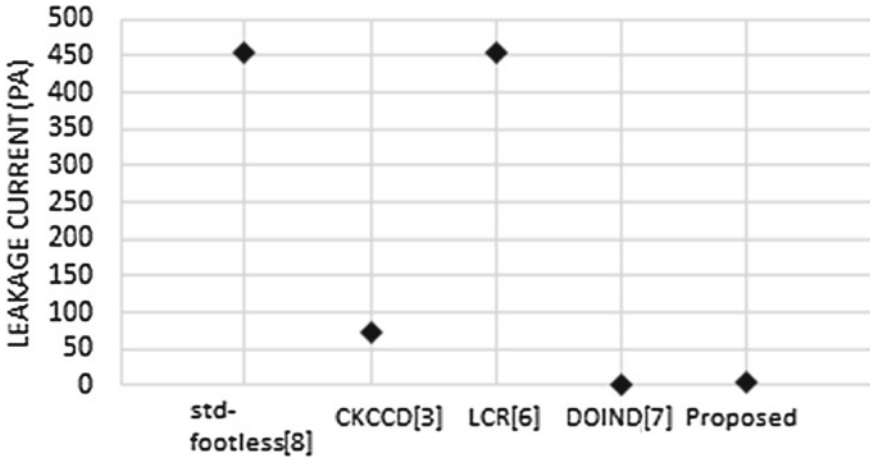
Fig. 4 Graphical comparison of leakage current of the proposed and existing domino logic circuits for 8 inputs

5 Conclusion

A design with low leakage, reduced power consumption with least possible delay called leakage-controlled keeper is proposed. The keeper of the design depends on the magnitude of drain current across PDN to reduce contention for improving the circuit performance. The simulation for various keeper-based domino logic designs and the proposed circuits is done using TCAD EDA 180 nm CMOS technology with 1.8 V supply for 8, 16, 32, and 64 inputs with OR gate as a benchmark circuit. From the simulated result, the circuit proposed has 20–29% decrement in power consumption and drain current of PDN transistor when PDN transistors are inactive of the circuit proposed that shows 98–99% reduction when compared to standard footless. The proposed circuit has the least EDP and PDP among Std-footless, CKCCD, LCR, and DIOND. Therefore, the proposed circuit shows better energy efficiency than the existing domino circuits.

References

1. Rabaey JM, Chandrakasan A, Nicolic B (2003) Digital integrated circuits: a design perspective, 2nd edn. Prentice-Hall, Upper Saddle River, NJ
2. Moradi F, Cao TV, Vatajelu EI, Peiravi A, Mahmoodi H, Wisland DT (2013) Domino logic designs for high-performance and leakage-tolerant applications. Integr VLSI J 46:247–254

3. Wang L, Krishnamurthy R, Soumyanath K, Shanbhag N (2000) An energy-efficient leakage-tolerant dynamic circuit technique. In: Proceedings of international ASIC/SoC conference, pp 221–225
4. Peiravi A, Asyaei M (2012) Robust low leakage controlled keeper by current-comparison domino for wide fan-in gates. Integr VLSI J 45(1):22–32
5. Asyaei M (2015) A new leakage-tolerant domino circuit using voltage-comparison for wide fan-in gates in deep sub-micron technology. Integr VLSI J 51:61–71
6. Oklobdzija VG, Montoye RK (1985) Design-performance trade-offs in CMOS domino logic. In: Proceedings of the IEEE custom integrated circuits conference, pp 334–337
7. Lih Y, Tzartzanis N, Walker WW (2007) A leakage current replica keeper for dynamic circuits. IEEE J Solid-State Circuits 42:48–55
8. Shah AP, Neema V, Daulatabad S (2016) DOIND: a technique for leakage reduction in nanoscale domino logic circuits. J Semicond 37
9. Namdev A, Gupta R, Kurmi US (2015) A survey on different domino logic circuit design for high-performance and leakage tolerant. Int J Adv Res Electr Electron Instrum Eng 4(4)
10. Ram M (2010) Reliability analysis of dynamic logic circuits under transistor aging effects in nanotechnology. MSc thesis, San Francisco, California
11. Fallah F, Pedram M (2005) Standby and active leakage current control and minimization in CMOS VLSI circuits. IEICE Trans Electron (Special Section on Low-Power LSI and Low-Power IP), E88-C(4):509–519

Drain-Doping Engineering and its Influence on Device Output Characteristics and Ambipolar Conduction on a Splitted-Drain TFET Model

Bijoy Goswami, Disha Bhattacharjee, Ayan Bhattacharya and Subir Kumar Sarkar

Abstract The study investigates the ramifications of drain-doping engineering for all characteristics and parameters of a 2D TFET model. The paper suggests that three different structures of the device consist of splitted-drain region with different doping concentrations. Splitted-drain structures exhibit major reduction in ambipolar conduction due to increase in the tunneling width at the channel–drain junction. The structures are named according to the relative position of the drain: Splitted-Drain Single-Gate TFET (SD-SG TFET: drain in upper location), Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET: drain in middle location), and Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET: drain in bottom location). All the fundamental device characteristics and parameters are analyzed for all the three structures, and their merits and drawback are recorded for optimal evaluation. All the simulations are done in Silvaco, Atlas.

1 Introduction

Apparent limitation of MOSFET technology conducts the widespread of TFET devices which have been revolutionizing the nanoscale industry due to its vari-

B. Goswami · D. Bhattacharjee · A. Bhattacharya (✉) · S. K. Sarkar
Department of ETCE, Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: ayanbrkmahs@gmail.com

B. Goswami
e-mail: bijoy.ete@aec.ac.in

D. Bhattacharjee
e-mail: dishabhatt.v@gmail.com

S. K. Sarkar
e-mail: su_sircir@yahoo.co.in

ous desirable interests. The primary interest of TFET is current conduction through modulation of quantum mechanical Band-to-Band Tunneling (BTBT) [1], which has advantages over the thermionic injection across an energy barrier for carrier transport in MOSFET. The subthreshold swing of a TFET is less than 60 mV/decade at room temperature and is resistant from short-channel effect [2, 3]. In addition, TFETs show a very small leakage current in the range of femtoamperes (fA) [4] making it suitable for low-power applications [5–6] along with analog and RF applications. Other advantages of TFETs compared to other alternative device concepts are simpler fabrication steps and compatible approach with standard CMOS processing. Moreover, TFETs do not rely on high energetic processes like impact ionization, accepted as scalable device, no punch-through effect, and the tunneling effect enhances the device operating speed and has a small threshold voltage.

The primary constraints concerning TFET devices are ambipolar conduction [7, 8], which originates from the BTBT at the channel–drain junction and other is low ON current, which is considerably poorer than MOSFETs [9]. To compensate the ambipolar conduction, the idea proposed is gate–drain underlap structure [10], which lowers the ambipolar conduction considerably but it reduces the current driving capacity [11]. Therefore, in this paper, we opt for a different approach, where 2-D TFET devices are simulated to analyze the influence of drain-doped engineering [12]. The structures are named according to the corresponding location of drain, Splitted-Drain Single-Gate TFET (SD-SG TFET), when drain is in relative top location, Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET) when drain is in absolute middle location, and Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET) drain in absolute bottom location. In all the structures, drain is splitted into heavily doped region above a lightly doped region. With this arrangement, we can find many desirable developments like suppressing the ambipolar conduction and low leakage current. After analyzing the effects of varying the gate voltage, channel thickness, and oxide permittivity on the drain characteristics, the ON current improves and records very low OFF current. These parameters including drain characteristics and band diagram are analyzed for all the above structures.

2 Device Structure and Parameters

The 2-D cross-sectional view of the p-channel TFET structures with absolute net doping profile is shown in Fig. 1, (a) Splitted-Drain Single-Gate TFET (SD-SG TFET), (b) Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET), and (c) Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET) using Silvaco, Atlas. The structures contain two separate drain regions of different concentrations instead of a single-doped drain region present in a conventional TFET. The highly doped drain region is above the lightly doped region. The doping specifications of the TFET models are P+ type source region (1×10^{20} cm^{-3}) and p-channel region (1×10^{17} cm^{-3}), which are kept constant for all the simulations. Such doping concentration creates a tunnel junction between source and channel where the phenomenon of interband

tunneling occurs for the TFET to conduct in N-mode. The doping specification of the upper N+ doped drain region is 5×10^{18} cm^{-3}, and drain electrode has the Ohmic contact with this drain region. The lower N− doped drain region is 1×10^{17} cm^{-3}. Both disparate drain regions are variably doped to attain an optimal doping profile after analyzing the simulated graphs of all the characteristics. The device is simulated with lateral dimensions of 70 nm, 50 nm, and 70 nm for source, gate, and drain, respectively, and a total device length of 250 nm. A 30 nm oxide overlapped depletion region is present in both sides of source and drain, and this region is to reduce impact ionization and improve ON current. The vertical dimensions of Si ground plane and SiO$_2$ spacer are 20 nm and 20 nm, respectively. The lengths of the both parted drain regions D1 and D2 are varied to get the optimal dimensions for better characterizations and in these structures, both drain parts of vertical lengths are 5 nm. The dielectric used for gate electrode is SiO$_2$, which has a thickness of t$_{ox}$ = 1 nm and dielectric constant is 3.9.

3   Simulation Setup

All the simulations are performed using Silvaco Atlas version 5.20.2.R by including appropriate models such as nonlocal BTBT (BBT.NONLOCAL), Lombardi mobility model (CVT), bandgap narrowing model, Fermi–Dirac statistics, and the Shockley–Read–Hall and Auger recombination models, and drift–diffusion carrier transport is used. The ON and OFF current is evaluated and compared for all the structures for optimal biasing range of V$_{ds}$ up to 1 V and device characteristics is plotted for different V$_{gs}$ = 0.4, 0.5, and 0.6 V, and their performance execution is discussed in later sections.

4   Results and Discussion

4.1   Drain Output Characteristics

Figure given below compares the drain output characteristics for the three structures, SD-SG TFET, MSD-SG TFET, and BSD-SG TFET, for effective channel length of 50 nm, 60 nm, and 70 nm, respectively. Drain voltage V$_{ds}$ is swept to 1 V and obtained characteristics for 0.4, 0.5, and 0.6 V gate bias voltage, and the device is turned ON after V$_{gs}$ = 0.5 V. From the graphs below, it can be observed that the structure of BSD-SG TFET has highest I$_{ON}$/I$_{OFF}$ ratio due to higher effective channel length and recorded OFF current, starting from 1×10^{-21} A indicating minimum leakage. MSD-SG TFET and SD-SG TFET also have much lower I$_{OFF}$, 1×10^{-19} and 1×10^{-18}, respectively, compared to other conventional TFETs, where I$_{OFF}$ starts around
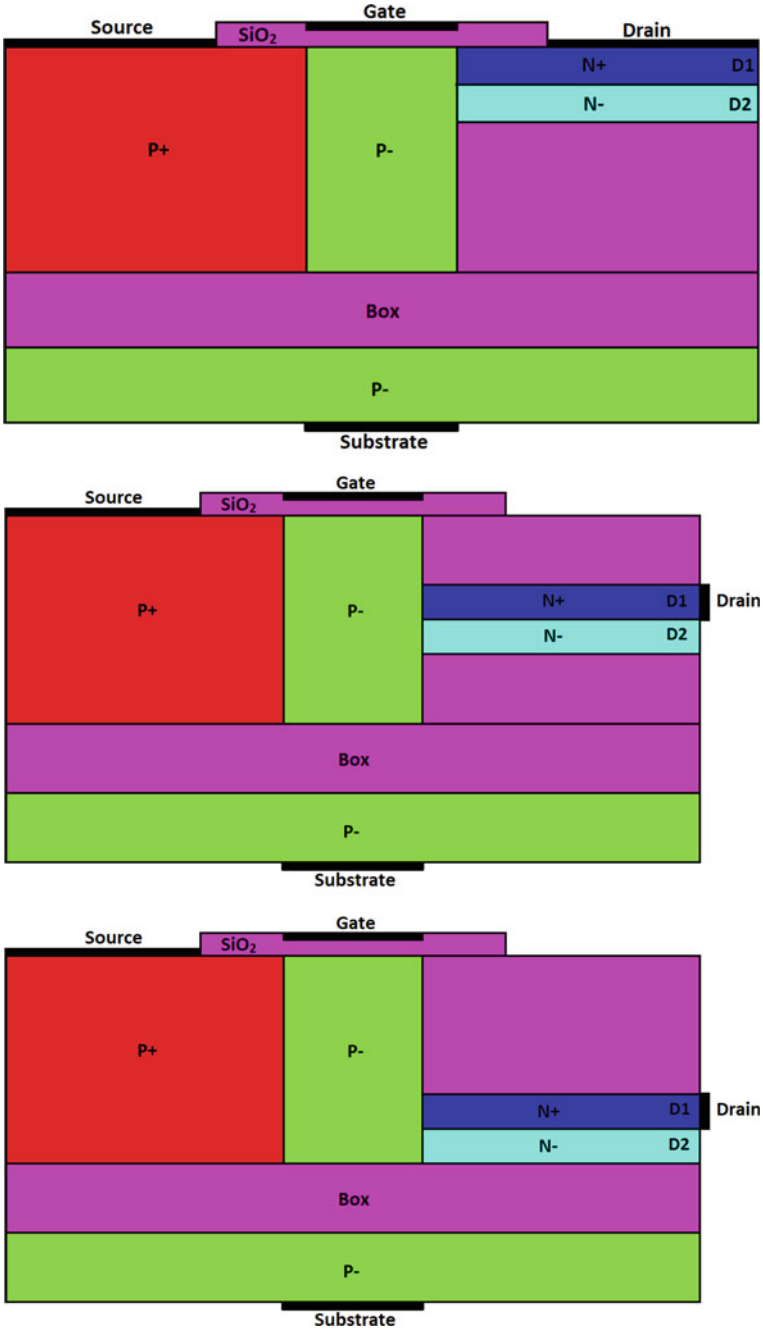
Fig. 1 Structure and doping profile of **a** Splitted-Drain Single-Gate TFET (SD-SG TFET), **b** Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET), and **c** Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET)

in the range of 1×10^{-16} A to 1×10^{-17} A. Therefore, all the proposed structures have higher I_{ON}/I_{OFF} ratio than another standard TFET (Fig. 2).

4.2 Optimization of Ambipolar Conductivity

TFET exhibits unwanted ambipolar conduction due to the inability to form tunnel junction at source–channel and channel–drain interfaces. To compensate the ambipolar conduction, gate–drain underlap structure is proposed but it lowers the current driving capacity; thus, these proposed drain engineered structures have advantage over those underlapped structures. If the above half of the drain region is highly doped than the lower half, the tunneling width at the channel–drain interface increases which in turn reduces the ambipolar current without hampering the ON state current in the TFET. The reduction of the ambipolar current can be observed from the density of states switch band diagram. Figure 3 shows the band diagram of the three structures. The subthreshold slope for tunneling depends on the steepness of the band edges, and in Fig. 3 the highest point of valence band is almost parallel to the lowest point of conduction band for BSD-SG TFET; as per the switching principle, there is minimum possibility of ambipolar conduction by modulating the tunnel barrier by drain-doping engineering.

5 Conclusion

In this proposed work, the effects of the drain-doping engineering on the characteristics and parameters of a TFET have been probed. A splitted-drain structure is used where the highest doped part is above the lower doped part. Three structures have been developed according to the relative position of the splitted-drain. For structure BSD-SG TFET, when the drain is located on bottom, shows the highest drive current, significantly low OFF current and better **I_{ON}/I_{OFF}** followed by MSD-SG TFET and then SD-SG TFET; all structures indicate better characterization than conventional TFETs. Also, the ambipolar conduction is reduced due to the proposed parted drain structure which can be observed from the density of states switch band diagram of the proposed structures. This outcome can be further explored and used in the field of low power, RF, and analog applications.

References

1. Khatami Y, Banerjee K (2009) Steep subthreshold slope n-and p-type tunnel-FET devices for low-power and energy-efficient digital circuits. IEEE Trans Electron Devices 56:2752–2761

(a)

(b)

(c)
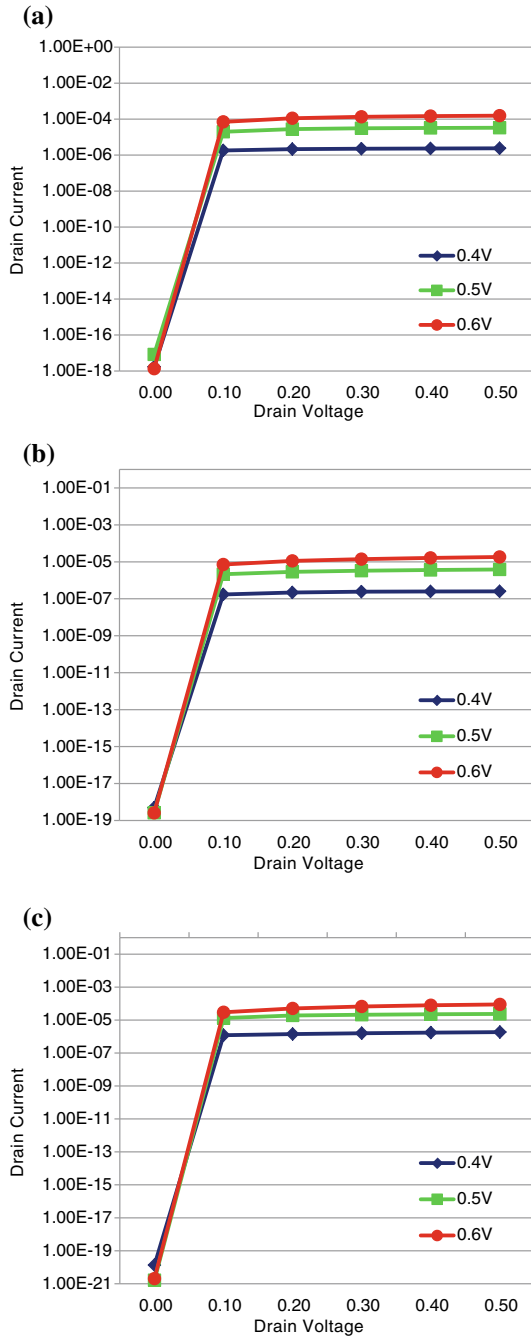
Fig. 2 Drain output characteristics of **a** Splitted-Drain Single-Gate TFET (SD-SG TFET), **b** Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET), and **c** Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET)

Fig. 3 Density of states switch band diagram of Splitted-Drain Single-Gate TFET (SD-SG TFET), Mesial-Splitted-Drain Single-Gate TFET (MSD-SG TFET), and Basal-Splitted-Drain Single-Gate TFET (BSD-SG TFET)
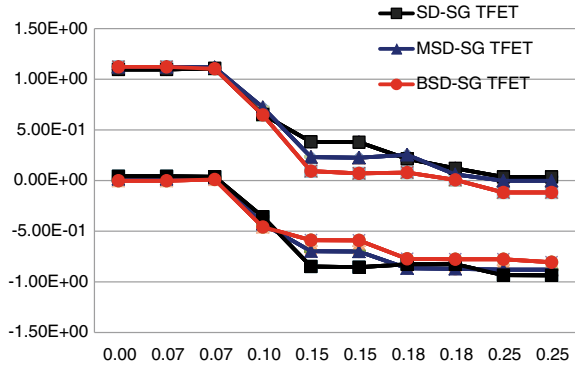
2. Ionescu AM, Riel H (2011) Tunnel field-effect transistors as energy efficient electronic switches. Nature 479(7373):329–337
3. Seabaugh AC, Zhang Q (2010) Low-voltage tunnel transistors for beyond CMOS logic. Proc IEEE 98(12):2095–2110
4. Appenzeller J, Lin Y-M, Knoch J, Avouris P (2004) Band-to-band tunneling in carbon nanotube field-effect transistors. Phys Rev Lett 93(19):196805
5. Saurabh S, Kumar MJ (2011) Novel attributes of a dual material gate nanoscale tunnel field-effect transistor. IEEE Trans Electron Devices 58(2):404–410
6. Lattanzio L, De Michielis L, Ionescu AM (2012) Complementary germanium electron–hole bilayer tunnel FET for sub-0.5-V operation. IEEE Electron Device Lett 33(2):167–169
7. Bagga N, sarkhel S, Sarkar SK (2015) Recent research trends in gate engineered tunnel FET for improved current behavior by subduing the ambipolar effects: a review. In: 2015 IEEE international conference on computing, communication and automation (ICCCA), pp 1264–1267
8. Sarkhel S, Bagga N, Sarkar SK (2015) Compact 2D modeling and drain current performance analysis of a work function engineered double gate tunnel field effect transistor. J Comput Electron 15(1):104–114
9. Bhuwalka KK, Schulze J, Eisele I (2005) Scaling the vertical tunnel FET with tunnel band gap modulation and gate work function engineering. IEEE Trans Electron Devices 52(5):909–917
10. Verhulst AS, Vandenberghe WG, Maex K, Groeseneken G (2007) Tunnel field-effect transistor without gate-drain overlap. Appl Phys Lett 91(5):053102
11. Raad B, Nigam K, Sharma D, Kondekar P (2016) Dielectric and work function engineered TFET for ambipolar suppression and RF performance enhancement. Electron Lett 52(9):770–772
12. Shaker A, El Sabbagh M, Mohammed M. El-Banna (2017) Influence of drain doping engineering on the ambipolar conduction and high-frequency performance of TFETs. IEEE Trans Electron devices 64(9)

Implementation of a Doped Pocket Region in Order to Enhance the Device Performance of MOSFET

Prasenjit Saha and Bijoy Goswami

Abstract The majority carrier concentration in the channel of an n-MOSFET is increased by inserting a moderately doped pocket region at the bottom of the substrate which is partially below the source and gate. Base of the substrate is insulated by SiO_2, so that there is least possibility of leakage through it. The device designed is of 20 nm channel length, and a simulation-based study has been carried out. In order to perform analysis and comparison, the same is designed with 30–40 nm channel length and conventional MOSFET of 30–40 nm is designed. Device characteristics obtained from the simulation shows that it follows characteristics of the conventional MOSFET. On comparing the drain current of the modified MOSFET with the conventional one, the modified structure performs better. On performing leakage current analysis, it is found that the modified MOSFET has a smaller value than the conventional MOSFET.

Keywords Conventional MOSFET · Modified MOSFET · Leakage current · Channel length

1 Introduction

The demand for faster cheaper and smaller devices has led to the downscaling of MOSFETs to well below 30 nm for VLSI application, and has successfully been fabricated and characterized [1, 2]. However, with a decrease in the channel length, the device performance is found to be continuously degrading, because many physical barriers arise like short-channel effects, drain-induced barrier lowering, punch through, etc. [3, 4]. One of the main technological roadblocks when we scale these devices beyond some limit is that the high mechanical leakage current penetrates through the gate oxide and it increases the power dissipation [5]. The direct tunneling current increases exponentially with the decrease in oxide thickness which is of

P. Saha (✉) · B. Goswami
Department of ETCE, Jadavpur University, Kolkata, India
e-mail: prasenj1ps@gmail.com

Fig. 1 Modified MOSFET
structure of 22 nm × 60 nm
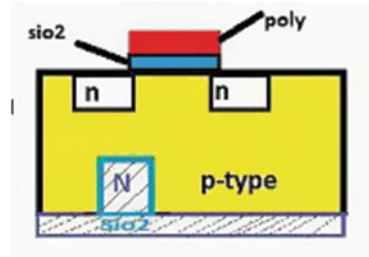
primary concern of MOS scaling [6, 7]. From research of recent years, it is seen that
the modified structures like double gate, triple structure, and gate around structures
are found to be more promising than the conventional model. The structure is shown
in Fig. 1.

2 Device Structure and Simulation Setup

In the n-MOSFET, the majority carriers in the channel are electrons. The drain current
at a corresponding gate and drain voltage can be increased if the electron density
in the channel increases, correspondingly the hole concentration decreases. In order
to enhance the electron concentration in the channel, an n-type moderately doped
region is inserted in the substrate [8]. The doped region in the substrate is positioned
in such a way that it is below the half of source and the channel. The holes in the
channel are attracted by the introduced moderately doped region in the substrate,
thereby increasing the electron concentration in the channel. Since the doped region
introduced is layering in the substrate, there may be some leakage. Therefore, the
base of the substrate is insulated by SiO_2, which is of 1 nm thickness, so that there is
no possibility of leakage through the substrate [9, 10]. Figure 2 shows the structure
formation.
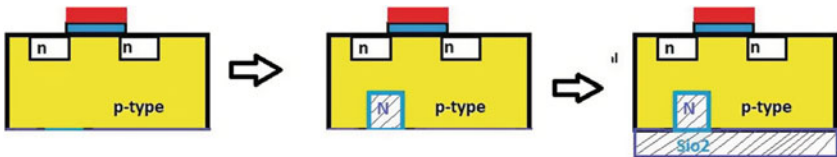
Fig. 2 Structure formation

2.1 Fabrication Steps

A moderately doped region is inserted in a p-type silicon substrate using N-well process. After the formation of the SiO_2 layer using oxygen ion implantation technology [11], the structure is first rotated 180° vertically [12] and horizontally. Now using the fabrication of conventional n-MOSFET, the complete structure can be developed. Figure 3 shows the fabrication steps.

2.2 Simulation Setup

A 20 nm n-MOSFET with the doped region in the substrate is designed using the ATLAS simulator of the Silvaco TCAD, and then the same device has been also designed for the channel length of 30–40 nm to a make comparison with varying channel length. 30–40 nm conventional MOSFET is designed to find the comparison with the new doping of MOSFET. Figures 4 and 5 show the structure designed.
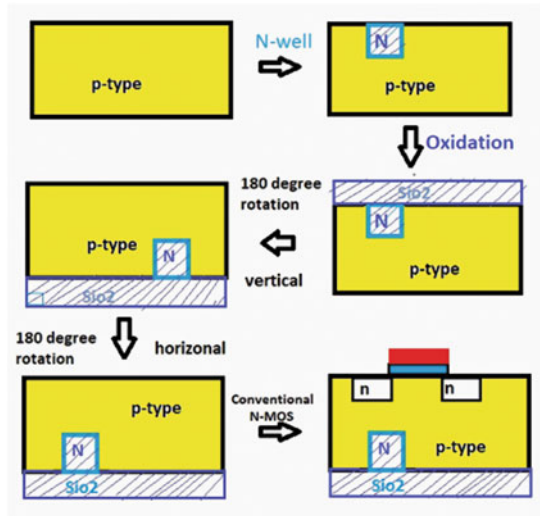
Fig. 3 Fabrication steps
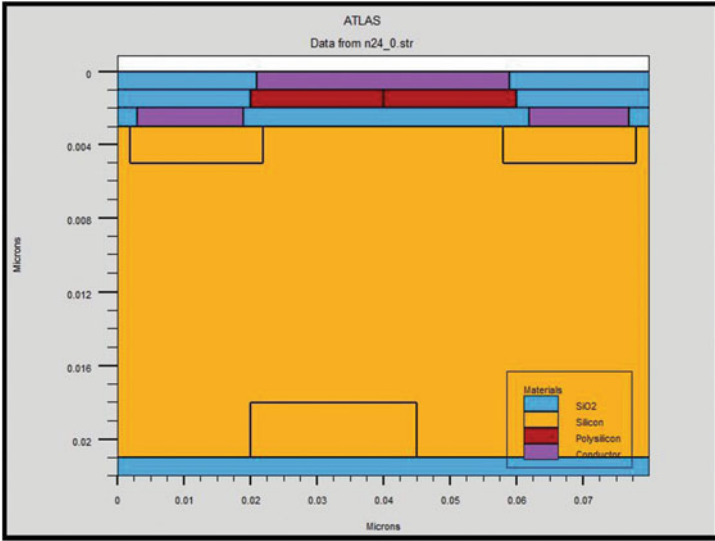
Fig. 4 40 nm modified MOSFET
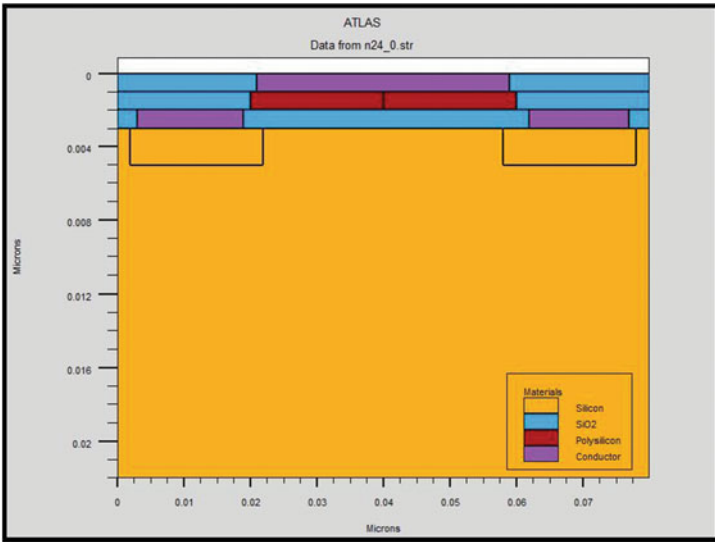
Fig. 5 40 nm conventional MOSFET

3 Results and Discussions

3.1 Drain Biasing

The drain is biased with $V_d = 0.1$ V and the gate voltage is varied with 0–3 V to obtain the I_d versus V_g characteristics for 20, 30, and 40 nm modified n-MOSFET. Figure 6 shows the I_d–V_g characteristics.

The graphs show the I_d–V_g characteristics of the modified n-MOSFET for 40, 30, and 20 nm channel length. On comparing them, we can observe that as channel length decreases the slope of the curve increases which follows the characteristics of the conventional MOSFET.

3.2 Gate Biasing

The gate is biased with $V_g = 1.1, 2.2$, and 3.3 V, and the drain voltage is varied from 0 to 3 V to obtain the I_d versus V_{ds} curve. And a comparison is done among 20, 30, and 40 nm.

The graphs showing the I_d–V_d characteristics of the modified n-MOSFET for 20 nm channel length is shown in Fig. 7. On comparing them, we can observe that
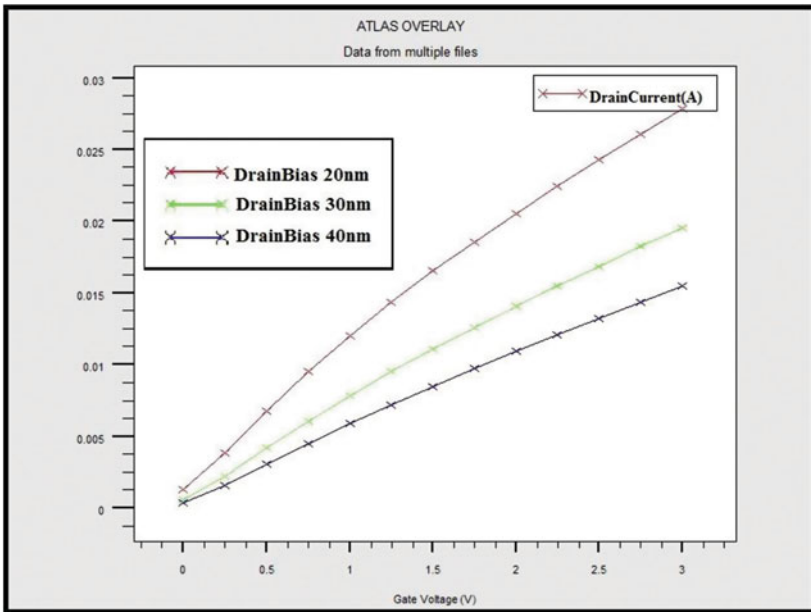
Fig. 6 Drain current obtained at $V_g = 3$ V
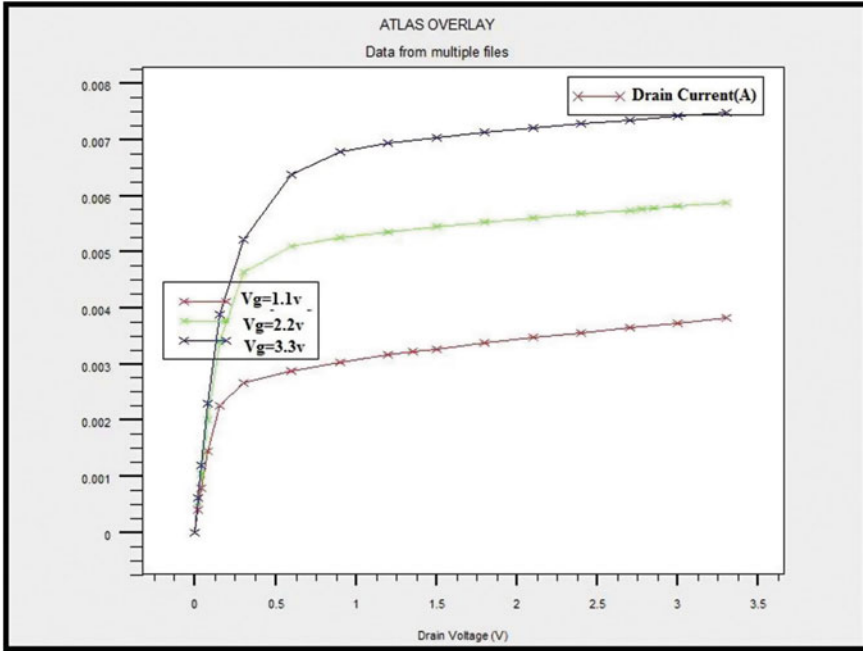
Fig. 7 I_d–V_d characteristics for 20 nm channel length

as channel length decreases the slope of the curve increases, thus yielding a higher drain current and it follows the characteristics of the conventional MOSFET.

3.3 Comparison of Conventional MOSFET and Proposed MOSFET

Drain Biasing: Drain biasing is performed on both the structures and a comparison is made for 40 nm channel length. Figure 8 shows the results obtained for the comparison on the drain biasing.

The slope of the characteristics is equal to 1/Ron. If we compare both, we can see that the new MOSFET has a greater slope than the conventional MOSFET; hence, it has a greater drain current than the conventional one.

Gate Biasing: The gate is biased with $V_g = 1.1$, 2.2, and 3.3 V, and the drain voltage is varied from 0 to 3 V to obtain the I_d versus V_{ds} curve. And a comparison is done among 40 nm. Both have a 40 nm channel length, which follows the characteristics of the conventional MOSFET. On comparing the characteristics of both the devices, we can see that the new MOSFET has a greater slope than the conventional
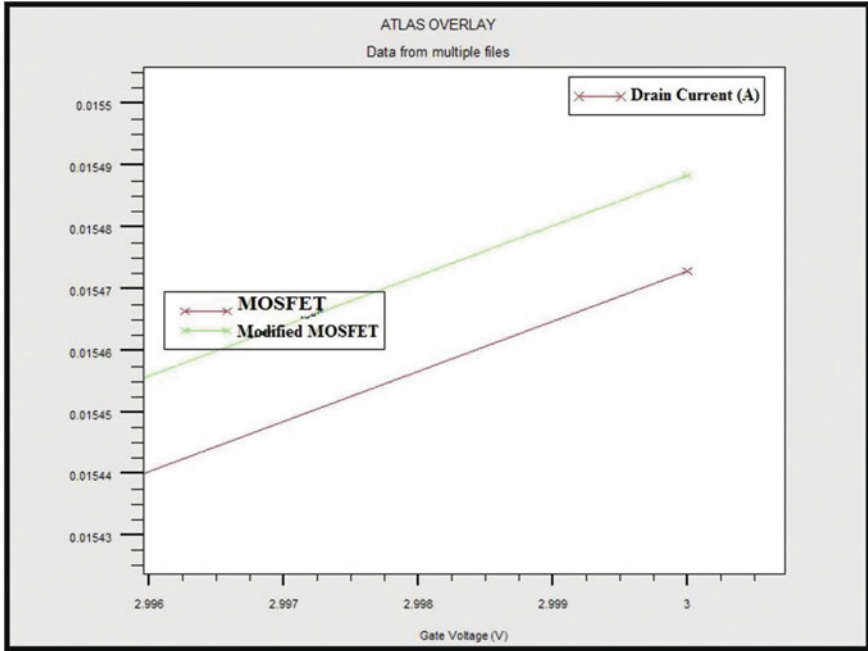
Fig. 8 I_d versus V_g curve for 40 nm

MOSFET in all gate voltages of 1.1, 2.2, and 3.3 V; hence, it yields a greater drain current than the conventional one.

Leakage Current Analysis: Leakage current analysis is performed on the MOS-FETS with channel length of 30 nm and 40 nm and a comparison is done between the devices at respective channel lengths. The results obtained for 30 nm are shown in Fig. 9.

It is observed that the leakage current is minimized; in case of 30 nm it is 1.154 mA less, and in case of 40 nm it is 5 uA less than the conventional MOSFET.

4 Conclusion

The results obtained from the drain and gate biasing signify that the MOSFET with the newly introduced doped region obeys the characteristics of the conventional MOS-FET. On comparing the drain current of modified MOSFET with the conventional MOSFET at different conditions of gate voltage, drain voltage, and channel length, the modified MOSFET has a higher value, and thus the gain will be higher. The leakage current of the modified MOSFET is found to be less than the conventional MOSFET, and hence the power consumption will be minimized.
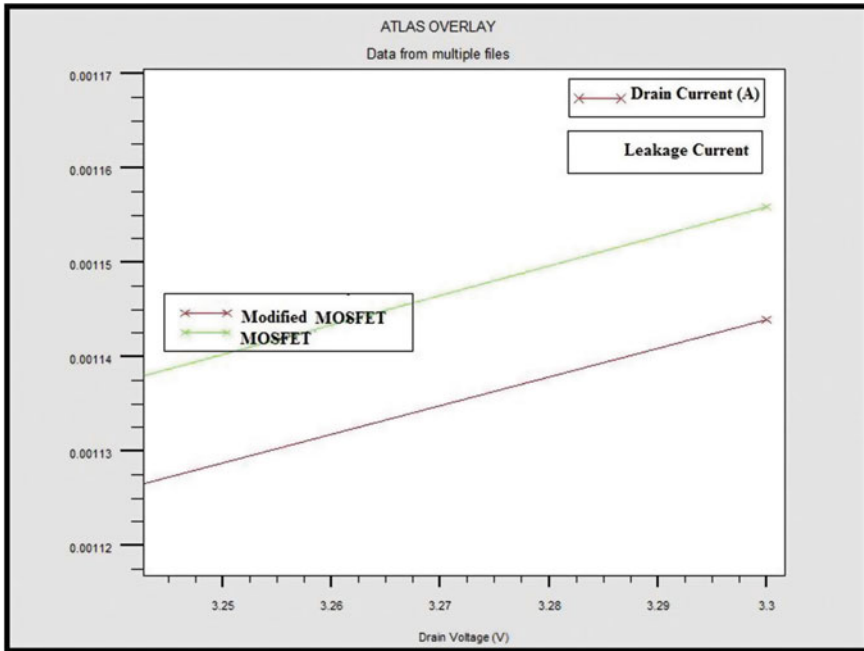
Fig. 9 Leakage current comparison at 30 nm channel length

References

1. Emerging research devices, International technological roadmap semiconductors (2009)
2. Packan P (1999) Pushing the limits. Science 285(5436):2079–2081
3. Banerjee P, Sarkar SK (2017) 3-D analytical modelling of dual-material triple gate silicon-on-nothing MOSFET. IEEE Trans Electron Devices 64(2)
4. Shrivastava A, Tripathi N, Agarwal E, Singh A (2015) Comparative study of double gate soi MOSFET and single gate soi MOSFET through simulation. IJEEE 07(01). ISSN: 2321-2055 (E)
5. Sacconi F, Carlo AD, Lugli P, Stadile M (1980) Full band approach to tunneling in MOS structures. IEEE Trans Electron Devices 22(11)
6. Chang L, Yeo YC, Polishchuk I, King TJ, Hu C (2002) Direct-tunneling gate leakage current in double-gate and ultrathin body MOSFETs. IEEE Trans Electron Devices 49(12)
7. Choi CH, Yu Z, Dutton RW (2001) Impact of gate tunneling current in scaled MOS on circuit performance: a simulation study. IEEE Trans Electron Devices 48(12)
8. Sze SM, Ng KK (2014) Physics of semiconductor devices (Chapter 7), 3 ed, p 376. ISBN: 978-8126517022
9. Sano E, Kasai R, Ohwada K, Ariyoshi H (1980) A two dimensional analysis for MOSFET's fabricated on buried SiO_2 layer. IEEE Trans Electron Devices 22(11)
10. Duan, B, Yang Y (2011) REBULF super junction MOSFET with N+ buried layer. Micro Nano Lett
11. Ohwada K, Omura Y, Sano E (1981) A high-speed buried channel MOSFET isolated by an implanted silicon dioxide layer. IEEE Trans Electron Devices 28(9)
12. Gandhi SK (2013) VLSI Fabrication and principles: silicon and gallium arsenide (Chapter 11), 2nd ed, pp 710–711. ISBN: 978-8126517909

Design of Digital-to-Analog Converter Using Dual Pair Differential Amplifier in 180 nm CMOS Technology

Preeti Prajapati and Trailokya Nath Sasamal

Abstract Designing Digital-to-Analog Converter (DAC) using dual differential pair as a current-to-voltage converter has been analysed in this paper. Current-steering topology is used since it enables high conversion rate and linearity. The proposed circuit uses binary-weighted architecture rather than segmentation as binary-weighted structure has high SFDR and less tolerant to poor design. It also uses rail-to-rail operation using Gilbert cell. This rail-to-rail circuit provides 60 dB gain with a phase margin of 63°. A current range of 89.79 μA–8.32 mA is obtained with total power consumption of 15 mW. The design and simulations are done using Cadence Virtuoso on 180 nm CMOS technology.

Keywords DAC · Cadence · Low power · Current steering · Gilbert cell · Rail-to-Rail operation

1 Introduction

Data converters are widely used in telecommunication system, video signal processing, cellular network and in many other applications. DAC performs the conversion between digital and analog [1]. DAC can be implemented in several ways that consist of active components and passive components. In early days, resistor array [2] and charge-redistribution DAC [3] are used, though it provides good linearity but it suffers from various drawbacks like large chip size and large power. Some examples of active component are current cell matrix [4], weighted current source [5] and switched-current D/A converters [6]. Though these converters occupy a large area,

P. Prajapati (✉)
School of VLSI Design and Embedded System, National Institute of Technology Kurukshetra, Kurukshetra 136119, India
e-mail: infoseeker1024@gmail.com

T. N. Sasamal
Department of Electronics and Communication Engineering, National Institute of Technology Kurukshetra, Kurukshetra 136119, India

they are implemented to consider device matching [7]. Because of the parallel input, it provides high speed but power consumption is large. The switched-current DAC has simple architecture and has low power consumption. Due to serial input, it has shortcoming at high speed.

Instead of using active and passive component, segmentation architecture can also be used. In [8], segmentation is done in such a way that 5 bits are unary and another 5 bits in binary-weighted form. Similarly, in [9], 12 bits are divided into 7 MSBs that is implemented in equally weighted current sources and the 5 LSBs in binary-weighted current sources. Segmentation depends on a number of bits used in unary coding, leads to complexity and large power, and also degrades the performance of the circuit instead of improving. So rather than using the concept of segmentation, the proposed circuit uses only binary-weighted structure that achieves high SFDR at high clock frequency. The power consumption is low and simple in design and achieves compatible static and dynamic performance.

The current-steering topology has number of advantages like high resolution, high accuracy, small chip area, compatibility with CMOS technology and optimum power consumption [10]. So, the proposed circuit uses current-steering binary implementation that has merits of both, viz., binary architecture and current-steering one. It describes digital part of the circuit.

In typical DACs, operational amplifier acts as a current-to-voltage converter but two-stage operational amplifiers comprise differential stage which provides high gain stage to the system. Two-stage op-amps have two poles. Due to this, system is unstable and phase margin is low that creates oscillations in the system. So instead of using op-amp, dual differential amplifier called Gilbert cell is used which performs the same function. Also, to maximize the dynamic range of the circuit and for a favourable SNR in the analog section of a system-on-chip, rail-to-rail operation is used that is achieved by placing dual differential pair of p-channel and n-channel MOSFET in parallel. It resolves the problem of limited voltage headroom and poor dynamic range since the signal has to be kept as wide as possible. Rail-to-rail operations find widespread application in video signal processing and multimedia systems [11, 12]. This depicts the analog part of the circuit.

The remainder of paper is organized in the following section: Sect. 2 describes the system architecture. In Sect. 3, simulation results are shown and compared with the other DACs. Finally, conclusion is described in Sect. 4.

2   System Architecture

Current-steering architecture offers high speed so it is used in almost any field. Since it requires no buffer and uses load resistor directly for current, the proposed DAC consists of analog part and digital part. Switch driver, switch, array and a current source array implement the analog part of the circuit, while decoder, the latches and a clock driver defines the digital part of the circuit. It is a binary-weighted DAC that

uses a repeating cascaded structure of current source with a dual differential amplifier which acts as a current-to-voltage converter.

2.1 Current Source

It is a basic unit of DAC that directly affects the performance. In the proposed circuit, cascade stage is used since it acts as a constant current source. It has high output impedance that is useful for static and dynamic performance. It also reduces transient voltage fluctuations at node M7 of the current source. This 10-bit DAC uses 10 current source cells. Every cell has cascaded current source with a latch circuit and then carries a current corresponding to 1LSB and increases in binary-weighted form. V_{b1}, V_{b2} and V_{b3} are bias voltages and their values find out by calculations are given as 0.85 V, 1.4 V and 1.65 V, respectively.

2.2 Switch Driver

The size of switch is small so that parasitic capacitance is less. The schematic of switch is shown in Fig. 1. The deglitching circuit on the current source adjusts the crossing point of two switch controls so that differential switches are not turned off simultaneously. Transistors M2 to M5 form the differential Goliath inverters and inverters U1, U2 form the differential David inverters [13]. Transistor M3, M6 acts as inverter for the right part, and transistor M1, M5 acts as inverter for left part of differential circuit. The left inverter buffers the digital input, so it acts as an inverting buffer. If size of the buffer is increasing, then switching speed also increases. So in this work, there is a trade-off for speed. This latch is faster, simpler and consumes low power. Transistors M2, M7, M9 to M12 act as a cascode stage. Input is *INL* and its complement is *INR*. *INL* and *INR* are the inputs that are applied through M1 and M8, respectively. *IL* and *IR* are the currents that are applied in NMOS and PMOS Gilbert cell as input.

2.3 Gilbert Cell Design

With reference to the conventional Gilbert cell topology, dual differential pair with active load was designed using p-channel and n-channel MOSFET as described in Fig. 2 and Fig. 3, respectively.

It uses the current mirror topology. The input stage of Gilbert cell shows a non-linear behaviour that becomes critical at lower bias current, and hence the linearity of Gilbert cell is not good. So we can achieve better linearity using current mirror since it shows linear behaviour regardless of the bias current [14]. The transistors M5, M6
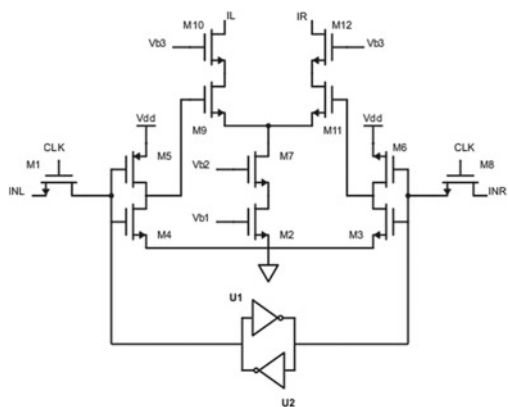
Fig. 1 Schematic of switch
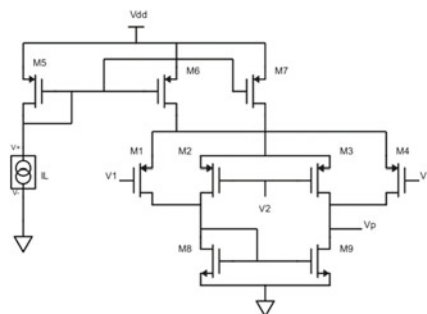
Fig. 2 MOS version of Gilbert cell using p-channel transistor
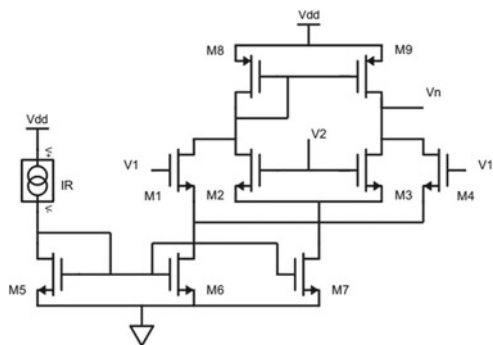
Fig. 3 MOS version of Gilbert cell using n-channel transistor
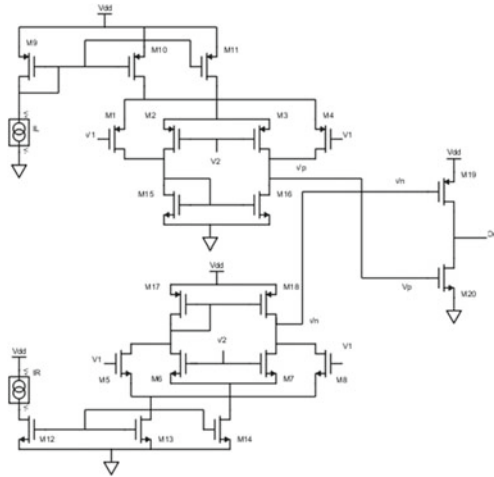
Fig. 4 Rail-to-rail Gilbert cell

and M7 act as a current mirror for n-type and p-type, respectively, for a wide range of input voltage. This results in a constant transconductance so that a good linearity performance can be achieved.

The current rail also provides sufficient current to drive Gilbert cell in saturation region. There are two differential inputs named as V1 and V2 applied to the cross-connected differential pairs. The above Gilbert cell has dual differential amplifier stages that consist of transistors M1–M4 for n-type and p-type in which transistors M5–M7 used for current mirror and transistors M8–M9 acts as active load depicts in Fig. 2 and Fig. 3, respectively. Currently, *IL* and *IR* are used to drive the input stage of PMOS and NMOS Gilbert cell. So *IL* and *IR* are the currents that split between the differential pairs. Gilbert cell has two important properties. The first one is that the tail current drives the small signal gain of the circuit. The second property tells that the two transistors in a differential pair provide a simple means of steering the tail current to one of the two destinations [15].

2.4 Rail-to-Rail Operation

The outputs V_n and V_p are connected in a cross-coupled common source configuration that allows full rail-to-rail input range. A p-channel pair of Gilbert cell is in parallel with an n-channel pair of Gilbert cell so that one pair is conducting at a time and covers full input common mode range as depicted in Fig. 4.

The input common mode range of n-type MOSFET is

$$V_{cn} > V_{ss} + V_{gsn} + V_{dsn}$$

where V_{gsn} represents the voltage drop between gate to source and V_{dsn} represents the voltage drop between drain–source of the NMOS transistor, respectively.

Similarly, the input common mode range of p-type MOSFET is given as follows:

$$V_{cp} < V_{DD} - V_{gsp} - V_{dsp}$$

where V_{gsp} is the voltage drop between gate and source and V_{dsp} represents the voltage drop between drain–source, respectively.

For full rail-to-rail input, at least one of the input pairs should be on that means

$$V_{cp} > V_{cn}$$
$$\text{i.e } V_{DD} - V_{SS} > V_{gsp} + V_{dsp} + V_{gsn} + V_{dsn} > 2V_{th}$$

There exist a common mode input voltages in which both pairs are on. In this case, it creates a serious issue. It can be observed that the transconductance is almost double when only one of the input pairs is on and is given as follows:

$$g_m = 2\sqrt{K\frac{W}{L}I}$$

where g_m is the transconductance, K' is the transconductance parameter, I is the drain current of the transistor, W is the width and L is the length of the transistor, which makes optimal frequency compensation very difficult. For improvement of the gain and to enhance signal-to-noise ratio, input stage should have large transconductance.

3 Simulation Results

We have used Cadence spectre for simulation using Cadence Virtuoso tool for standard 180 nm CMOS process under 1.8 V supply voltage. We have also included current mirror circuit in our simulations. Threshold voltages of -0.5 V and 0.45 V are considered for PMOS and NMOS, respectively. The transient output waveform of the circuit is shown in Fig. 5. The plot of spurious-free dynamic range (SFDR) versus frequency is depicted in Fig. 6. The net power consumption is 15 mW. Frequency response of rail-to-rail circuit is also shown in Fig. 7. This rail-to-rail circuit provides 60 dB gain with a phase margin of 63°. Table 1 summarizes the comparison of results of proposed circuit from previously reported DAC.
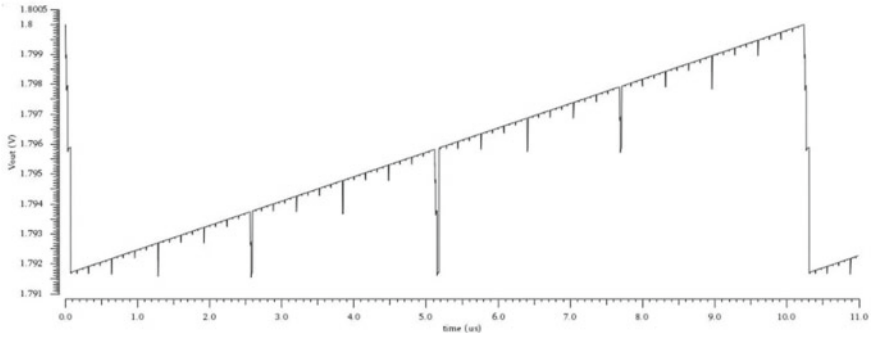
Fig. 5 Transient output waveform

Fig. 6 Measured SFDR versus frequency
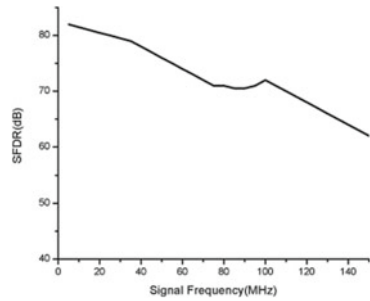
Fig. 7 Frequency response of rail-to-rail circuit
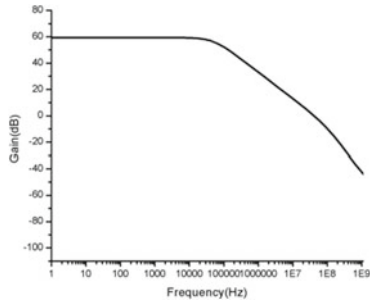
Table 1 Comparison of results from previously reported DAC

Parameters	This work	Ref. [16]	Ref. [17]
Technology (nm)	180	180	180
Supply voltage (V)	1.8	1.8	1.8
Resolution (Bits)	10	10	10
Sample rate (MS/s)	100	250	250
SFDR (dB)	82	58	61.2
Power (mW)	15	19	22
FOM (10^9)	6826	13473	11636

4   Conclusion

In this brief, we developed a new architecture for implementing digital-to-analog converter using dual differential pair. From the analysis, it can be inferred that binary-weighted architecture has very simple decoder and due to this, it consumes less power. Our result also shows that the proposed circuit has SFDR of more than 50 dB.

References

1. Myderrizi Indrit, Zeki Ali (2010) Current-steering digital-to-analog converters: functional specifications, design basics, and behavioral modeling. IEEE Antennas Propag Mag 52(4):197–208
2. Pelgrom, Marcel (1990) A 50 MHz 10-bit CMOS digital-to-analog converter with 75 Omega buffer. In: 1990 37th IEEE international solid-state circuits conference, ISSCC 1990. Digest of technical papers. IEEE
3. Tsai C-C et al (2005) 10-bit switched-current digital-to-analogue converter. IEE Proc-Circuits Devices Syst 152(3):287–290
4. Nakamura, Yasuyuki, et al (1991) A 10-b 70-MS/s CMOS D/A converter. IEEE J Solid-State Circuits 26(4):637–642
5. Fong, Leong K, Salama CAT (1993) A 10 bit semi-automatic current mode DAC. In: 1993 IEEE international symposium on circuits and systems ISCAS'93. IEEE
6. Traff H, Holmberg T, Eriksson S (1991) Application of switched-current technique to algorithmic DA-and AD-converters. IN: IEEE international symposium on circuits and systems. IEEE
7. Tsai C-C et al (2005) 10-bit switched-current digital-to-analogue converter. IEE Proc-Circuits Devices Syst 152(3):287–290
8. den Bosch V, Anne et al (2001) A 10-bit 1-GSample/s Nyquist current-steering CMOS D/A converter. IEEE J Solid-State Circuits 36(3):315–324
9. Sullivan O, Kevin et al (2004) A 12-bit 320-MSample/s current-steering CMOS D/A converter in 0.44 mm/sup 2. IEEE J Solid-State Circuits 39(7):1064–1072
10. Reda Ahmed, Ibrahim Mohamed F, Farag Fathi (2016) Input–output rail-to-rail CMOS CCII for low voltage–low power applications. Microelectron J 48:60–75
11. Moldovan L, Li HH (1997) A rail-to-rail, constant gain, buffered op-amp for real time video applications. IEEE J Solid-State Circuits 32(2):169–176
12. Wu W-CS et al (1994) Digital-compatible high-performance operational amplifier with rail-to-rail input and output ranges. IEEE J Solid-State Circuits 29(1):63–66
13. Deveugele J, Steyaert MSJ (2006) A 10-bit 250-MS/s binary-weighted current-steering DAC. IEEE J Solid-State Circuits 41(2):320–329
14. Wan Qiuzhen, Wang Chunhua (2011) A low-voltage low-power CMOS transmitter front-end using current mode approach for 2.4 GHz wireless communications. Microelectron J 42(5):766–771
15. Razavi B (2001) Design of analog CMOS integrated circuit
16. Chou F-T, Chen Z-Y, Hung C-C (2014) A 10-bit 250MS/s low-glitch binary-weighted digital-to-analog converter. In: 2014 27th IEEE international system-on-chip conference (SOCC). IEEE
17. Deveugele J, Steyaert MSJ (2006) A 10-bit 250-MS/s binary-weighted current-steering DAC. IEEE J Solid-State Circuits 41(2):320–329

An Analytical Surface Potential Model for Highly Doped Ultrashort Asymmetric Junctionless Transistor

Nipanka Bora and Rupaban Subadar

Abstract Here, we present an analytical solution for surface potential of a heavily doped ultralow channel Double-Gate Asymmetric Junctionless Transistor (DG AJLT). The gate-oxide-thickness and flatband voltage asymmetry were taken into considerations; further, while solving 2D Poisson's equation both fixed and mobile charges in the silicon region regions were considered. To solve the 2D Poisson equation for the asymmetric DG junctionless transistor, we separate the solution of the channel potential into basic and perturbed terms. The equations derived from a general symmetric DG junctionless transistor are considered as basic terms, and using Fourier series a solution related to the perturbed terms for the asymmetric structures was obtained. The electrical characteristics predicted by the analytical model shows an excellent agreement with that of commercially available 3D numerical device simulators.

Keywords Analytical model · Perturbation technique · Asymmetric double-gate junctionless transistor (ADG JLT) · Surface potential

1 Introduction

A Junctionless Transistor (JLT) does not have the very high doping concentration gradients or junctions at the source and drain junctions, thereby making fabrication processes easier than MOSFETs with junctions. JLTs has many other advantages like low OFF-state currents, near ideal subthreshold slope (SS ~ 60 mV/dec), high ON-state to OFF-state current ratio, DIBL effects, etc. [1]. Thus, JLTs are potential candidate for sub 20 nm technology nodes and beyond. There are several models for JLT available for the Symmetric Double-Gate (SDG) structures. Colinge et al. [2] proposed the first JL transistor. Duarte et al. [3] studied the electrostatic behavior

N. Bora (✉) · R. Subadar
Department of Electronics and Communication Engineering, North-Eastern Hill University,
Shillong 793022, Meghalaya, India
e-mail: nbora@nehu.ac.in; nipankabora@gmail.com

of JL DG transistor using analytical modeling scheme. Subsequently, Lin et al. [4] developed an analytical model of an electric potential of a double-gated fully depleted junctionless transistor. A universal model for symmetric double-gate JL transistor has been reported [5]. However, all these models have been derived by assuming both gates to be perfectly symmetric but in reality, this may not be possible due to process variations and uncertainties which can affect surface potential and other parameters of SDG junctionless transistor. The JLT can be seen as an Asymmetric Double-Gate (ADG) device structure. Thus, a suitable analytical model is very much essential in incorporating these effects. Lu and Taur [6] proposed analytical models for an asymmetric DG MOSFET to reflect these variations. However, these analytical models work with a 1D Poisson equation, which is valid for the potential distribution of long-channel devices; such models are not suitable for the characteristics of asymmetric short-channel devices. A very few analytical models are available in the literature like the model by Jin et al. [7] for asymmetric model based on symmetric structure for junctionless transistors which cannot be considered as a full analytical model ADG JLT. Since the working principle for junctionless transistor is different, same models are not valid for ADG JLT and such models are rare in available literature. Moreover, to develop a JLT without source and drain junction regions is to make the transistor as small as possible which will increase the chip functionality. Because of all these reasons, models which can predict the behavior of ADG JLTs should be developed.

2   Model Derivation

The coordinates, x and y, are as shown in Fig. 1. The Poisson's equation in the silicon region considering both mobile and bulk fixed charges can be written as
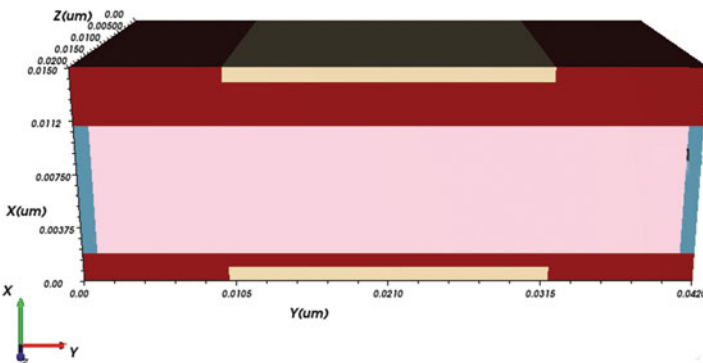
Fig. 1 A 3-D asymmetric junctionless DG MOSFET structure

$$\frac{\partial^2 \psi(x, y)}{\partial x^2} + \frac{\partial^2 \psi(x, y)}{\partial y^2} = \frac{q}{\varepsilon_{si}} \left(-N_b + N_b e^{\frac{q(\psi - \psi_f)}{kT}} \right) \tag{1}$$

Equation (1) has no direct analytical solution. Since n-type JL MOSFET is assumed, the body is doped with n-type impurity and the source and the drain are also heavily doped with n-type impurity, where N_b is the bulk doping concentration, $\psi(x, y)$ is the channel potential, ε_{si} is the silicon permittivity, ψ_f is the electron quasi-Fermi potential, and q is the electron charge. As Eq. (1) is not directly solvable, we will divide the 2D potential as

$$\psi(x, y) = \psi_I(y) + \psi_{II}(x, y) \tag{2}$$

where $\psi_I(y)$ is a 1-D function of y and $\psi_{II}(x, y)$ is a 2-D function of both x and y. Here, $\psi_{II}(x, y)$ is the potential where asymmetric nature has been included. The boundary conditions at the interface between the gate oxide and the silicon body for asymmetric JLT can be written as

$$V_{gs} - V_{fb} + \Delta_{fb} - \psi\left(\frac{t_b}{2}, y\right) + (t_{ox} - \Delta t_{ox})\frac{\varepsilon_{si}}{\varepsilon_{ox}} \frac{\partial \psi(t_b/2, y)}{\partial x} = 0$$

$$V_{gs} - V_{fb} - \psi\left(-\frac{t_b}{2}, y\right) - t_{ox}\frac{\varepsilon_{si}}{\varepsilon_{ox}} \frac{\partial \psi(t_b/2, y)}{\partial x} = 0 \tag{3}$$

where V_{gs}, V_{fb} are gate to source and flat band voltage, respectively, ε_{ox} is the oxide permittivity, t_{ox}, t_b are oxide, and body thickness, respectively, Δt_{ox}, Δ_{fb}, are the asymmetries in the oxide thickness and flatband voltage, respectively. The boundary condition near the drain/source contact can be written as

$$\psi(x, L) = V_{ds} + V_{bi}$$

$$\psi(x, 0) = V_{bi} \tag{4}$$

where L is the gate length, V_{bi} is the built-in voltage between the silicon body and source/drain contact. These two functions should satisfy the following equations and the corresponding boundary conditions, respectively. Now the Poisson's equation for $\psi_I(y)$ can be written as

$$\frac{\partial^2 \psi_I(y)}{\partial y^2} = \frac{q}{\varepsilon_{si}} \left(-N_b + N_b e^{\frac{q(\psi - \psi_f)}{kT}} \right) \tag{5}$$

The solution of this equation for JLTs which normally operates in the depletion region can be given as [8]

$$\psi_I(y) = V_{gs} - V_{th} - \frac{q N_b T_b}{8 C_{si}} - V_{th} \, \text{Lambert W} \left[\frac{q^2 N_b T_b}{4 C_{ox} K T} e^{\left(\frac{q(V_{gs} - V_{th} - \Psi_f)}{KT} \right)} \right] \quad (6)$$

where Lambert W is the Lambert W-function, V_{th} is the threshold voltage, and C_{si}, C_{ox} are Silicon and oxide capacitors, respectively.

The 2D Poisson's equation for $\psi_{II}(x, y)$ is where asymmetric conditions are used can be given as

$$\frac{\partial^2 \psi_{II}(x, y)}{\partial x^2} + \frac{\partial^2 \psi_{II}(x, y)}{\partial y^2} = 0 \quad (7)$$

The solution of the above equation can be obtained as

$$\psi_{II}(x, y) = \sum_{n=1}^{\infty} \left[K_n \cos\left(\frac{n \pi x}{L} \right) + R_n \cos\left(\frac{n \pi x}{L} \right) \right] \sin\left(\frac{n \pi y}{L} \right) \quad (8)$$

where K_n and R_n are the coefficients. They can be expressed as

$$K_n = \frac{Num1 - Num2}{Den1 - Den2} \quad (9)$$

$$Num1 = \left(1 + \frac{n \varepsilon_{si} \pi t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{2L} \right) A_n$$

$$Num2 = \left(1 - \frac{n \varepsilon_{si} \pi t_{ox1}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{2L} \right) B_n$$

$$Den1 = \left(1 + \frac{n \pi \varepsilon_{si} t_{ox1}}{\varepsilon_{ox} L} \right) \left(1 + \frac{n \pi \varepsilon_{si} t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{L} \right)$$

$$Den2 = \left(1 + \frac{n \pi \varepsilon_{si} t_{ox1}}{\varepsilon_{ox} L} \right) \left(1 + \frac{n \pi \varepsilon_{si} t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{L} \right)$$

$$R_n = \frac{Num11 - Num22}{Den11 - Den22} \quad (10)$$

$$Num11 = \left(1 - \frac{n \varepsilon_{si} \pi t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{2L} \right) A_n$$

$$Num22 = \left(1 + \frac{n \varepsilon_{si} \pi t_{ox1}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{2L} \right) B_n$$

$$Den11 = \left(1 - \frac{n \pi \varepsilon_{si} t_{ox1}}{\varepsilon_{ox} L} \right) \left(1 - \frac{n \pi \varepsilon_{si} t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{L} \right)$$

$$Den22 = \left(1 + \frac{n \pi \varepsilon_{si} t_{ox1}}{\varepsilon_{ox} L} \right) \left(1 + \frac{n \pi \varepsilon_{si} t_{ox2}}{\varepsilon_{ox} L} \right) \cos\left(\frac{n \pi t_b}{L} \right) \quad (11)$$

where

$$A_n = \frac{2qN_bL^2[\cos(n\pi)-1]}{\varepsilon_{si}n^3\pi^3} + 2V_{ds}\frac{\cos(n\pi)}{n\pi} - 2(V_{gs1}-V_{fb1}-V_{bi})\frac{\cos(n\pi)-1}{n\pi}$$

$$B_n = \frac{2qN_bL^2[\cos(n\pi)-1]}{\varepsilon_{si}n^3\pi^3} + 2V_{ds}\frac{\cos(n\pi)}{n\pi} - 2(V_{gs2}-V_{fb2}-V_{bi})\frac{\cos(n\pi)-1}{n\pi}$$

(12)

K_n and R_n are decided by the material parameters such as ε_{si} and ε_{ox} and design parameters such as L, N_b, t_{ox1}, t_{ox2}, t_b and also influence of biases such as V_{ds} and $V_{gs1}-V_{fb1}$, $V_{gs2}-V_{fb2}$. For the solution of $\psi_{II}(x,y)$ till third-order terms taken to get higher accuracies; once $\psi_I(y)$, $\psi_{II}(x,y)$ are calculated, the final expression for potential in channel for ADG JT becomes

$$\psi(x,y) = \psi_I(y) + \psi_{II}(x,y)$$

(13)

3 Model Verification and Discussions

To validate the analytic model for surface potential, we had considered the 3D device as shown in Fig. 1. To include the dependence on the impurity concentrations as well as the transverse and longitudinal electric field values, Lombardi mobility model is employed. For leakage currents issue, the Shockley–Read–Hall (SRH) recombination model is included in the simulation. Fermi–Dirac carrier statistics without impact ionization is utilized in the simulations, but Quantum effect is not considered. Doping concentration in channel N_D of 5×10^{19} and 1×10^{20} cm^{-3}, silicon body thickness (T_b)= 10 nm are considered for TCAD simulation. Channel width (W) is 10 nm. In addition, p-type polysilicon is used having doping concentration 10^{22} cm^{-3}.

A comparative surface potential variation along the channel direction for SDG and ADG JLTs is shown in Fig. 2. Here, for SDG, the oxide thicknesses (t_{ox}) are kept same at 1 nm for both top and bottom gates, while for ADG JLT, oxide thicknesses for top gate were kept at similar to SDG JLT, i.e., 1 nm while that of bottom gate kept at 2 nm, the channel lengths for both the transistors were kept at 20 nm. From Fig. 2, we can observe that the characteristic curve for the potential variation is same for both top and bottom surfaces for SDG JLT and also the top surface where we had kept t_{ox} same (1 nm). The bottom surface for ADG JLT shows a lower potential curve which is evident from higher t_{ox} (2 nm) which leads to lower capacitive action from bottom gate and thus low surface potential compared to the top gate. Moreover, the variations in surface potential curves occur mainly due to two reasons. First, when the device is not in use, nonzero oxide field is present in the silicon dioxide layer. Second, the potential difference between the gate and the source and the potential difference between gate and the drain are different. These two factors contribute to the bending
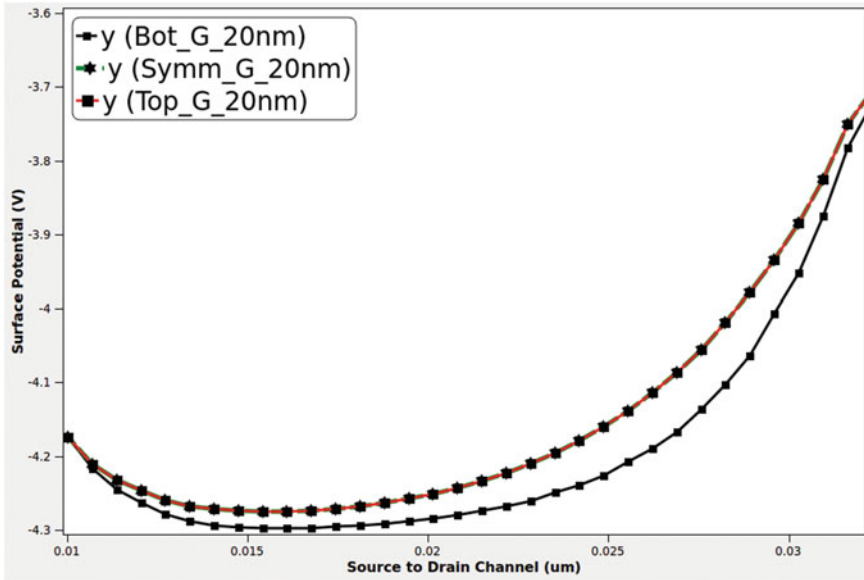
Fig. 2 Potential variation along Y-direction of the SDG/ADG device

of the potential curve between the source and the drain. With the decrease in the length of the device, the band bending increases which is shown in Fig. 3. Potential variations from source to drain for different channel lengths (20, 10 nm) and similar oxide thicknesses (top 1 nm, back 2 nm) of the ADG JLT are plotted in Fig. 3. From Fig. 3, we can observe that the band bending for channel length ($L_g = 20$ nm) shown is lesser than that in case for channel length ($L_g = 10$ nm). In case of the device ($L_g = 10$ nm) compared to that of device with ($L_g = 20$ nm), the source and the device are in greater proximity [2]. To observe the severity on surface potential with change in oxide thickness, we have considered different thicknesses of the oxide layers. For first case, we had taken t_{ox} top as 1 nm and that of bottom as 2 nm, while for the second case we had correspondingly taken 1 nm and 5 nm, respectively. From Fig. 4, it was evident that for top surface potential, both the cases were almost similar but that of bottom surface potential changes drastically as we go for 2–5 nm. Figure 5 shows the variations of surface potential of ADG JLT for different bias voltages (V_{GS}) with channel length 20 nm and t_{ox} front 1 nm and bottom 2 nm. This curve shows that with increase in gate potential, the differences in top and bottom surface potential bending are reduced. This can be explained by the fact that with increase in gate potential more number of charge carriers will be available in the channel; moreover, in the bottom surface, the t_{ox} is more, so lesser will be surface potential. This difference gets reduced correspondingly as we increase the gate voltages.
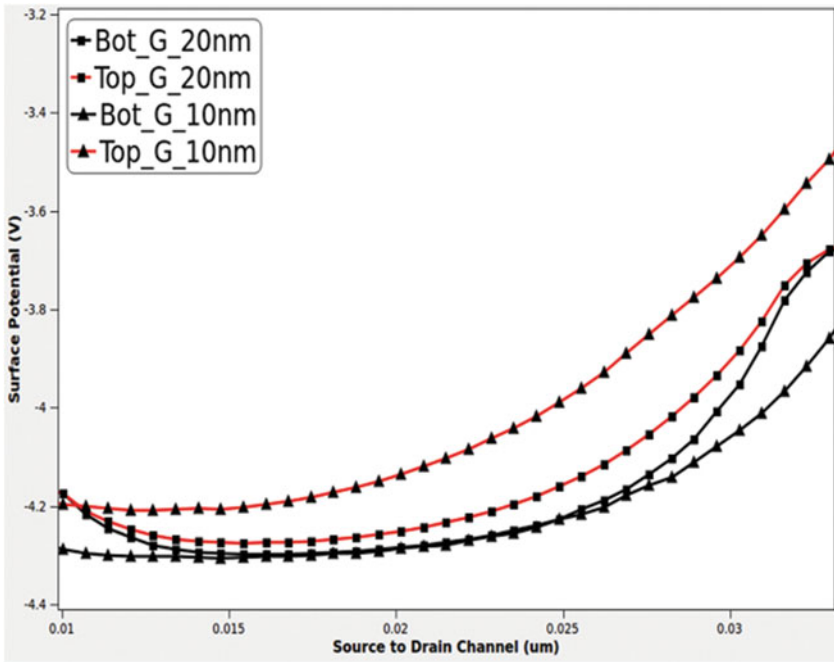
Fig. 3 Potential variation along Y-direction for different L for ADG JLT
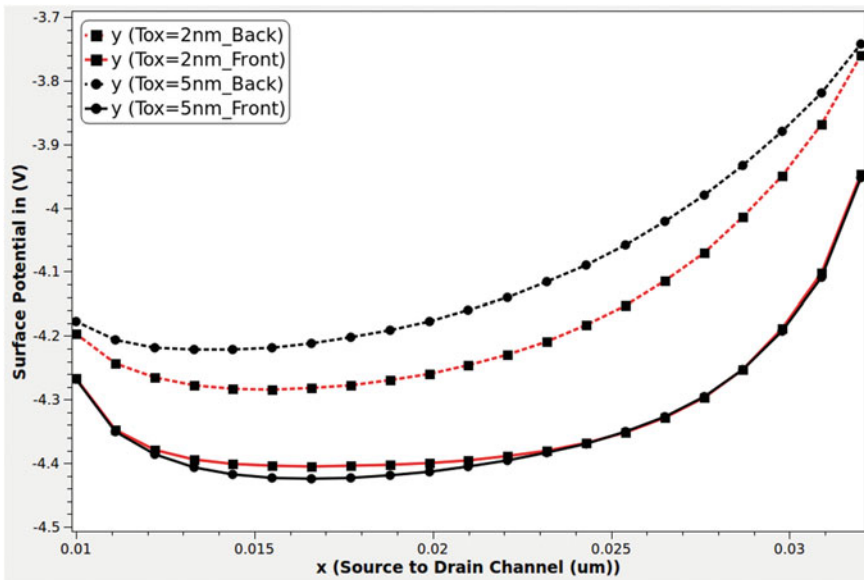
Fig. 4 Surface potential for different t_{ox} for a 20 nm ADG JLT device
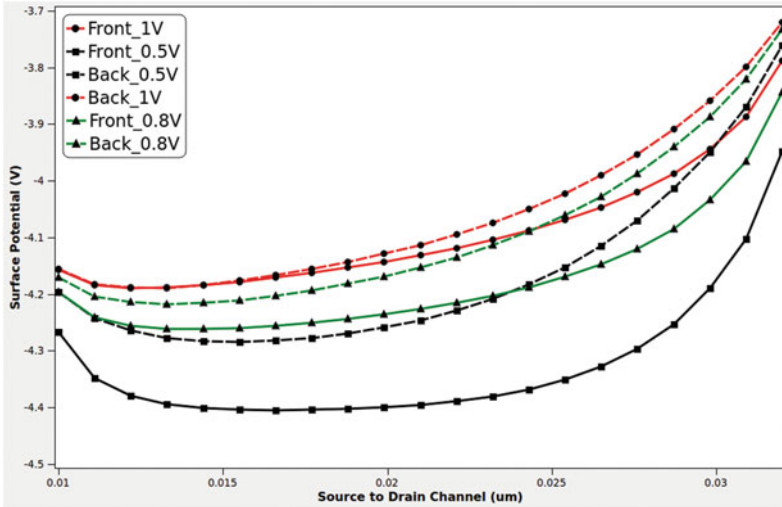
Fig. 5 Potential variation along Y-direction for different V_{gs} for ADG JLT

4   Conclusion

Here, we have proposed an analytical model of surface potential for Double-Gate Asymmetric Junctionless Transistor (DG AJLT). The model can accurately predict the surface potential including the flatband asymmetry and gate-oxide-thickness asymmetry. The variations of electrical characteristics because of structural asymmetry like the differences of gate oxide thicknesses and gate biases between the top gate and bottom gate oxide can be explained. Further, the models predict the variation of design parameters like body thickness and channel length variations with high accuracy. The models proposed show a very good agreement with the results obtained from 3D TCAD device simulation.

References

1. Colinge J-P, Lee C-W, Afzalian A, Akhavan ND, Murphy R (2010) Nanowire transistors without junctions. Nature Nanotech 5:225–229
2. Lee C-W, Afzalian A, Akhavan ND, Yan R, Ferain I, Colinge J-P (2009) Junctionless multigate field effect transistor. Appl Phys Lett 94:0535111–0535112
3. Duarte JP, Choi S-J, Choi Y-K (2011) A full-range drain current model for double gate junctionless transistors. IEEE Trans Electron Devices 58(12):4219–4225
4. Lin Z-M, Lin H-C, Liu K-M, Huang T-Y (2012) Analytical model of subthreshold current and threshold voltage for fully depleted double-gated junctionless transistor. Jpn J Appl Phys 51(2S):BC14-7
5. Bora N, Das P, Subadar R (2016) An analytical universal model for symmetric double gate junction-less transistors. J Nano Electron Phys Ukraine 8(2):02003–02007

6. Lu H, Taur Y (2006) An analytic potential model for symmetric and asymmetric DG MOSFETs. IEEE Trans Electron Devices 53(5):1161–1168
7. Jin Xiaoshi, Liu Xi, Kwon Hyuck-In, Lee Jung-Hee, Lee Jong-Ho (2013) A subthreshold current model for nanoscale short channel junctionless MOSFETs applicable to symmetric and asymmetric double-gate structur. Solid-State Electron 82(5):77–81
8. Baruah RK, Paily RP (2016) A surface-potential based drain current model for short-channel symmetric double-gate junctionless transistor. J Comput Electron 15:45–52

Analytical Modeling of Surface Potential for Double-Gate MOSFET

Jitendra Prasad, Amit Agarwal, P. C. Pradhan and B. P. Swain

Abstract In this paper, an analytical surface potential of a double-gate MOSFET is modeled using the two-dimensional Poisson equation, and the impact of gate voltage (V_{GS}) and drain voltage (V_{DS}) on the surface potential is studied. The physical dimensions of the bulk MOSFETs have been aggressively scaled down to its limit and are not showing any improvement in device performance on further scaling. So in order to enhance the performance of the device, we need a new architecture. This new architecture is called double-gate MOSFET or DG MOSFET and it is a promising candidate for 40 nm technology nodes. DG MOSFET provides better control of the channel inversion by applying two gates across the channel.

Keywords MOSFETs · DG-FET · Short-channel effects (SCEs) · Surface potential

1 Introduction

The process of scaling in the semiconductor industry has gained much attention in the last few decades [1]. The main advantage of scaling is that reduction in the size of the transistors increases the speed and reduces the cost [2]. When we make the circuits smaller, the capacitance of the circuit reduces, thereby increasing the operating speed.

J. Prasad · A. Agarwal (✉) · P. C. Pradhan
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India
e-mail: amiteng2007@gmail.com

J. Prasad
e-mail: prasadjitendra95@gmail.com

B. P. Swain
Centre for Material Science and Nanotechnology, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India

The capacitance is the ability of the device to store the electric charge. The product of resistance and capacitance known as time constant characterizes the rate of charging and discharging of a capacitor, and when this time constant is smaller the charging and discharging rate of the capacitor increases and vice versa. However, with great reduction comes great problem known as Short-Channel Effects (SCEs) [3]. These short-channel effects have adverse effect on the performance of the device. These short-channel effects occur when the channel of the MOSFET becomes same order as the depletion layer width of the source and drain. When these source and drain depletion regions come close to each other and start interacting with each other, the transistor starts behaving differently, which impacts the performance, modeling, and reliability of the device and degrades the performance of the device. Also, due to this short-channel effect, the gate controllability of the channel is reduced [4].

The gate controllability over the channel can be increased by adding extra gates in the device [1]. Buvaneswari et al. studied that this additional gate would help to strengthen the immunity of the channel and provide benefits such as high-speed and low-power configuration to the device [4].

2 Double-Gate MOSFET

Double-gate transistor is the first to the multigate transistor family. MOSFET containing two different gates placed on the opposite side of the body such that a gate-oxide-body-oxide-gate stack is formed called double-gate MOSFET [5]. The advantage of a double-gate MOSFET is that it has two gates and efficiently controls the channel with two gates placed on the opposite side of the channel. Controlling the channel by multiple gates has advantage of better control over the channel inversion; so, the Short-Channel Effect (SCE) is reduced [6].

The structure of a DG MOSFET is based on Silicon on Insulator (SOI) technology [7]. Two gates are systematically used to control the electrostatic coupling, so the amount of current flow in the channel is properly modulated by electric field. The control from the back gate enables higher transconductance and minimizes the SCE [8]. The voltage applied on the gate terminals controls the electric field and determines the amount of current flow through the channel.

Basically, there are two types of DG MOSFET: symmetrical and asymmetrical [9, 10]. Symmetrical double-gate MOSFET has same gate electrode material and the same voltage is applied at both gates, whereas asymmetrical DG MOSFET is made up of different gate electrode materials.

The schematic diagram of a DG MOSFET is shown in the below figure, where t_{ox}, t_{si}, and L are the thickness of gate oxide, thickness of silicon, and length of the channel (Fig. 1).

Fig. 1 Double-gate
MOSFET [13]
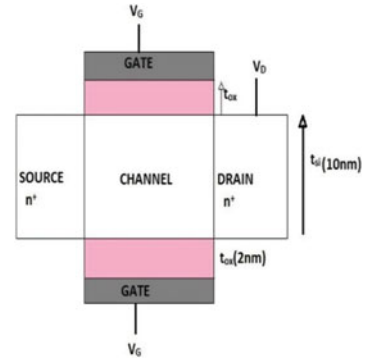
3 Derivation of Surface Potential of Double-Gate MOSFET

The Poisson's equation of potential ψ is given by

$$\frac{d^2\psi(x, y)}{dx^2} + \frac{d^2\psi(x, y)}{dy^2} = \frac{qN_a}{\epsilon_{si}} \tag{1}$$

Using Young's parabolic potential distribution [11] along the vertical direction, we get

$$\psi(x, y) = a(x) + b(x)y + c(x)y^2 \tag{2}$$

where $a(x)$, $b(x)$, and $c(x)$ are coefficients and determined using the boundary conditions, and continuity of electric flux at the Si–SiO$_2$ interfaces as follows:

$$\psi(x, t_{si}) = \psi_b(x) = a(x) + b(x)t_{si} + c(x)t^2 \tag{3}$$

$$\frac{\partial\psi(x, y)}{\partial y} = \frac{\epsilon_{ox}}{\epsilon_{si}}\frac{(\psi_f(x) - (V_{GS} - V_{FB}))}{t_{ox}}$$

$$\frac{\partial\psi(x, y)}{\partial y} = b(x) \quad \text{for } y = 0 \tag{4}$$

$$\frac{\partial\psi(x, y)}{\partial y} = \frac{\epsilon_{ox}}{\epsilon_{si}t_{si}}\frac{((V_{GS} - V_{FB}) - \psi_b(x))}{t_{ox}}$$

$$\frac{\partial \psi(x, y)}{\partial y} = b(x) + 2t_{si}c(x) \text{ for } y = t_{si} \tag{5}$$

where V_{GS} is the applied gate voltage and V_{FB} is the flat band voltage, whereas $\psi_f(x)$ and $\psi_b(x)$ are front and back surface potentials, respectively. Using the symmetry condition that is $\psi_f(x) = \psi_b(x)$ and solving Eqs. (4) and (5), we get

$$c(x) = \frac{\epsilon_{ox}}{\epsilon_{si}} \frac{((V_{GS} - V_{FB}) - \psi b(x))}{t_{ox}} \tag{6}$$

Substituting the value of a(x), b(x), and c(x) in Eq. 2, we get the value of potential as follows [12]:

$$\psi(x, y) = \psi_f(x) + \frac{\epsilon_{ox}}{\epsilon_{si}} \frac{(\psi_f(x) - (V_{GS} - V_{FB}))}{t_{ox}} y + \frac{\epsilon_{ox}}{\epsilon_{si}} \frac{((V_{GS} - V_{FB}) - \psi b(x))}{t_{ox}} y^2 \tag{7}$$

where ϵ_{ox} and ϵ_{si} are the permittivity of gate oxide and silicon, and $\psi_f(x)$ and $\psi_b(x)$ are the front and back surface potentials [14–18].

4 Results and Discussions

In this section, we are going to discuss the results which are obtained from the theoretical model of double-gate MOSFET and compared the surface potential value with channel length for different varying parameters like gate oxide, silicon thickness, drain-to-source voltage, and drain-to-gate voltage.

Figure 2 show the variation of surface potential along the channel length for various V_{GS}. It is observed from the figure that as gate voltage increases, the surface potential along the source and drain side increases; therefore, surface potential increases in the channel region and the immunity to manage the SCEs is enhanced.

From Fig. 3, we can conclude that when positive drain voltage is applied, it shows the parabolic curve; at 0 V, the curve is flat, whereas on application of negative drain voltage, the parabolic curve of surface potential is inverted.

Figure 4 shows the variation of surface potential along the channel length for different values of the oxide thickness whose value ranges from 10 to 1 nm. The scaling of the oxide thickness increases the gate control over the channel region, but the oxide thickness value should not be scaled down to a very small value because if oxide thickness value is very small the tunneling of electrons through the oxide increases which further increases the SCEs, and hot carriers effect becomes prominent.

Figure 5 shows the variation of surface potential along the channel length for various silicon thicknesses that is ranging from 10 to 5 nm. It may be observed from the figure that as the value of silicon thickness is increased, the curve is parabolic in nature.

Fig. 2 Surface potential
versus channel length for
different V_{GS}
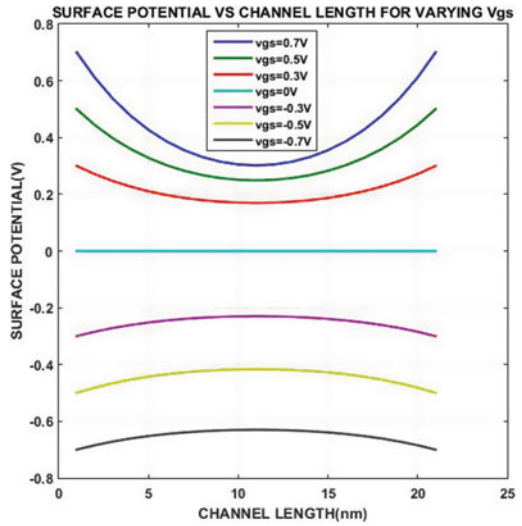
Fig. 3 Surface potential
versus channel length for
different V_{DS}
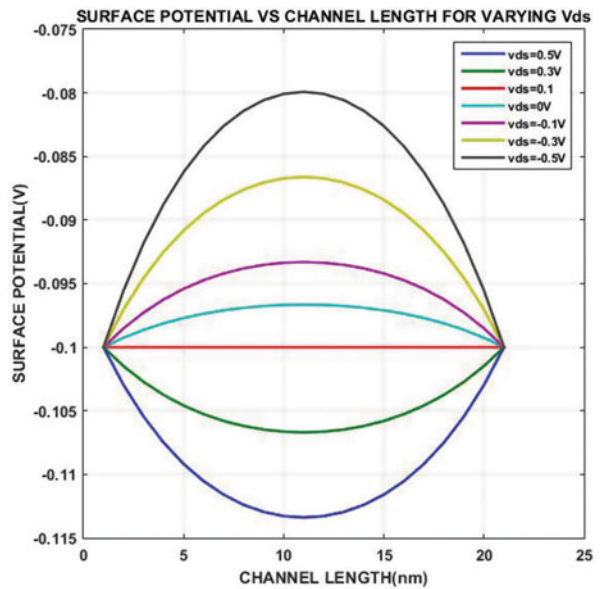
Fig. 4 Surface potential versus channel length for different t_{OX}
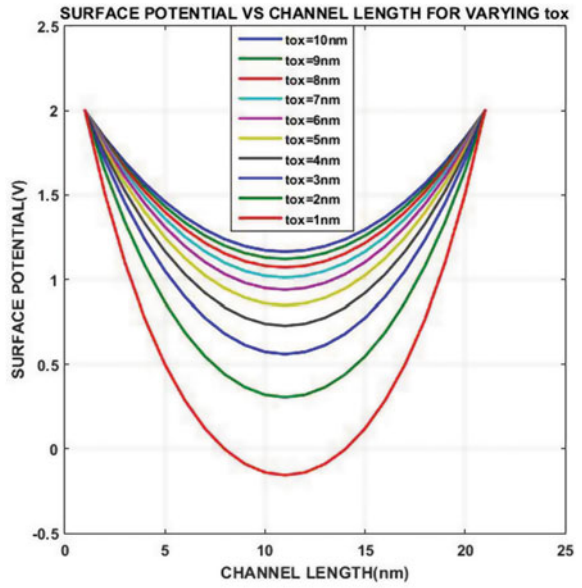
Fig. 5 Surface potential versus channel length for different t_{Si}
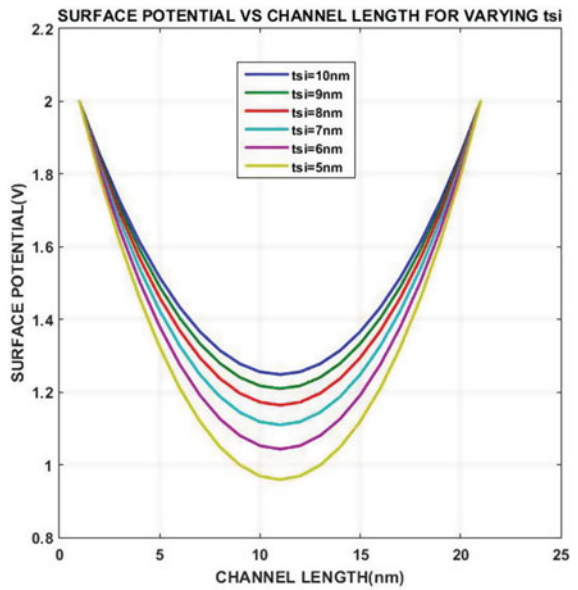
5 Conclusion

The operation of the double-gate MOSFET brings many advantages such as further scalability of device, reduced short-channel effects, high current drive, and increase in threshold voltage. DG MOSFET is the perfect candidate for replacing planar bulk MOSFET which suffers from severe short-channel effects such as drain-induced barrier lowering (DIB), impact ionization, etc. Also, in this paper, we studied the analytical surface potential model of DG MOSFET using Young's parabolic approximation of the channel. The variation of surface potential along the channel length for different parameters such as gate oxide, gate voltage, and drain voltage is studied and its effect is studied.

References

1. Auth CP, Plummer JD (1997) Scaling theory for cylindrical, fully depleted surrounding gate MOSFET's. IEEE Electron Device Lett 18(2):74–76
2. Ortiz-Conde A, Garcia-Sanchez FJ, Muci J, Malobabic S, Liou JJ (2007) Review of core compact models for undoped double-gate SOI MOSFETs. IEEE Trans Electron Devices 54(1)
3. Young KK (1989) Short-channel effect in fully depleted SOI MOSFETs. IEEE Trans Electron Devices 36(2):399–402
4. Oh S-H, Monro D, Hergenrother JM (2000) Analytic description of short-channel effects in fully-depleted double-gate and cylindrical surrounding-gate MOSFETs. IEEE Electron Device Lett 21(9)
5. Buvaneswari B (2014) A survey on multi gate MOSFETS. Int J Innov Res Sci Eng Technol 3(3):783–788
6. Bansal J, Sharma N, Kumar SP, Chaujar R, Gupta M, Gupta RS (2008) Two-dimensional analytical sub-threshold model of double gate MOSFET with gate stack. In: Proceedings of international conference on microwave
7. El Hamid HA, Guitart JR, Iniguez B (2007) Two-dimensional analytical threshold voltage and subthreshold swing models of undoped symmetric double-gate MOSFETs. IEEE Trans Electron Devices 54(6)
8. Suzuki K, Tosaka Y, Sugii T (1996) Analytical threshold model for short channel double-gate SOI MOSFET's. IEEE Trans Electron Devices 43:1166–1168
9. Su P, Fung SKH, Tang S, Assaderaghi F, Hu C (2000) Partial-Depletion SO1 MOSFET model for deep-submicron CMOS designs. In: IEEE custom integrated circuits conference
10. Nandi A, Saxena AK, Dasgupta S (2013) Analytical modeling of a double gate MOSFET considering source/drain lateral gaussian doping profile. IEEE Trans Electron Devices 60(11)
11. Reddy GV, Kumar MJ (2005) A new dual material double gate (DMDG) nanoscale SOI MOSFET: two dimensional analytical modeling and simulation. IEEE Trans Electron Devices 4(2):260–268
12. Colinge JP (2004) Multiple-gate SOI MOSFETs. Solid State Electron 48(6):897–905
13. Vaddi R, Agarwal R, Dasgupta S (2012) Compact modelling of a generic double gate MOSFET with gate S/D underlap. IEEE Trans Electron Devices 59(10):2846–2849
14. Widiez J, Lolivier J, Vinet M, Poiroux T, Previtali B, Dauge F (2005) Experimental evaluation of gate architecture influence on DG SOI MOSFETs performance. IEEE Trans Electron Devices 52(8)
15. Rahman A, Lundstrom MS (2002) A compact scattering model for the nanoscale double-gate MOSFET. IEEE Trans Electron Devices 49(3):481–489

16. Chen Q, Harrell EM, Meindl JD (2003) A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs. Electron Devices 50(7):1631–1637
17. Park JT, Colinge JP (2002) Multiple-gate SOI MOSFETs: device design guidelines. IEEE Trans Electron Devices 49(12):2222–2228
18. Lu H, Taur Y (2006) An analytic potential model for symmetric and asymmetric DG MOSFETs. IEEE Trans Electron Devices 5:1161–1168

Grid Impedance Measurement in Low-Voltage Network

Gaurav Trivedi, Rahul Sharma and Awnish Kumar Tripathi

Abstract The performance of the power system can be affected by the variation in grid impedance. Hence, continuous monitoring of the grid impedance is very essential. This paper presents a method for the measurement of grid impedance variation in low-voltage network. The estimation method is computationally very simple, and it is dependent on the variance in voltage of the grid at two successive time sampling instants. A quadratic equation is formed using grid voltage. The proposed method is modeled in MATLAB Simulink to measure the grid impedance. Simulation results obtained in MATLAB Simulink are showing or demonstrating the efficacy of the measurement technique for discerning the impedance in low-voltage network.

Keywords Impedance measurement · Grid voltage · Point of common coupling · Harmonic distortion · Quadratic equation

1 Introduction

It is very important to know about the impedance of the network at the time of controlling and modeling of the system. Without knowing the structure of the power system network and the impedance values that is alone used for the arrangement, it is inconceivable to mitigate the harmonic components of the network. This information is necessary for maintaining power quality under variation in load to a network or upgradation of the system. If the line impedance of the network is known, then it is

G. Trivedi (✉) · A. K. Tripathi
Department of Physics, National Institute of Technology Kurukshetra, Kurukshetra 136119, India
e-mail: gauravtrivedi687@gmail.com

A. K. Tripathi
e-mail: awnish1982@gmail.com

R. Sharma
Department of Electrical Engineering, National Institute of Technology Kurukshetra, Kurukshetra 136119, India
e-mail: rahulsharmaknit.2006@gmail.com

possible to reduce the voltage disturbance in a network by rearranging the sensitive loads at the place of low harmonic distortion [1]. Hence, the line impedance of the network, in particular, is of the interest.

Transmission line impedance variation is very common, and it happens because of various reasons such as environmental abnormalities, transmission cables of very long distance, or the harmonics associated with the variable loads connected to the grid, such as distributed generation systems. The discrepancy in supply impedance too influences the ripples on the measured voltage at the point of common coupling. The control performance of the system connected to the point of common coupling can be affected by a mismatched impedance value alongside a distorted PCC voltage. Therefore, it is very important that the disparity in supply impedance should be considered for the proper control implementation. Also, it would be significantly advantageous, if a better online method for the estimation of the supply impedance Z_s is provided. It is used to keep posted the overall value of the impedance on the ac side for the control scheme. The line impedance measurement is widely needed for proper decoupling between the active and or/ reactive power, accurate power distribution, and proper grid synchronization and to improve the system stability. The existing line impedance estimation methods are completely based on harmonic or impulse injections and can measure impedance value mostly but the existing methods are very complex and add high computational burden on the inverter control algorithm. However, both inductive and resistive parts of the line impedance cannot be estimated separately and/or accurately.

There are a number of methods suggested in the literature with both offline and online implementations for the assessment of the supply impedance. The grid impedance measurement technique proposed in [2] is a steady-state technique in which a non-characteristic harmonic current is injected into the power plant network and the measurement of the change in voltage response is done. Fourier analysis of the particular injected harmonic current is done further for the processing of the results. In [3], fast Fourier transform is being applied to estimate the distinct harmonic impedances. A voltage transient is injected at the point of common coupling which results in a transient deviation in the voltage and current values at the point of common coupling. Nevertheless, a high computational effort is requisite for the application of an online FFT. Also, there are few more methods for the evaluation of grid impedance which requires variation of real and reactive powers [4]. In this method, some form of disruption is inserted onto the energized network for the measurement. Response to this inserted signal is used to estimate the impedance value. The impedance is estimated in [5] by inserting a small amount of sinusoidal current signal at various frequencies and then evaluating the phase and amplitude of the voltage and current at the insertion point. Another active method as suggested in [6] involves the insertion of a small current spike into the network. It is important that the current spike signal should have adequate spectral content for the suitable frequency range. Inserted current as well as voltage response transients is being measured for the estimation of the impedance in the frequency domain. Virtual-flux-based control method was also proposed to estimate impedance in [7]. In [8], the author has proposed a technique in which the grid impedance is calculated with the help of existed

PV-inverter sensors and logic control. An analytical approach for the calculation of the coupling impedance in DPC of active rectifiers is also suggested in [9, 10].

This work proposes the simplified mathematical approach to estimate the line resistance and inductance separately and/or accurately. To estimate the line impedance, the proposed method required only local information and the designed mathematical equations. Any harmonic injection and complex computational burden are not required. The proposed method is implemented on the low-voltage islanded microgrid network. The proposed approach works on the principle of assumption that the magnitude of grid voltage does not change between two consecutive sampling time instants [10, 11]. The simulation results of the estimation compared to the actual line impedance are shown in different conditions to validate the proposed mathematical approach.

2   Proposed Method

In this paper, the total inductance is estimated by the given proposed method which is fed to the model-based predictive controller. By considering the value of total estimated inductance, the deviation in the supply inductance L_s can be easily observed and we can also calculate the grid voltage using the estimated value inside the controller (Figs. 1 and 2).

Herong et al. [12] shows that rise in frequency value leads to the decrement in the magnitude of voltage harmonics, as a result of which high-frequency harmonics has restricted effects on the grid impedance. In addition to this, the high-frequency values of grid impedance do not show any appreciable effects on the predictive control. Hence, a low-frequency grid impedance model is adopted here. The esti-
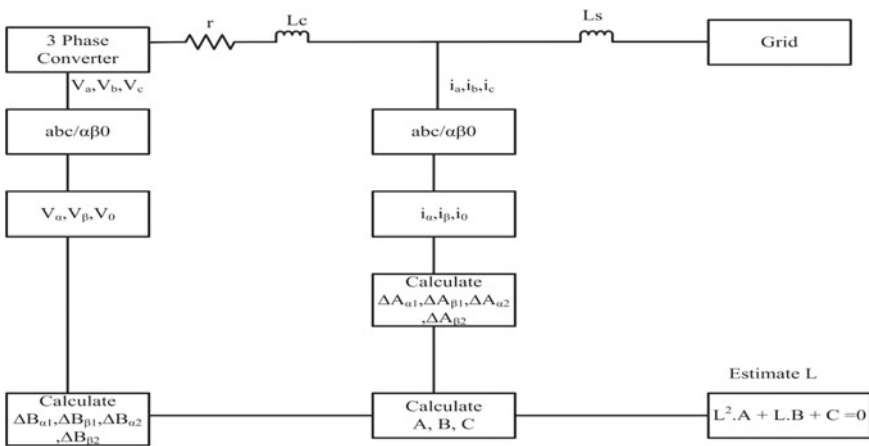
Fig. 1 Block diagram of Simulink model
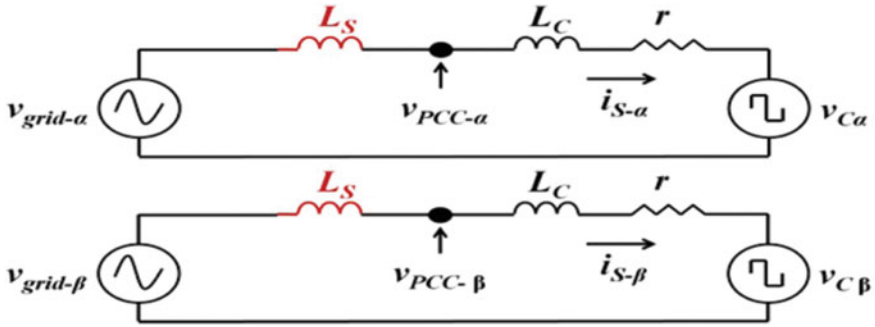
Fig. 2 α and β reference frames model
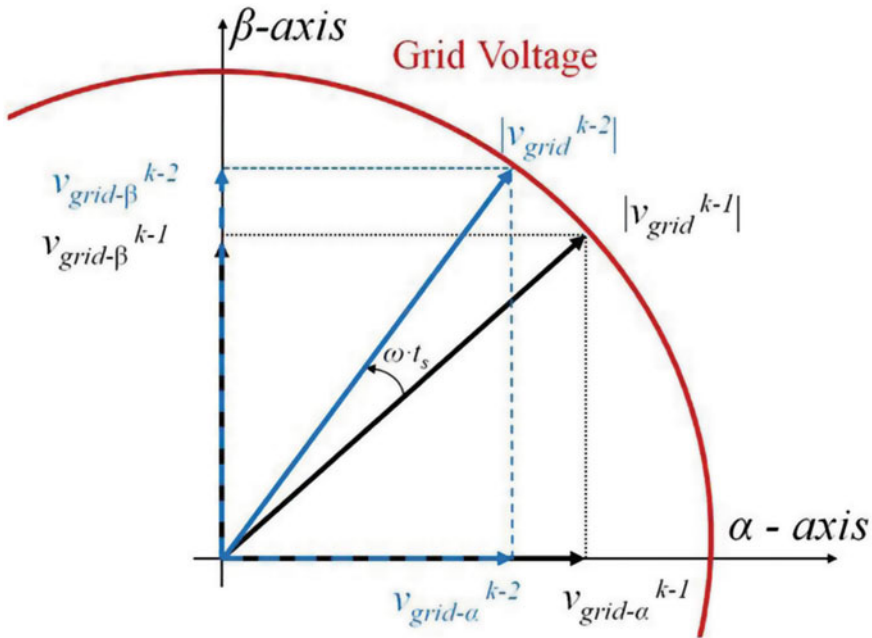
Fig. 3 Representation of grid voltage in $\alpha - \beta$ reference frame

mation method is based on the assumption that the magnitude of grid voltage does not change between two consecutive sampling time but a phase shift of "$\omega \cdot t_s$" is observed. However, the grid voltage vector will not change significantly which can be seen in Fig. 3.

In $\alpha - \beta$ frame of reference, the square of the magnitude of GV at any instant k can be represented as

$$\left|v_{grid}^k\right|^2 = \left(v_{grid-\alpha}^k\right)^2 + \left(v_{grid-\beta}^k\right)^2 \tag{1}$$

From Figs. 1 and 2, the magnitude of GV at the time k is expressed taking the value of v_{grid}^k obtained from the system model in $\alpha - \beta$ reference frame

$$\left(v_{grid-x}^k\right)^2 = \left(L \cdot \frac{di_{s-x}}{dt} + r \cdot i_{s-x}^k + v_{c-x}^k\right)^2 \tag{2}$$

Hence, the square of the magnitude of GV at time k can be represented as

$$\left|v_{grid}^k\right|^2 = \left(L \cdot \frac{di_{s\alpha}}{dt} + r \cdot i_{s\alpha}^k + v_{c\alpha}^k\right)^2 + \left(L \cdot \frac{di_{s\beta}}{dt} + r \cdot i_{s\beta}^k + v_{c\beta}^k\right)^2 \tag{3}$$

Similarly, the square of the magnitude of GV at the previous time instant $k - 1$ can be represented as

$$\left|v_{grid}^{k-1}\right|^2 = \left(L \cdot \frac{di_{s\alpha}}{dt} + r \cdot i_{s\alpha}^{k-1} + v_{c\alpha}^{k-1}\right)^2 + \left(L \cdot \frac{di_{s\beta}}{dt} + r \cdot i_{s\beta}^{k-1} + v_{c\beta}^{k-1}\right)^2 \tag{4}$$

$v_{c-\alpha}$ and $v_{c-\beta}$ in the above-mentioned equations are extracted from [11]. The total inductance L is calculated by equating Eqs. (3) and (4)

$$\left|v_{grid}^k\right|^2 - \left|v_{grid}^{k-1}\right|^2 = 0 \tag{5}$$

On solving Eq. (5), we can develop a quadratic Eq. (6), with the parameters *A, B, and C.*

$$L^2 \cdot A + L \cdot B + C = 0 \tag{6}$$

A, B, and *C* can be expressed as below:

$$A = (\Delta A_{\alpha 1})^2 + \left(\Delta A_{\beta 1}\right)^2 - (\Delta A_{\alpha 2})^2 - \left(\Delta A_{\beta 2}\right)^2 \tag{7}$$

$$B = 2 \cdot \left(\Delta B_{\alpha 1} \cdot \Delta A_{\alpha 1} + \Delta B_{\beta 1} \cdot \Delta A_{\beta 1} - \Delta B_{\alpha 2} \cdot \Delta A_{\alpha 2} - \Delta B_{\beta 1} \cdot \Delta A_{\beta 2}\right) \tag{8}$$

$$C = (\Delta B_{\alpha 1}) + \left(\Delta B_{\beta 1}\right)^2 - (\Delta B_{\alpha 2})^2 - \left(\Delta B_{\beta 2}\right)^2 \tag{9}$$

where

$$\Delta A_{\alpha 1} = \frac{i_{s\alpha}^{k+1} - i_{s\alpha}^k}{Ts}, \Delta A_{\beta 1} = \frac{i_{s\beta}^{k+1} - i_{s\beta}^k}{Ts}$$

$$\Delta A_{\alpha 2} = \frac{i_{s\alpha}^k - i_{s\alpha}^{k-1}}{Ts}, \Delta A_{\beta 2} = \frac{i_{s\beta}^k - i_{s\beta}^{k-1}}{Ts}$$

Table 1 Simulation parameters

Grid line voltage (V, rms)	270
Frequency of grid (Hz)	50
Rating of grid (KVA)	500
Inductance, L_s (mH)	0.45
Filter capacitance, C_f (uF)	170

$$\Delta B_{\alpha 1} = r i_{s\alpha}^k + \upsilon_{c\alpha}^k, \ \Delta B_{\beta 1} = r i_{s\beta}^k + \upsilon_{c\beta}^k$$

$$\Delta B_{\alpha 2} = r i_{s\alpha}^{k-1} + \upsilon_{c\alpha}^{k-1}, \ \Delta B_{\beta 2} = r i_{s\beta}^{k-1} + \upsilon_{c\beta}^{k-1}$$

After substituting (7)–(9) in (6), the total inductance is evaluated as

$$L_{estimation} = \frac{1}{2} \cdot \frac{B}{A} \cdot \left[-1 + \sqrt{1 - \frac{4 \cdot C \cdot A}{B^2}} \right]$$

3   Results and Discussion

The system is modeled in MATLAB Simulink to validate the algorithm. The system data are shown in the Table 1.

Figure 4 represents the waveform of the power flow through the filter inductance and line inductance. The power waveform with impedance estimation shows lesser ripples and better dynamics as compared to the waveform without impedance estimation.

The current waveform also clearly reflects the improvement in power quality with the impedance estimation as compared to the current waveform without impedance estimation. This shows the effects of the line impedance (Fig. 5).

Similarly, voltage waveform shows the better result in terms of dynamics and ripple content with impedance estimation algorithm as compared to without estimation (Fig. 6).
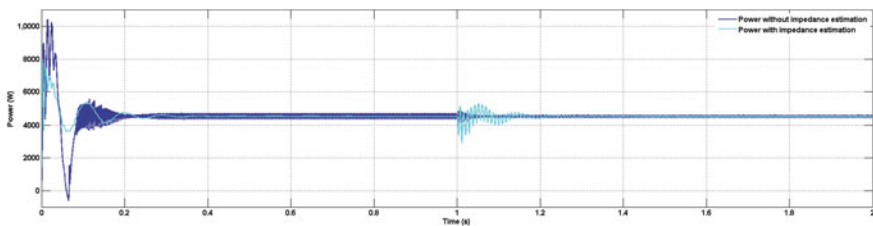
Fig. 4 Waveform of the power flow
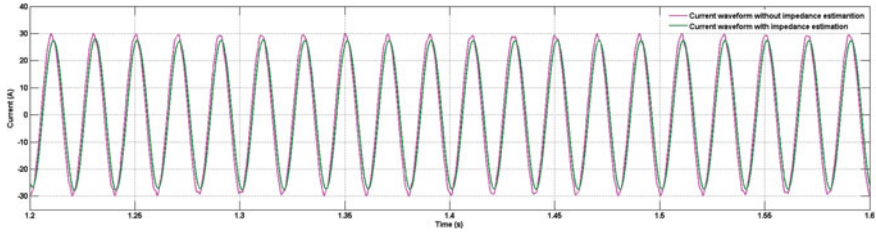
Fig. 5 Waveform of current
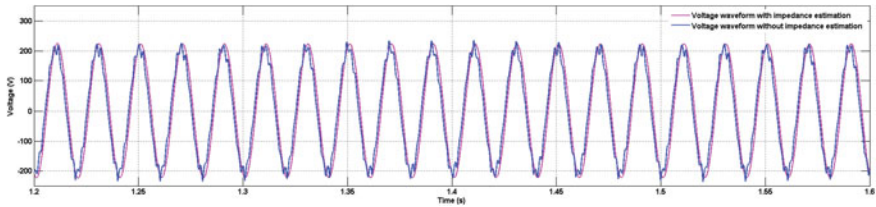
Fig. 6 Waveform of voltage
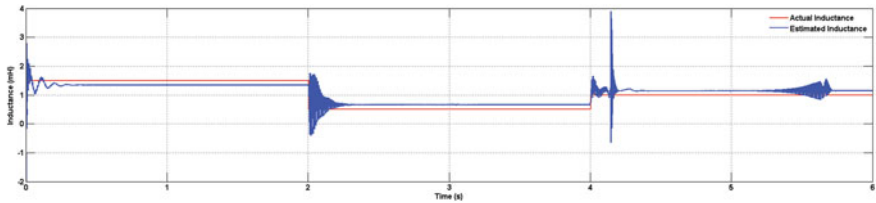
Fig. 7 Waveforms of actual and estimated inductance

Figure 7 shows the waveforms of the line inductance in actual and estimated with algorithm. The estimated algorithm is closely tracking the actual line inductance value. This shows the usefulness of the impedance estimation method.

4 Conclusions

The proposed method is designed to estimate the impedance of line in the network. The results have validated the usefulness of the estimation method. The waveforms have shown the improved power quality with the estimation method against the without estimation method results. The line impedances also affect the control method in the weak grid and microgrid connected system as concluded by the result analysis. Therefore, the line impedance estimation is a must for proper control and improves power quality of the low-voltage network. This proposed method may be applied in the distributed energy systems to improve the control schemes such as microgrid,

weak grid, etc. Further, the dynamic and transient response of the line impedance estimation method may be improved as the scope of the future study.

References

1. Dugan RC, McGranaghan MF, Beaty HW (1996) Electrical power systems quality, 1st ed. McGraw Hill
2. Timbus A, Teodorescu R, Blaabjerg F, Borup U (2006) Online grid impedance measurement suitable for multiple PV inverters running in parallel. In: Proceedings of APEC'06, 19–23 March 2006, 5 pp
3. Sumner M, Palethorpe B, Thomas DWP, Zanchetta P, Piazza MC (2002) A technique for power supply harmonic impedance estimation using a controlled voltage disturbance. IEEE Trans Power Electron 17(2):207–215
4. Timbus AV, Rodriguez P, Teodorescu R, Ciobotaru M (2007) Line impedance estimation using active and reactive power variations. In: Proceedings of the IEEE power electronics specialists conference, Jun 2007, pp 1273–1279
5. Rhode JP, Kelley AW, Baran ME (1995) Line impedance measurement: a nondisruptive wide-band technique. In: Presented at IAS annual meeting, Orlando, Florida, USA
6. Palethro B, Summer M, Thomas D (2000) Power system impedance measurement using a power electronic converter. In: Proceedings of harmonics and quality of power, 2000, vol 1, Orlando, FL, pp 208–213
7. Antoniewicz P, Kazmeirkowski MP (2008) Virtual-flux-based predictive direct power control of AC/DC converters with online inductance estimation. IEEE Trans Ind Electron 55(12):4381–4390
8. Asiminoaei L, Teodorescu R, Blaabjerg F, Borup U (2004) A new method of on-line grid impedance estimation for PV inverters. In: Proceedings of APEC'04, vol 3, pp 1527–1533
9. Norniella JG et al (2011) Analytical and iterative algorithms for online estimation of coupling inductance in direct power control three-phase active rectifiers. IEEE Trans Power Electron 26(11):3298–3307
10. Norniella JG et al (2013) New strategies for estimating the coupling inductance in grid-connected direct power control-based three-phase active rectifiers. In: Proceedings of the IEEE PEC, July 2013, pp 1–5
11. Arif B, Tarisciotti L, Zanchetta P, Clare J, Degano M (2014) Integrated grid inductance estimation technique for finite set model predictive control in grid-connected converters. In: Proceedings of the IEEE ECCE, Sep 2014, pp 5797–5804
12. Herong G, Xiaoqiang G, Deyu W, Weiyang W (2012) Real-time grid impedance technique for grid-connected power converters. In: Proceedings of the IEEE ISIE, May 2012, pp 1621–1626

Fault Localization on the Transmission Line Using FDOST and RBFNN

Manini Sarkar, Abdul Rahim and Bikash Patel

Abstract This paper presents a fault localization technique based on fast discrete orthogonal Stockwell transform (FDOST) and radial basis function neural network (RBFNN) on the transmission line. A part of the transmission network of WBSETCL, West Bengal is designed and simulated in MATLAB Simulink for the fault investigation. The fault current signals are recorded at one end of the transmission line with a sampling frequency of 50 kHz, and FDOST energy is extracted as fault feature from each of the three fault current signals. These features are fed to the RBFNN for fault localization on the transmission line. The proposed algorithm is found accurate for different types of faults, fault resistances and fault inception angles (FIA) at different locations on the transmission line.

Keywords Fast discrete orthogonal Stockwell transform (FDOST) · FDOST energy · Radial basis function neural network (RBFNN) · Fault inception angle (FIA) · Faults

1 Introduction

The power system protection is very important aspect for uninterrupted quality power supply to the consumers. Fast and accurate detection and localization of faults on the transmission line which is the heart of the power system is very important to restore the power system into normal condition. Different techniques are found in the literature for fault localization on the transmission line. The phasor estimation technique

M. Sarkar · A. Rahim · B. Patel (✉)
Electrical Engineering Department, Kalyani Government Engineering College, Kalyani, West Bengal, India
e-mail: biks.ee@gmail.com

M. Sarkar
e-mail: manini.sarkar.ee@gmail.com

A. Rahim
e-mail: abdulrahim0790@gmail.com

using discrete Fourier transform [1] and fast discrete Stockwell transform [2] locates faults on the transmission line accurately, although the travelling wave method [3, 4] is faster than this technique. Wavelet transform has better time–frequency resolution than discrete Fourier transform and is applied for fault localization [5, 6]. The features extracted using signal processing tools are used to train the machine learning tools for the estimation of fault location. Machine learning techniques like support vector machine [4], Fuzzy logic, radial basis function neural network [7, 8], and back propagation neural network [5, 9] are popular for fault localization.

Wavelet packet decomposition (WPD), a generalized form of wavelet transform, has a better time–frequency resolution than the wavelet transform. The drawbacks of WPD are the dilemma of mother wavelet selection and poor noise immunity. The Stockwell transform (ST), which is free from these limitations, can be used as the signal processing tool for fault localization [9–11]. S-transform has high redundancy and computational complexity. The high redundancy of S-transform can be improved by N number of unit length orthogonal basis vector incorporated in discrete orthogonal S-transform (DOST) [11, 12].

The present article proposes a fault localization method using fast discrete orthogonal S-transform (FDOST) and radial basis function neural network (RBFNN). The FDOST coefficients have been extracted from the fault current signals to train radial basis function neural network (RBFNN) for fault localization. The proposed method is tested for different fault resistances and fault inception angles (FIA) at different locations on the transmission line.

2 Fast Discrete Orthogonal Stockwell Transform

The S-Transform (ST), proposed in 1996, represents a signal in the time–frequency domain with a frequency-dependent resolution which is suitable for non-stationary signal analysis.

$$S(\tau, f) = \int_{-\infty}^{+\infty} h(t) \frac{|f|}{\sqrt{2\pi}} \exp\left(-\frac{(\tau - t)^2}{2f^2}\right) \exp(-i2\pi f t) dt \tag{1}$$

where f is the frequency of the signal, and t and τ are the time variables. The discrete S-Transform (DST) can be represented as

$$S[j, n] = \sum_{m=0}^{N-1} H(m + n) \exp\left(-\frac{2\pi^2 m^2}{n^2}\right) \exp\left(\frac{i2\pi m j}{N}\right) \quad \text{for } n \neq 0$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} h[k] \quad \text{for } n = 0 \tag{2}$$

where $H[.]$ is the discrete Fourier transform (DFT) of $h[.]$.

For a discrete signal with the length N, DST has N^2 number of coefficients and large computational complexity of $O(N^3)$. The computational complexity can be reduced by introducing N number of orthogonal basis vectors of unit length. This DST is known as discrete orthogonal S-transform (DOST), and it generates N numbers of coefficients with a computational complexity of $O(N^2)$. By introducing FFT, the computational complexity of DOST reduces to $O(N \log N)$. The basis function used in this algorithm is defined in a fixed window. A generalized window-dependent basis function is developed for more efficient time–frequency representation of S-transform using FDOST algorithm. The FDOST coefficients are calculated by inner product of the signal $h(k)$ and the basis function $E^{\varphi}_{[p,\tau]}(k)$ as given in (3).

$$S^{\varphi}_{[p,\tau]} = \left\langle h(k), E^{\varphi}_{[p,\tau]}(k) \right\rangle = \left\langle F^{-1} R^{\varphi} H, D[k]_{[p,\tau]} \right\rangle = \left\langle F^{-1} R^{\varphi} H, D[k]_{[p,\tau]} \right\rangle \quad (3)$$

where H is the DFT of $h(k)$ and R^{φ} is a sequence function and

$$E^{\varphi}_{[p,\tau]}(k) = \frac{1}{\sqrt{\beta(p)}} \sum_{j=0}^{\beta(p)-1} \left[C^{\varphi}_{[p,j]}(\nu(p)) \right]^{-1} \exp\left(2\pi i (\beta(p) + j)\left(\frac{k}{N} - \frac{\tau}{\beta(p)} \right) \right) \quad (4)$$

$$D[k]_{[p,\tau]} = \frac{1}{\sqrt{\beta(p)}} \sum_{j=0}^{\beta(p)-1} \exp\left(i2\pi (\beta(p) + j)\frac{k}{N} \right) \exp\left(-i\frac{2\pi \tau j}{\beta(p)} \right) \quad (5)$$

The computational complexity remains same as $O(N \log N)$ but only N numbers coefficients are generated. These coefficients are localized in both time and frequency domains. Each FDOST coefficient measures a specific feature of a signal in different window-dependent bases. The energy of FDOST coefficients of a signal can be used to represent the feature of the signal and it is defined as the sum of square of absolute values of the coefficients.

$$E = \sum_{k=1}^{N} \left| S^{\varphi}_{k[p,\tau]} \right|^2 \quad (6)$$

3 Radial Basis Function Neural Network (RBFNN)

The radial basis function neural network consists of three layers: input layer, hidden layer and output layer. The hidden layers consist of radial basis function and receive input vector via unit connection weight. The hidden layer calculates the Euclidean distance between centre and input vector, and it maps the input vector into output space by nonlinear transformation. The output layer uses linear combiner with adjustable weight parameter. Here, the parameters can be determined using least

square method. If the input space is p dimensional and output space is m dimensional, the complete transformation can be written as

$$d_i = \omega_{0i} + \sum_{j=1}^{h} \omega_{ij} \varnothing \left(\left\| x - c_j \right\|, \beta_j \right) \tag{7}$$

where $d_i \rightarrow i$th output, $x \in R^p \rightarrow$ input vector, $\omega_{0i} \rightarrow$ biasing term, $\omega_{ij} \rightarrow$ weight between ith hidden node and jth output node, $c_j \in R^p \rightarrow$ centre of jth hidden node, $\beta_j \rightarrow$ real constant (spread factor), $\varnothing(\cdot) \rightarrow$ Nonlinear function and $i = 1, \ldots, m$ (number of output layer) and $j = 1, \ldots, h$ (number of hidden layer).

4 System Description and Fault Simulation

A part of West Bengal State Electricity Transmission Company Limited (WBSETCL) is designed in MATLAB Simulink for the fault investigation as shown in Fig. 1. The transmission lines are modelled using positive, negative and zero sequence components of resistance, inductance and capacitance. Total amount of load is 400 MW including 150 MW at Bus-3 and 250 MW at Bus-2. The specifications of the transmission lines are given in Table 1.

The transmission line 1 between Bakreswar and Arambag is considered to test the proposed algorithm. Four types of faults like line to ground fault (LG), line to line fault (LL), line to line to ground fault (LLG) and three-phase fault (LLLG) are generated at 12 locations for training purpose and another 7 locations for testing purpose. All these faults are simulated for fault resistances (R_f) of 10 and 100 Ω as well as for fault inception angles (FIA) of 0^0 and 90^0. Three-phase fault current
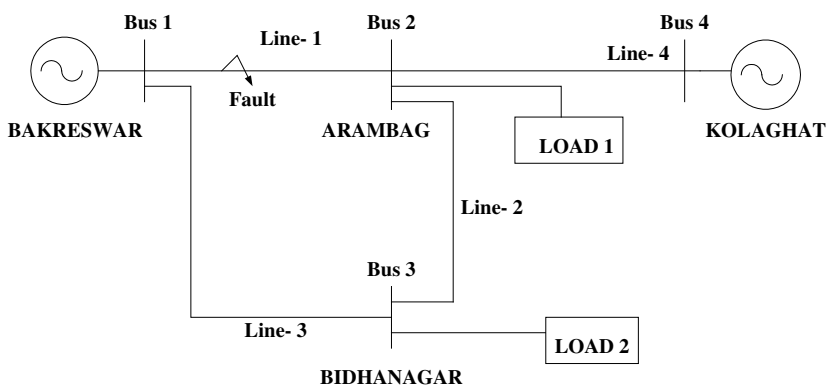
Fig. 1 The single-line diagram of the practical power system under investigation

Table 1 Specifications of the transmission lines

Line	Description	Length of the line (km)	Voltage (KV)
Line-1	Bakreswar to Arambagh	130	400
Line-2	Arambagh to Bidhannagar	114	400
Line-3	Bakreswar to Bidhannagar	40	220
Line-4	Kolaghat to Arambagh	64	400
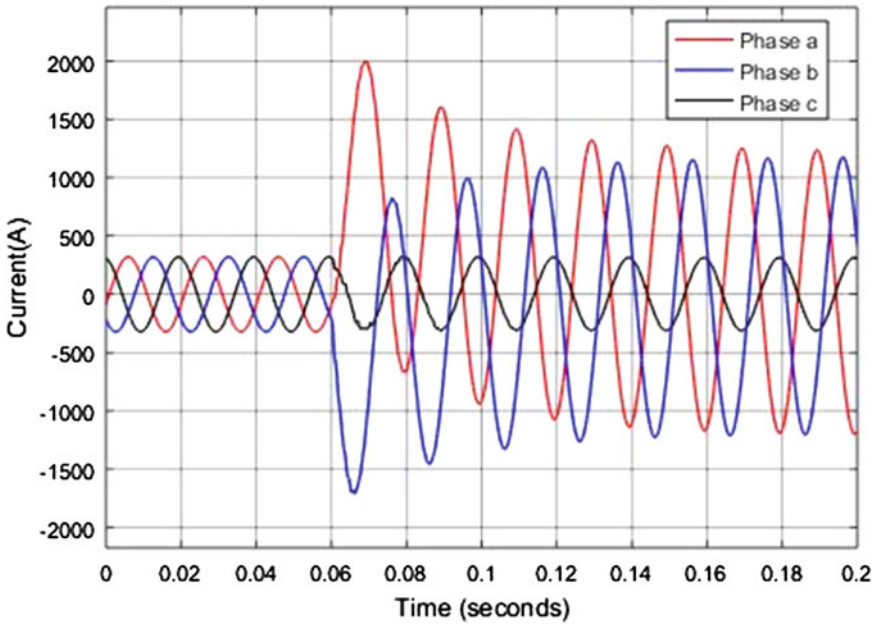
Fig. 2 Three-phase current signals for LLG fault at 70 km with fault resistance 10 Ω, FIA 0°

signals are recorded at Bus-1 with sampling frequency of 50 kHz. Fault current signals for three-phase fault at 70 km are shown in Fig. 2.

5 Feature Extraction and Fault Localization

The FDOST energy is calculated from half-cycle post-fault three-phase current signals, recorded for different conditions using Eq. (6). The FDOST energy is used as the fault feature. The extracted features are fed to RBFNN after normalization.

Fig. 3 Flowchart of the proposed fault localization algorithm
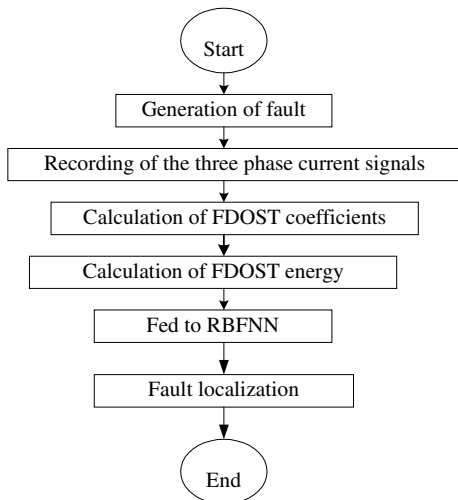
The RBFNN is trained with 12 sets of features for faults at 12 locations, and then it is tested for 7 fault locations on transmission line. The RBFNN is designed with three input vectors (features for three-phase currents) with a spread factor of 0.8. The complete flowchart is shown in Fig. 3.

6 Result and Analysis

Applying the proposed algorithm, the fault locations are estimated on transmission lines and the percentage error is calculated with respect to total length of the transmission line.

$$\text{Percentage Error} = \frac{\text{Estimated Location} - \text{Actual Location}}{\text{Total Length of Transmission Line}} \times 100 \quad (8)$$

The results of fault localizations for different types of faults, fault resistances (R_f) and fault inception angles (FIA) are tabulated in Table 2. From this table, it is clear that the proposed algorithm is accurate to locate faults at different conditions. Fault inception angle depends on the time of fault initiation and mostly unpredictable. Thus, the nature of fault transients changes with the time of fault inception, although the accuracy of the proposed algorithm is independent of FIA as shown in Table 2. The nature of fault transients also depends on the fault resistances but the normalized values of FDOST-based fault features diminish the impact of fault resistance on fault localization on the transmission line.

Table 2 Result for fault localization for different conditions

Fault condition	Fault type	Actual fault location (km)	Estimated fault location (km)	% Error
$R_f = 10\ \Omega$ and FIA $= 0°$	LG	7	7.0509	0.0391
		25	24.9870	−0.01
		85	85.0064	0.0049
	LLG	47	46.9764	−0.018
		103	103.1240	0.095
		122	121.9142	0.066
	LL	85	84.9936	−0.005
		103	103.1118	0.086
		122	122.0298	0.023
	LLLG	64	64.0064	0.005
		85	85.0094	0.007
		103	102.9802	−0.015
$R_f = 10\ \Omega$ and FIA $= 90°$	LG	7	7.271	0.208
		25	24.814	−0.143
		85	85.546	0.420
	LLG	47	46.655	−0.265
		103	102.743	−0.198
		122	122.392	0.302
	LL	85	85.421	0.324
		103	102.964	−0.028
		122	122.149	0.115
	LLLG	64	63.823	−0.136
		85	84.694	−0.235
		103	102.951	−0.038
$R_f = 100\ \Omega$ and FIA $= 0°$	LG	7	6.8356	−0.126
		25	25.0873	0.067
		85	85.0539	0.041
	LLG	47	46.4955	−0.388
		103	103.189	0.145
		122	122.194	0.149
	LL	85	85.0143	0.011
		103	103.0632	0.049
		122	121.8884	−0.086
	LLLG	64	63.9334	−0.051
		85	85.0387	0.030
		103	103.0615	0.047

7 Conclusion

This paper presents a fault localization technique based on fast discrete orthogonal S-transform (FDOST) and radial basis function neural network (RBFNN) on transmission line. A part of WBSETCL, West Bengal is designed in MATLAB Simulink for the fault investigation. FDOST energy is extracted as fault feature from each of the three fault current signals. The normalized values of these features are fed to the RBF neural network for fault localization at different distances on the transmission line. The accuracy of the proposed algorithm is very high for different types of faults, fault resistances and fault inception angles (FIA).

References

1. Ji L, Booth C, Kawano F, Beaumont P (2014) Improved fault location through analysis of system parameters during auto reclose operations on transmission lines. IEEE Trans Power Deliv 29:2430–2438
2. Samantaray SR (2013) Fast S-transform based distance relaying in transmission line. Electr Power Syst Res 95:268–274
3. Hamidi RJ, Livani H (2017) Traveling wave-based fault location algorithm for hybrid multi-terminal circuits. IEEE Trans Power Deliv 32(1):135–144
4. Livani H, Evrenosoglu CY (2014) A machine learning and wavelet-based fault location method for hybrid transmission lines. IEEE Trans Smart Grid 5(1):51–59
5. Ngaopitakkul A, Bunjongjit S (2013) an application of a discrete wavelet transform and a back propagation neural network algorithm for fault diagnosis on single-circuit transmission line. Int J Syst Sci 44(9):1745–1761
6. Chatterjee S, Chakravarti S, Roy CK, Dey D (2008) Wavelet network-based classification of transient using dominant frequency signature. Electr Power Syst Res 78:21–29
7. Bhalija B, Maheshwari RP (2007) An adaptive distance relaying scheme using radial basis function neural network. Electr Power Compon Syst 35:245–259
8. Patel B (2018) A new FDOST entropy based intelligent digital relaying for detection, classification and localization of fault on transmission line. Electr Power Syst Res 157:39–47
9. Roy N, Bhattacharya K (2015) Detection, classification and estimation of fault location on an overhead transmission line using S-transform and neural network. Electr Power Compon Syst 43(4):461–472
10. Stockwell RG, Mansinha L, Lowe RP (1996) Localization of the Complex Spectrum: the S-transform. IEEE Trans Signal Proc 44(4):998–1001
11. Segal R, Kothari ML, Madnani S (2000) Radial basis function (RBF) network adaptive power stabilizer. IEEE Trans Power Syst 15(2):722–727
12. Reddy MJB, Gopakumar P, Mohanta DK (2016) A novel transmission line protection using DOST and SVM. Eng Sci Technol Int J 19:1027–1039

Low-Voltage Hardware-in-Loop Test Model Using Real-Time Digital Simulator for Single-Phase Converter

Sumangal Bhaumik, Banibrata Mondal and Jitendranath Bera

Abstract This paper describes the development of low-voltage model of different stages of power converter for making the performance evaluation of it under different operational conditions. The operational conditions vary depending on the network to which the converters are to be interfaced—the types as well as the number of operational conditions to be considered are maximum, in particular, when they are to be tied with micro-grid or smart grid systems. The remedial measures of the effect of these changes in operational conditions are to be incorporated into the low-signal control schemes for its reliable, secured and long-term operation. The various real system operational conditions are simulated in real-time digital simulator (RTDS) and the prototype of the developed hardware in low voltage level is interfaced with it.

1 Introduction

Hardware-in-loop (HIL) test is becoming popular day by day in almost every field of engineering for the ease of its performance evaluation. The hardware could be tested under real-like environment and the performance can also be evaluated in real time. It is highly recommended in [1] to perform HIL testing for applications where live testing is merely possible and discontinuation in service is restricted. A power hardware in the loop testing is always performed before installation of a new device

S. Bhaumik (✉) · B. Mondal · J. Bera
Department of Applied Physics, University of Calcutta, Kolkata, India
e-mail: sumangalstar26@gmail.com

B. Mondal
e-mail: banibrata.mondal@gmail.com

J. Bera
e-mail: jitendrabera@rediffmail.com

like SVC, STATCOM, etc. Normally what is practiced is to build the hardware and test it under simulated environment using Power HIL (PHIL) test facilities. Thus the data of the test are collected and analysed to make necessary changes in the hardware if required during unsatisfactory performance. For almost every PHIL testing the device under test which is to be installed is directly tested with the HIL test facility and no separate low-voltage model is fabricated for performance testing, making modification in the hardware or replacements in some parts of the hardware if required [1–4].

These complications could be averted if the same hardware is replicated into a low-voltage model and by testing it with the help of HIL, thus analysing the performance of the low-voltage model instead of the high-voltage one which is costly and time consuming to fabricate in case of unsatisfactory performance. The low-voltage model is always easy to build and less time consuming to fabricate. During the HIL test, for unsatisfactory performance the modification in the model can be done quickly and easily. Once the model gives satisfactory performance the required hardware can be built. It is seen that with the development of low-voltage model HIL testing, the time, complexity, required modification cost, etc. is decreased. It is described in [4], that how software simulation is helpful for the purpose of HIL simulation.

PHIL test on power systems has been discussed in [5] in this article it has been shown how HIL testing gives better result as compared to normal simulation. The output of HIL test has been compared with the software simulation output which clearly shows the difference.

Based on these previous works, the authors are motivated to take the challenges in PHIL testing with low-voltage model. This paper is organized as follows. Section 2 describes the importance of HIL testing along with the previous work done. The basic building blocks for low-voltage HIL testing of converters and the step-by-step process to perform the test are described in Sect. 3. Section 4 describes the case studies for the output and input signals through various I/O devices of RTDS with different kinds of signal generations as well as various analog feedback. Section 5 concludes about the development.

2 Review Work

This literature describes how low-cost real-time simulation could be performed with the help of VTB-RT which is a real-time extension of virtual test bed. Boost converted and H- bridge inverter has been used as the device under test. The comparison between the theoretical result and the practical test result using VTB-RT shows the competency of this approach [2]. The process of the experimentation, simulation result, related mathematics are well explained and part of the algorithm used is also described. In [3], a general framework for FPGA-based emulation of an electrical machine has been detailed. Difference between MATLAB/Simulink simulation and real-time simulation is shown. Comparison of rotor speed versus time curve with FPGA-based emulation and experimentally measured results are also depicted in the
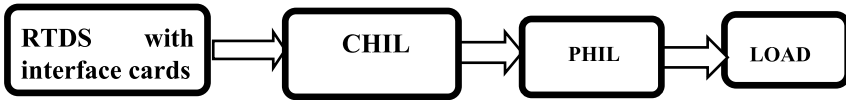
Fig. 1 Basic building blocks of HIL

article. Literature [4] shows the basic HIL process, interfacing card details, power amplifier requirements for the purpose of power hardware in loop test. Various DAQ cards their different connection schemes and interface algorithm has been detailed. Problems of conventional buck-based amplifier and the possible solution is also described. Testing of reactive load and PHIL operation during a fault is addressed. PHIL simulation in power system network and the advantage of conjoin software hardware simulation, hybrid configuration of the simulator is well addressed in this article [5]. Switched mode power amplification, linear power amplification and generator type power amplification process is described in detail. In [6], a three-level buck converter and its modulation have been investigated. Ripple minimization, Startup modulation, Level shifting modulation has been well described and experimental support is provided. Electric thermal photovoltaic modelling with the use of power hardware in loop test is described in literature [7]. The electrical and thermal characteristics of a PV cell model are depicted with experimental result as support.

3 Materials and Methods

The block diagram shown in Fig. 1 is the scheme which is elaborated in this paper, RTDS facility is used to perform the control HIL test on the control circuit of the single-phase converter. After satisfactory performance of the CHIL test, the PHIL test is performed on the single-phase converter circuit which is then connected with the load.

The power hardware in loop test of the single-phase inverter is proposed and the software simulation is performed.

3.1 Block Schematic of Proposed HIL

The block diagram shown in Fig. 2 is the scheme of the experiment where the boost converter is fed from the 12 V DC supply. The duty cycle of the boost converter is controlled by the duty cycle controller circuit using TL494. DDAC card also provides duty cycle control signal as a real-time output from simulated hardware with in RTDS. Either the signal from DDAC card or from the TL494 is used to operate the boost converter. DOPTO card takes feedback from the output of the duty cycle controller.
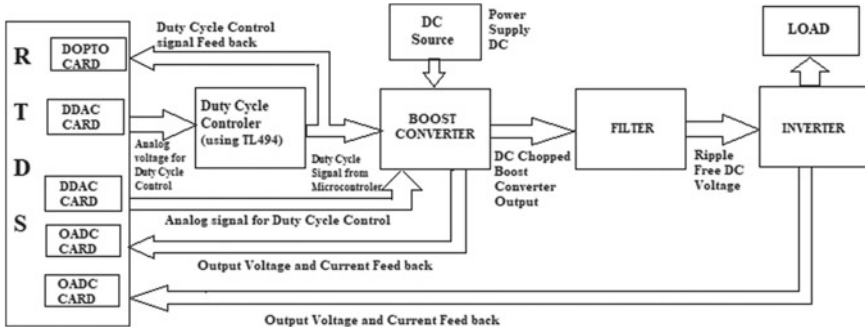
Fig. 2 Schematic of proposed HIL rest

The output of the boost converter after filtering by the LC filter is fed back to the RTDS through OADC card for its real-time simulation. The RSCAD software simulates the functioning of boost converter with this real converter and keeps track of its output through PWM signal generation. This DC voltage is then fed to the single phase inverter. The inverter output voltage is also fed back to the RTDS facility with the help of OADC card.

3.2 Mathematical Model of Boost Converter

The circuit shown in Fig. 3a is of a boost converter, where 'V_{in}' is the DC supply voltage, 'S' is the chopper switch (MOSFET), 'D' for diode, 'L' for inductor, 'C' is capacitor and 'R' represents the resistive load. During switch on condition, the diode does not conduct, the current 'I_L' flow through the inductor and the equivalent circuit looks like as shown in Fig. 3b.

When switch is turned off the equivalent circuit is shown in Fig. 3c.

$$\text{Output voltage, } V_{out} = (1/(1 - \alpha)) * V_{in}, \tag{1}$$

where, duty cycle $\alpha = (T_{on}/T)$, T_{on} is turn on time of the chopper switch and T is total time period.

The transfer function of the boost converter used for the modelling is given by

Fig. 3 Schematic of boost converter along with on and off switching states
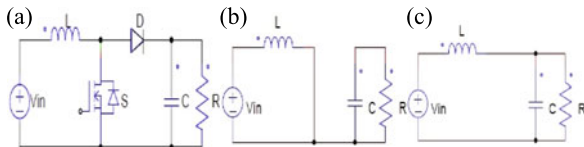
$$G(s) = \frac{V_o}{1 - \alpha} * \frac{1 - Ls}{(1 - \alpha)^2 * R} * \frac{(1 - \alpha)^2 * R}{RLCs^2 + Ls + (1 - \alpha)^2 * R} \qquad (2)$$

where, $L = \frac{V_{in} * \alpha}{f * \Delta I_L}$

3.3 Converters-RTDS Interfacing at Low Voltage Level

Step-by-step development of the boost-converter hardware and interfacing various cards with the RTDS test facility is described as follows.

For high-voltage testing, RTDS with power amplifier is to be interfaced with the converters. As low-voltage model testing is of our interest, power amplifier is not required and it could be directly interfaced with the existing DDAC or OADC cards of the RTDS. The various kinds of basic functionalities of the converters such as effect of change of load on their voltage and frequency regulation, dynamic performances under faulty conditions, the output power quality, etc. can be studied at the low-voltage levels. On the basis of satisfactory performance it can be tested with high-voltage systems. Any deviations from the desired performances of the converters can be analysed and corrective measures can be taken at the low-voltage level with greater flexibility and without the need for high voltage equipment.

4 Experimentation

4.1 Duty Cycle Controller Using TL494

Details of boost converter hardware and duty cycle controller using 'TL494' are discussed. As per the working principle of TL494, an analog output voltage is fed in its feedback pin to get the PWM output duty cycle which is shown in Fig. 4 with the help of DSO. The DDAC card is used to generate the analog voltage in order to vary the duty cycle as per the desired output and this signal is used to drive the IGBT/MOSFET of the fabricated Boost converter.

Fig. 4 HIL test on boost converter

4.2 HIL Test of the Boost Converter

In the first step, the converter duty cycle ratio is controlled with the TL494 and the filtered output voltage of the converter is fed to the RTDS through OADC card. The data for the voltage are collected and stored for comparison purpose. In the second part which is the extended part of the previous one, the duty cycle controller is replaced with the RSCAD software built model that could generate the duty cycle which is fed to the boost converter chopper switch using the DDAC. So the model of the duty cycle controller which is developed in the RSCAD software acts as a replica of the hardware built with TL494. Thus the hardware part becomes a part of the simulation and the simulation model is so developed that it acts like a real hardware one.

4.3 Case Studies and Simulation Results

A. Generation of Triangular wave using DDAC card and other required components

The DDAC card is configured and used to generate a ramp signal which could be used as a carrier signal for PWM generation and the same is controlled and monitored with the help of RSCAD software.

Figures 5 and 6 represent a draft schematic and the runtime mode of operation of RSCAD software for the ramp signal generation. The draft file shows one DDAC output component and three sliders which are used to control the voltage magnitude, frequency and phase of the ramp signal. From the RUNTIME mode the sliders for controlling of the voltage magnitude, frequency and phase are shown.

From the draft file, 3PC processor is assigned (Fig. 7) to generate the sawtooth waveform (Fig. 8). The sliders in the RUNTIME file are adjusted to get desired output from the DDAC card.

Fig. 5 RSCAD draft
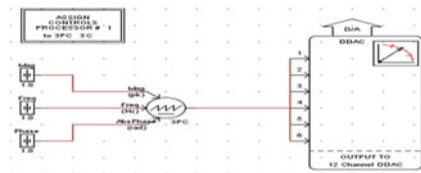
Fig. 6 RUNTIME modes
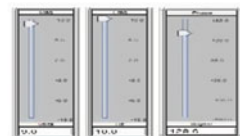
Fig. 7 Processor selection block
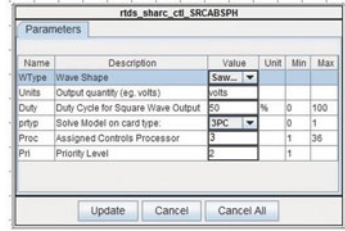
Fig. 8 Triangular wave

Fig. 9 RSCAD draft file
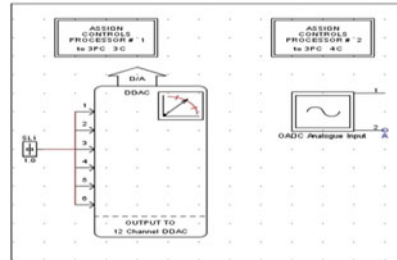
B. Receiving analog functions as input using DDAC card and observing the same through OADC card in the RSCAD software

The DDAC and OADC cards are used for interface as analog output and analog input respectively of a circuit.

Figure 9 shows the control blocks of DDAC and OADC cards of RTDS from RSCAD software and their respective parameters setting blocks are shown in Fig. 10. The DDAC card is used with the slider to act as a variable voltage source and the OADC card is used to take feedback of the analog output of hardware under test (HUT).

Hardware connections are made between one of the six channels of the DDAC card and channel 2 of the OADC card to obtain the same analog signal generated by the DDAC card (Fig. 11). The slider position (Fig. 12) and the corresponding output is shown in Fig. 11. The interfacing and hardware connection between the DDAC and OADC card can now be used for HUT.

C. The basic HIL testing using DDAC card, OADC card and voltage divider

The basic HIL testing is described with the help of a potential divider of using two 1 kΩ resistors. The output of divider is connected with the DDAC card and the

Fig. 10 Parameter selection block
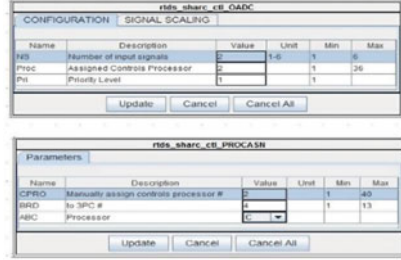
Fig. 11 RSCAD RUNTIME file
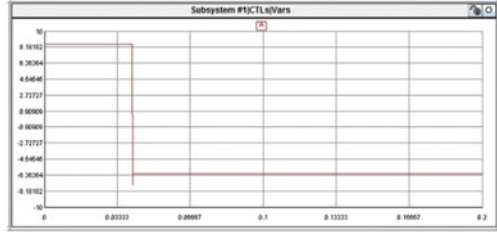
Fig. 12 RSCAD RUNTIME file
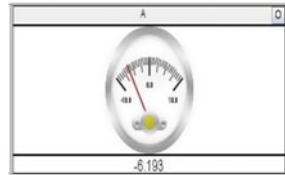
feedback is taken across one resistor. Channel 2 of the OADC card is used to take feedback from the voltage divider.

Like the principle shown in Fig. 9, a slider is used to control the voltage across the resistor string which can be anything between +10 to -10 V. The divider output is measured through OADC as 4.368 V for +10 V output. This is the last step before the fabricated boost converter is interfaced with the HIL test facility.

5   Conclusion

The basic setup for the HIL testing of the converters using RTDS at low-voltage levels are developed so that the required functionality testing of the converters can be tested.

The procedures to generate and accept various digital and analog signals to and from the HUT through RTDS cubicle and RSCAD software are described which are the integral part of HIL testing. The input and output signals as well as control

trigger pulses of converters are to be directly interfaced with these I/O cards for their performance study.

The above experimentation helps to understand and debug the problems related to the single-phase power converters which are to be tied with smart grid or microgrid. The required remedial measures can also be taken by the controller to avert these problems. The control hardware in loop test and power hardware in loop test procedures for maximum fault condition can also be analysed.

References

1. Mudrov M, Zyuzev A, Konstantin N, Valtchev S, Valtchev S (2017) Hardware-in-the-loop system numerical methods evaluation based on brush DC-motor model. In: optimization of electrical and electronic equipment (OPTIM) & 2017 International Aegean conference on electrical machines and power electronics (ACEMP), pp 418–433
2. Lu B, Wu X, Figueroa H, Monti A (2007) A low-cost real-time hardware-in-the-loop testing approach of power electronics controls. IEEE Trans Ind Electron 54(2):919–931
3. Tavana NR, Dinavahi V (2015) A general framework for FPGA-based real-time emulation of electrical machines for HIL applications. IEEE Trans Ind Electron 62(4):2041–2053
4. Jha K, Mishra S, Joshi A (2015) Boost-amplifier-based power-hardware-in-the-loop simulator. IEEE Trans Ind Electron 62(12):7479–7488
5. Lauss GF, Omar Faruque M, Schoder K, Dufour C, Viehweider A, Langston J (2016) Characteristics and design of power hardware-in-the-loop simulations for electrical power systems. IEEE Trans Ind Electron 63(1):406–417
6. Carstensen C, Biela J (2016) A three-level buck converter with a wide voltage operation range for hardware-in-the-loop test systems. IEEE Trans Ind Electron 31(9):6176–6191
7. Mai XH, Kwak S-K, Jung J-H, Kim KA (2017) Comprehensive electric-thermal photovoltaic modeling for power-hardware-in-the-loop simulation (PHILS) applications. IEEE Trans Ind Electron 64(8):6255–6264

Power Quality Improvement by Active Shunt Filter with Hysteresis Current Controller

Manas Mukherjee and Abhik Banerjee

Abstract This paper envisages the method to improve power quality of utility grid. Power drawn by nonlinear and unbalanced load contains higher order of harmonics components which in turns increases the total harmonics distortion level of supply current. We have used active shunt filter to improve the power quality with the help of hysteresis current controller. The p-q theory has been used to calculate the reference current to be compared with the load current drawn. With the help of hysteresis controller the gate triggering pulses for voltage source inverter used as active shunt filter have been generated. In our work Clarke's transformation matrix has been used for conversion of three phase voltage and current into two-axis α-β component. The whole system has been simulated in MATLAB software with a specific objective to improve the quality of power by reducing total harmonics distortion level and we have achieved that goal.

Keywords Power quality · Nonlinear and unbalanced load · Active shunt filter (ASF) · P-Q theory · Hysteresis current controller · Total harmonic distortion (THD)

1 Introduction

Rapid industrialization and consumer's demand of power has increased the generation capacity with a significant consideration of power quality issue. The use of nonlinear load and electronics devices such as computer, electronic bulb, inverter, etc., is a common trend in today's society [1–5]. As the repercussion these types of load inject harmonics components in the transmission line which pollutes the supply power and degrades the power quality [1–10]. As a result several disturbances

M. Mukherjee (✉) · A. Banerjee
National Institute of Technology Arunachal Pradesh, Yupia 791112, Arunachal Pradesh, India
e-mail: manaselectrical.91@gmail.com

A. Banerjee
e-mail: abhik_banerjee@rediffmail.com

such as excessive vibration of motor, damage of capacitor, appliances deterioration, excessive heat formation, etc., happen frequently [1]. So here the term good power and bad power come in today's power dispatch scenario. By the term good power we mean the frequency of supply current should be 50 Hz and the deviation should be in the range of $\pm 5\%$. Besides this the voltage profile should be in the desired level. There should be neither swell nor sag of the supply voltage and the supply waveform should be pure sinusoidal. But today's load profile does not permit to maintain such supply profile due to the inherent nonlinear characteristics of loads. To mitigate this so many methods are applied with the help of advanced power electronics based technologies like Dynamic Voltage Restorer (DVR), Static compensator (STATCOM), Unified Power Flow Controller (UPFC), Passive filter like LC, LCL, etc. [9, 10]. Here we have a novel approach to improve the power quality and minimize the THD by using active shunt filter. The drawback of passive filter is resonance problem [1]. Besides this the components in passive filter show aging problem. They are bulky and require more space [2]. Active filter is free from such type of problem and shows good performance in respect of speed and accuracy [2]. So many techniques are employed to control the current of active shunt filter like sinusoidal PWM [1], Fuzzy algorithm based current controller [5], etc. With the increase in switching speed of PWM the power waveform is also improved but switching loss is increased. Advanced method like adaptive fuzzy hysteresis band current controller has been proposed by [5] which minimize the uncertainty of system. In this paper we have simulated hysteresis current controller based active shunt filter with our main focus to improve supply current waveform with low THD profile.

2 Active Shunt Filter

A pictorial representation of active shunt filter is provided in Fig. 1 The basic structure of active power filter consists of a voltage source inverter [1, 4, 5].

The voltage source inverter is controlled by current. Its main function is to inject compensating current to the supply side in phase opposition to the harmonics components of load current. It injects compensating current in accordance with the current drawn by nonlinear and unbalanced load and cancels the harmonics introduced by nonlinear load. It also enables source to deliver only active power by injecting reactive power demand and maintains supply current sinusoidal [8–10]. In Fig. 1 R is resistance of the transmission line, L denotes transmission line inductance, Lf is inductance of filter connected to inductor. Rnl_load denotes nonlinear loads. R1, R2, R3, L1, L2, L3, C1, C2, C3 represents resistance, inductance and capacitance of unbalanced load in each phase.
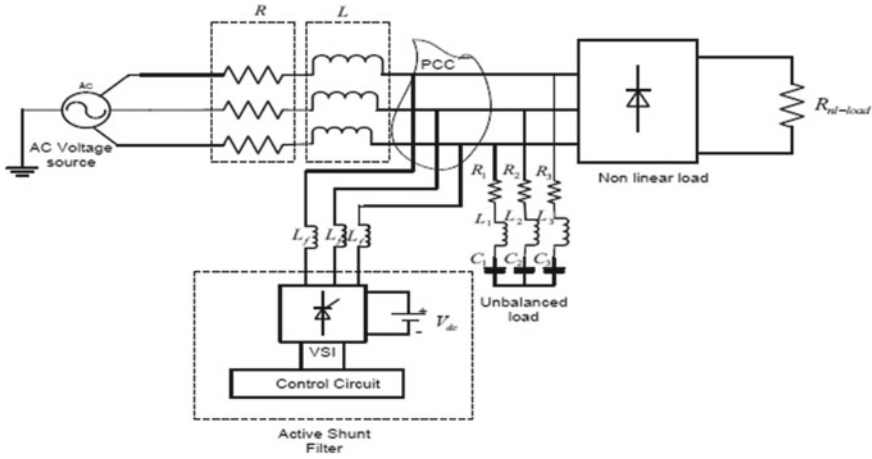
Fig. 1 Nonlinear and Unbalanced load with active shunt filter

2.1 Calculation of Reference Current by P-Q Theory

From Clarke's transformation matrix equation we get two axis component of three-phase system. Three-phase voltage and current can be transformed into voltage and current of d-axis and q-axis component [6, 7].

For voltage transformation in three-phase three wire system from Clark's transformation matrix equation [5, 10] we get (1)–(4)

$$v_\alpha = \sqrt{\frac{2}{3}}\left(v_a - \frac{v_b}{2} - \frac{v_c}{2}\right) \tag{1}$$

$$v_\beta = \sqrt{\frac{2}{3}}\left(\frac{\sqrt{3}v_b}{2} - \frac{\sqrt{3}v_c}{2}\right) \tag{2}$$

For current transformation we get from [5, 10]

$$i_\alpha = \sqrt{\frac{2}{3}}\left(i_a - \frac{i_b}{2} - \frac{i_c}{2}\right) \tag{3}$$

$$i_\beta = \sqrt{\frac{2}{3}}\left(\frac{\sqrt{3}i_b}{2} - \frac{\sqrt{3}i_c}{2}\right) \tag{4}$$

The active and reactive power can be obtained from Eqs. (5) and (6)

$$p = v_\alpha i_\alpha + v_\beta i_\beta \tag{5}$$

$$q = v_\alpha i_\beta - v_\beta i_\alpha \tag{6}$$
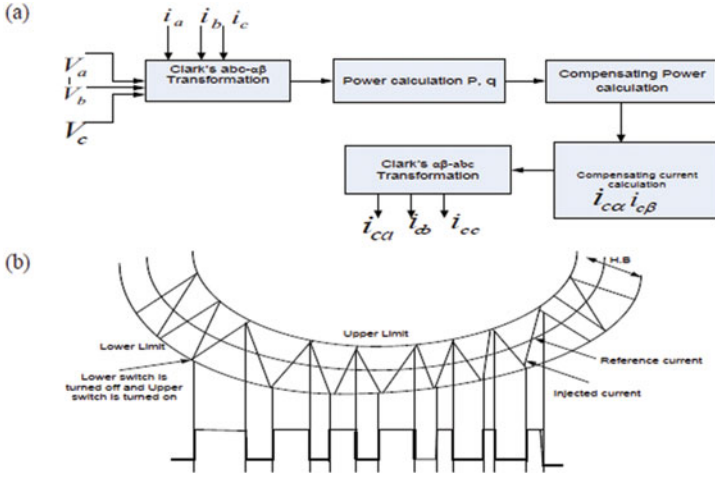
Fig. 2 **a** Block diagram of reference current calculation. **b** Hysteresis current controller

Now from these equations obtained from Clarke's abc-αβ transformation the reference current can be calculated [6, 7]. The generation of reference current is delineated in Fig. 2a. At first supply voltage and current are transformed into two axis components by Clark's transformation. The active and reactive powers are calculated from Eqs. (5) and (6). The calculated active power contains fundamental power called average power and higher order harmonics power called oscillating power. The oscillating power is separated from the average power by using a high pass filter. Now we can derive reference current $i_{c\alpha}$ and $i_{c\beta}$ from calculated oscillating power, reactive power, and two-axis component of supply voltage. Clarke's inverse matrix transformation is taken into consideration to calculate the reference current [8].

$$\begin{bmatrix} i_{c\alpha} \\ i_{c\beta} \end{bmatrix} = \frac{-1}{v_\alpha^2 + v_\beta^2} \begin{bmatrix} v_\alpha & v_\beta \\ -v_\alpha & v_\beta \end{bmatrix} \begin{bmatrix} p_{oscillating} \\ q \end{bmatrix}$$

These two axis reference currents are further converted to three-phase current [8].

$$\begin{bmatrix} i_{ca} \\ i_{cb} \\ i_{cc} \end{bmatrix} = \frac{\sqrt{2}}{\sqrt{3}} \begin{bmatrix} 1 & 0 \\ \frac{-1}{2} & \frac{\sqrt{3}}{2} \\ \frac{-1}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} i_{c\alpha} \\ i_{c\beta} \end{bmatrix}$$

Here i_{ca}, i_{cb} and i_{cc} are three-phase reference current.

2.2 Hysteresis Current Controller

Figure 2b shows hysteresis controller. For the purpose of controlling the compensating current hysteresis current controller method is a novel one. It has advantages of quick dynamic response as well as simplicity [1]. Besides these it provides smooth switching frequency for fixed switching frequency [2, 5]. A current control VSI is connected in parallel with the load in order to generate reference compensating current. When the injecting filter current through the coupling inductor crosses the reference current out of either upper or lower band limit the upper and lower switches of inverter are tripped according to hysteresis controller logic [1, 2, 5, 10]. The switching control strategy follows [1] and is shown in Fig. 2b.

S (Switching State) = 0: if $i_f(t) > i_f^*(t) + \frac{H \cdot B}{2}$
The upper switch is off and lower switch is on, $S = -v_{dc}$
$S = 1$: if $i_f(t) < i_f^*(t) - \frac{H \cdot B}{2}$
The upper switch is on and lower switch is off, $S = +v_{dc}$.
Thus it happens to maintain injected current to be in hysteresis band limit.

3 MATLAB Simulink Model

The MATLAB Simulink model in Fig. 3 shows that a nonlinear rectifier load and an unbalanced three-phase load are connected to a three-phase voltage source through circuit breaker and three phase V-I measurement unit. In parallel to nonlinear and unbalanced load an active shunt filter is connected. Active shunt filter comprises of an IGBT-based voltage source inverter which is connected to PCC (Point of common coupling) through coupling inductor. At the input of the VSI two capacitors are connected in series. A PI controller is connected to the input dc voltage of VSI with a reference voltage. From p-q theory [6, 7] all equations are formed using mathematical functional blocks in MATLAB software. After the reference current calculation is done the Boolean NOT gate and relational operator have been used for implementation of hysteresis current controller to generate gate pulses. The generated gate pulses have been given to the gate input of VSI. In hysteresis current controller circuit model the reference current has been compared with the measured supply current and according to the error the filter current ramps up and down through inductor and provides compensating current to the supply side in phase opposition to the harmonics components eliminating harmonics introduced by nonlinear and unbalanced load.
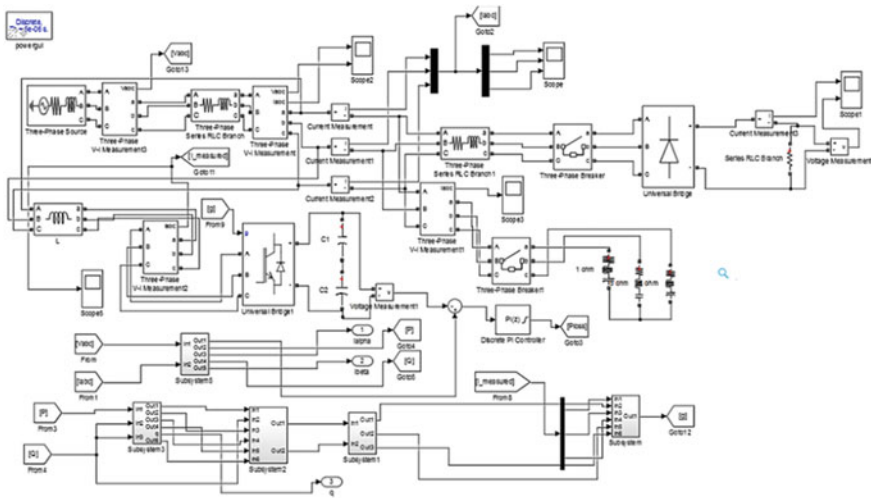
Fig. 3 MATLAB Simulink model

4 Result and Discussion

Before connecting the shunt active power filter supply voltage and current, total harmonics distortion (THD) of supply current, DC load voltage and current and unbalanced load current are observed. As the two circuit breakers (one each connected to nonlinear load and unbalanced load) are switched after 0.2 s the THD is measured after starting time 0.2 s for 100 cycles. The two loads draw current which shows 30.26% THD and the supply current is highly distorted. After the active filter is connected parallel to the load the THD goes down and comes to 5.32%. Before connecting active filter the fundamental current component was 64.03 A and after connecting the filter the fundamental current is 65.78 A. Besides this the supply current shows sinusoidal waveform. The compensating filter current is observed and along with all the measured quantities is shown in Figs. 4 and 5. Figure 4a shows supply voltage and current without filter. The supply current is not pure sinusoidal. Figure 4b shows that the supply current contains 30.26% THD. The fundamental component has frequency of 50 Hz and the fundamental component contains the magnitude of 64.03 A. Figure 4c shows the three-phase unbalanced current which contains higher notches. Figure 4d shows voltage and current waveform after the shunt filter is connected. This shows pure sinusoidal current waveform which was a distorted one in Fig. 4a. Figure 4e shows the THD in supply current with active shunt filter. Without active shunt filter Fig. 4b shows that the THD level is 30.26% where as Fig. 4e shows that the THD is only 5.32%. This shows that the THD level has been reduced significantly to a lower margin. Figure 5a represents the waveform of compensating current flowing through coupling filter which ramps filter current up and down to remain within the hysteresis band and trace the reference current.
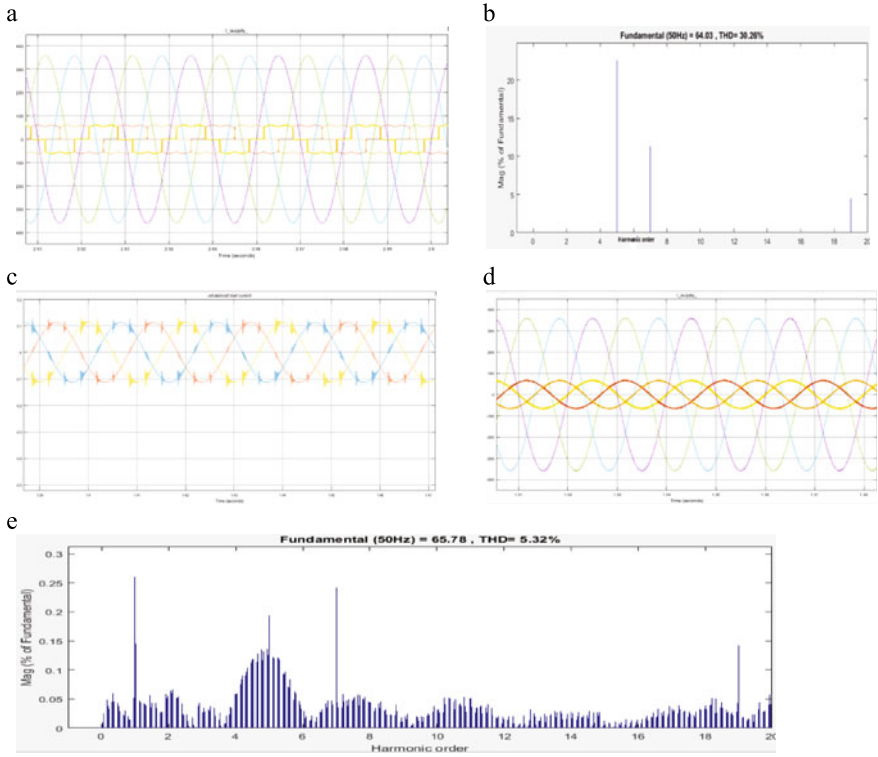
Fig. 4 **a** Supply voltage and current without filter, **b** THD of supply current without filter, **c** unbalanced load current without filter, **d** supply voltage and current with filter, **e** THD of supply current with filter
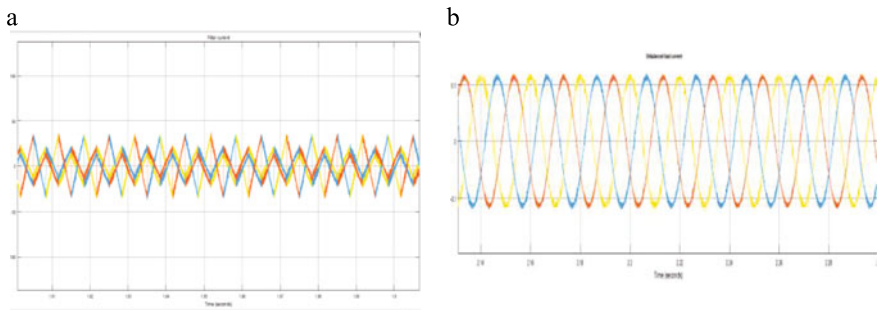
Fig. 5 **a** Compensating current through coupling inductor, **b** unbalanced load current with SAF

Figure 5b shows the unbalanced load current with active power filter and this contains fewer notches whereas Fig. 4c shows higher notches.

Table 1 Value of various parameters which were used in the system modeling

Name	Value
Ph–Ph voltage of three phase voltage source	440 V
Circuit breaker switching time	0.2 s
Nonlinear Dc load	10 Ω
Unbalanced RLC load	Phase A: R = 1 Ω, L = 1 mH, C = 1 μF; Phase B: R = 10 Ω, L = 1 mH, C = 1 μF; Phase C: R = 17 Ω, L = 1 mH, C = 1 μF
Filter inductance	0.39 mH
Capacitor used in VSI	50 μf

5 Conclusion

From the above result it is concluded that we have achieved our goal in reducing Total Harmonic Distortion of supply current and improving the power quality. By using hysteresis current controller based shunt active filter we have improved the shape of supply current waveform. This shows the transfer of a better quality of power. However more advanced methodologies of controlling method have been introduced such as adaptive fuzzy hysteresis band—current controller, space phasor based hysteresis current controller [3] etc. In future our focus is to improve power quality issues with the help of more advanced controlling methodology.

Appendix

See Table 1.

References

1. Kale M, Ozdemir E (2005) An adaptive hysteresis band current controller for shunt active power filter. Electr Power Syst Res 73(2):113–119
2. Suresh Y, Panda AK, Suresh M (2012) Real time implementation of adaptive fuzzy hysteresis-band current control technique for shunt active power filter. IET Power Electron 5(7):1188–1195
3. Chauhan SK, Shah MC, Tiwari RR, Tekwani PN (2013) Analysis, design and digital implementation of a shunt active power filter with different schemes of reference current generation. IET Power Electron 7(3):627–639
4. Hareesh M, Kumar GS, Sreenivasarao D (2015) Power quality improvement by shunt active filter for low cost applications. In: 2015 international conference on computation of power, energy information and communication (ICCCPEIC), 22 Apr 2015, pp 0153–0158. IEEE
5. Panda AK, Patel R (2015) Adaptive hysteresis and fuzzy logic controller-based shunt active power filter resistant to shoot-through phenomenon. IET Power Electron 8(10):1963–1977

6. Akagi H, Kanazawa Y, Fujita K, Nabae A (1983) Generalized theory of the instantaneous reactive power and its application. Trans Inst Electr Eng Jpn B 103(7):483–490
7. Akagi H, Kanazawa Y, Nabae A (1984) Instantaneous reactive power compensators comprising switching devices without energy storage components. IEEE Trans Ind Appl 625–630
8. Shah A, Vaghela N (2014) Shunt active power filter for power quality improvement in distribution system. Int J Eng Dev Res 1(2):23–27
9. Poonia A, Dey A (2016) Space phasor based improved hysteresis current controller for shunt active power filter using 3–level inverter. In: 2016 18th European conference on power electronics and applications (EPE'16 ECCE Europe), 5 Sep 2016, pp 1–10. IEEE
10. Ahuja RK, Chauhan A, Sharma S (2012) Hysteresis control for current harmonics suppression using shunt active filter. Int J Adv Res Technol 1(6):17–24

Evaluation of Energy Power Spectral Distribution of QRS Complex for Detection of Cardiac Arrhythmia

Akash Kumar Bhoi, Karma Sonam Sherpa and Bidita Khandelwal

Abstract The proposed approach involves detection of QRS complex and energy power spectral distribution analysis of the segmented QRS complex to establish the presence of arrhythmic beats in Electrocardiogram (ECG). The methods consist of three steps: (i) the baseline drift and high-frequency artifacts could seriously affect the detection performance, so Moving Average Filtering (MAF) and Stationary Wavelet Transform (SWT) are implemented at preprocessing stage. (ii) Localization of R-peaks by implementing FFT-based windowing and thresholding techniques. Then Q and S points are detected using search interval method based on the medical definition. (iii) The segmented QRS complex is analyzed with period-gram and Continuous Wavelet Transform using FFT (CWTFT) to obtain time–frequency domain power and energy of the complex. (iv) Statistical analysis has been proposed using one-way ANOVA to differentiate the healthy and arrhythmic QRS complex. The proposed QRS detection and analysis methodologies are evaluated with MIT-BIH Arrhythmia Database (MITDB) and FANTASIA database. The detection performance, i.e., Sensitivity $S_e(\%)$ and the Specificity $S_p(\%)$ for FANTASIA 100% each respectively, where as $S_e = 100\%$ and $S_p = 98.18\%$ for MITD. The failed detection percentage, $F_d(\%) = 0$ for FANTASIA and $F_d(\%) = 1.85\%$ for MITDB. The energy power distributed parameters obtained from PSD and CWTFT are statistically analyzed with one-way ANOVA and the p-value are found to be <0.05 (i.e., CI = 95%) for healthy and arrhythmia QRS complex which certainly signifies that the energy

A. K. Bhoi (✉)
Department of Electrical & Electronics Engineering, Sikkim Manipal Institute of Technology (SMIT), Sikkim Manipal University, Majitar, India
e-mail: akash.b@smit.smu.edu.in; akash730@gmail.com

K. S. Sherpa
Sikkim Manipal University, Gangtok, India
e-mail: karmasherpa23@gmail.com

B. Khandelwal
Department General Medicine, Central Referral Hospital and SMIMS, Sikkim Manipal University, Gangtok, India
e-mail: drbidita@gmail.com

power features of the arrhythmic QRS complex are different than the normal QRS complex.

Keywords QRS complex · CWTFT · One-way ANOVA · Energy power distribution · MITDB · FANTASIA

1 Introduction

Time domain features and morphological fiducial points can add crucial inputs towards decision-making and accurate diagnosis in electrocardiogram (ECG) [1]. Arrhythmia analysis can also be done by shape and time change observation of QRS complex [2]. Difference Operation Method (DOM) was proposed by Yun-Chi Yeh, et al. with a simple and computational-free algorithm for accurate detection of QRS complex [3]. The wavelet transform is widely used in many of the earlier proposed methods for QRS detection and analysis with the limitation of selection of proper mother wavelet and decomposition level [4–9]. Few well-known QRS detection approaches which are more complex *such as* Neural Network (NN) [5], Support Vector Machines (SVM) [10], Genetical algorithms [11], Hidden Markov Models (HMM) [12], K-Nearest Neighbor (K-NN) [13], parallel functioning of different algorithms [14, 15], Pan and Tompkins (PT) [16], Fractal Dimension Transformation (FDT) [17]. Unpredictable QRS Potentials (UIQP) were analyzed using Finite-Impulse-Response (FIR) based prediction model to identify ventricular tachycardia with patients undergoing high-risk ventricular arrhythmias [18]. Soroor and Jafarnia [19] proposed Multiresolution Wavelet with Thresholding Method and achieved 98.2% accuracy for QRS complex detection. QRS complexes could also be detected performing fixed structure Mathematical Morphology (MM) operators [20]. Atiyeh and Reza [21] have discussed a P-QRS-T waves-based detection approach, which is simple and accurate having very less response time during its real-time operation.

As from the literature survey, it is also clear that arrhythmia can be identified by analyzing QRS complex changes. Here the authors propose an algorithm by taking such problem into account. Classification between normal QRS and arrhythmic QRS of ECG is presented with efficient detection techniques and energy power spectral distribution of this complex. The performance evaluation is done with FANTASIA database and the MIT-BIH Arrhythmia Database (MITDB).

2 Materials and Methods

2.1 Datasets

The proposed methodologies have been tested over 30 selected healthy (Fantasia records) data having 1 row (signal) and 900,000 columns (samples/signal) with the duration of 1 h having sampling frequency: 250 Hz and sampling interval: 0.004 s [22]. 30 out of 48 lengthy arrhythmic records from MITDB, with sampling frequency 360 Hz and 11-bit resolution ranging over 10 mV are considered for this analysis. The proposed methodologies have been tested over these 30 selected arrhythmic signals (total beats considered: 4,000) having first row (signal) and 650,000 columns (samples/signal) with the duration of 1 h [23].

2.2 Preprocessing of Noisy ECG Signal

Moving average filter [24] and Stationary Wavelet Transform [25] were implemented to remove baseline drift and signal noise (Fig. 1).

2.3 R-Peaks Localization

Step 1: Discrete Fourier Transform (DFT) is computed from Fast Fourier Transform (FFT) and the Fourier transform is an integral (4) of the form

$$F(u) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi ux}dx \tag{4}$$
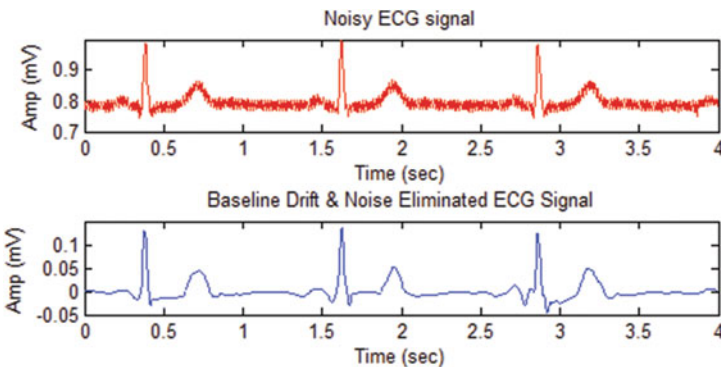
Fig. 1 Noise cancelation for FANTASIA data #f1o04 (4 s data for better visualization)
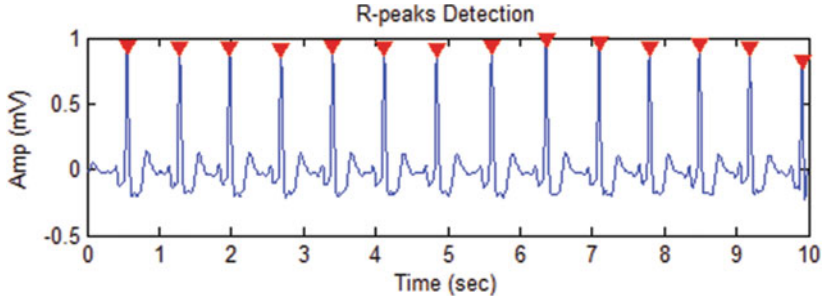
Fig. 2 R-peaks localization for MITDB data #105 (10 s data for better visual representation)

Low-frequencies component from the ECG signal is removed using FFT.

Step 2: Restoration of ECG signal is done with Inverse Fourier transform (5) and the expression for inverse Fourier transform is

$$f(x) = \int_{-\infty}^{\infty} F(u)e^{i2\pi ux}du \qquad (5)$$

Step 3: Windowed filter with default size for localization of maxima (only maximum values are being considered and other values are ignored).

Step 4: Implementation of threshold filters to remove small values and preserve significant ones.

Step 5: Repeat Step 3 with adjusting the size of the windowed filter to improve filtering performance. Then R-peaks are detected (Fig. 2).

Step 6: In case of negative QRS complexes, localization of minima is performed for detection of R-peaks and the other above steps remains same.

2.4 Q and S Inflection Points Localization

Step 7: Taking R point as standard, the search interval-1 locates 20 sampling points from prior and succeeding of R point. The least value found ahead of R point marked as Q1 and the same found after, marked as S1 [3].

Step 8: Search interval-2 is defined such that, it covered 40 sampling points prior and succeeding of R point. The least sampling point forth to R is Q2 and succeeding to R is S2 [3].

Step 9: Localization of Q by checking the location and amplitude values of Q1 and Q2. If Q1 and Q2 located at different points and Q1 amplitude > Q2 amplitude, then Q1 is the location of Q or vice versa.

Step 10: Localization of S by checking the position of S1 and S2. (i) If the position of S1 = position of S2, then S1 = S2 = S. (ii) If VS2 > VS1, then S = S1; else S = S2, where VSi = amplitude of Si and i = 1, 2 (Fig. 3).
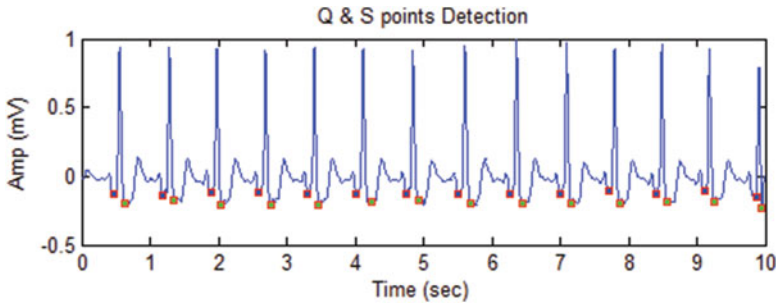
Fig. 3 Q & S point's localization for MITDB data #105 (10 s data for better visual representation)

The failed detection rates are calculated for both FANTASIA and MITDB databases. Failed detection is calculated by [3]:

$$F_d(\%) = \frac{FP + FN}{Total\ Beats} \times 100 \tag{6}$$

The failed detection percentage, $F_d(\%) = 0$ for FANTASIA and $F_d(\%) = 1.85\%$ for MITDB.

$$S_p(\%) = \frac{TP}{TP + FP} \times 100 \tag{7}$$

$$S_e(\%) = \frac{TP}{TP + FN} \times 100 \tag{8}$$

where, TP = True positive (properly detected beats), FN = False negative (failed to detect a real beat), FP = False positive (detects a beat when no beat is present).

The evaluation of the detection performance is computed using Specificity (S_p) and Sensitivity (S_e) [3], which are normally computed using (7) and (8). For FANTASIA both S_e and S_p found to be 100% where as $S_e = 100\%$ and $S_p = 98.18\%$ for MITDB (Table 1).

3   Feature Extraction from QRS Complex

3.1   Continuous Wavelet Transform Using FFT Algorithm (CWTFT)

The mean of the detected QRS complexes is segmented from the ECG signal and analyzed using CWTFT. CWT can be computed using the product of Fourier Transforms (Inverse Fourier Transform (IFT)), from the below (9) equation

Table 1 Energy power spectral distribution of QRS complex

FANTASIA	Power		Energy		MITDB	Power		Energy	
	F_domain	T_domain	Max (%)	Sca_Max		F_domain	T_domain	Max (%)	Sca_Max
f1o01m	0.9666	0.9135	0.4812	39.2000	101	0.0661	0.0890	0.5387	36.8000
f1o05m	0.9717	0.8902	0.4889	39.2000	105	0.3123	0.1436	0.6048	50
f1o08m	0.9672	0.9126	0.3985	39.2000	107	0.2900	0.2371	0.5552	38.2000
f1o10m	0.3629	0.2291	1.1681	19.4000	108	0.0705	0.0310	0.5373	36.3000
f1y01m	0.4764	0.2122	1.2507	17.4000	109	0.1344	0.0977	0.7259	50
f1y02m	0.2925	0.1944	1.1319	18.4000	111	0.0954	0.1010	0.7132	50
f1y03m	0.3654	0.1717	1.4919	15.5000	112	0.1459	0.1452	0.5266	36.8000
f1y04m	0.1725	0.1440	0.6246	29	117	0.2111	0.2336	0.5990	31
f1y05m	0.1102	0.1124	0.5895	29	119	0.0548	0.0698	0.6413	50
f1y06m	0.4842	0.2460	1.6021	14.5000	121	0.3989	0.1880	0.6975	47.4000
f1y07m	0.1835	0.1557	0.8417	23.2000	123	0.0918	0.1408	0.4991	40.7000
f1y08m	0.4324	0.2092	1.4796	15.5000	124	0.2284	0.1754	0.7395	50
f1y09m	0.2414	0.1127	0.5425	35.8000	200	0.0613	0.0639	0.5255	42.6000
f1y10m	0.5812	0.2779	1.4890	15.5000	202	0.2512	0.1345	0.6968	50
f2o01m	0.4469	0.2339	1.2997	16.5000	203	0.2671	0.1339	0.5146	33.4000
f2o02m	0.5237	0.2477	0.9863	23.2000	205	0.1201	0.1080	0.4952	39.7000
f2o03m	0.1437	0.1732	0.7645	26.1000	207	0.5461	0.2787	0.6694	34.8000

(continued)

Table 1 (continued)

FANTASIA	Power		Energy		MITDB	Power		Energy	
	F_domain	T_domain	Max (%)	Sca_Max		F_domain	T_domain	Max (%)	Sca_Max
f2o04m	0.4680	0.2551	1.2254	17.4000	208	0.1184	0.1106	0.5224	44.5000
f2o06m	0.4842	0.2460	1.6021	14.5000	210	0.2389	0.1445	0.6647	50
f2o07m	0.2220	0.1670	0.8252	25.2000	212	0.2795	0.1707	0.5330	41.6000
f2o09m	0.4420	0.2716	1.1316	18.4000	213	0.0902	0.1233	0.5099	46.5000
f2o10m	0.2584	0.1711	0.8801	23.2000	214	0.1906	0.1198	0.7076	50
f2y01m	0.4297	0.2505	1.3612	15.5000	221	0.2310	0.0988	0.4771	35.3000
f2y02m	0.5237	0.2477	0.9863	23.2000	222	0.1022	0.1557	0.6405	30
f2y03m	0.5182	0.2448	1.4546	15.5000	223	0.1322	0.1730	0.5940	46.5000
f2y06m	0.1144	0.1471	0.6274	30	228	0.2221	0.1288	0.5590	35.8000
f2y07m	0.2789	0.1803	0.9732	21.3000	230	0.0292	0.0654	0.4499	45.5000
f2y08m	0.2898	0.1679	0.8171	25.2000	232	0.0223	0.0580	0.4796	45.5000
f2y09m	0.2507	0.1620	0.7709	26.1000	233	0.2602	0.2651	0.7185	31
f2y10m	0.2082	0.1060	2.0152	11.6000	234	0.2442	0.1433	0.5171	41.6000
Database	Total beats	Total TP (beats)	Total FP (beats)	Total FN (beats)		Total failed detection (%)	Sensitivity (%)	Specificity (%)	
FANTASIA	4000	3989	11	09		0.50	99.72	99.77	
MITDB	4000	3971	17	13		0.75	99.67	98.18	
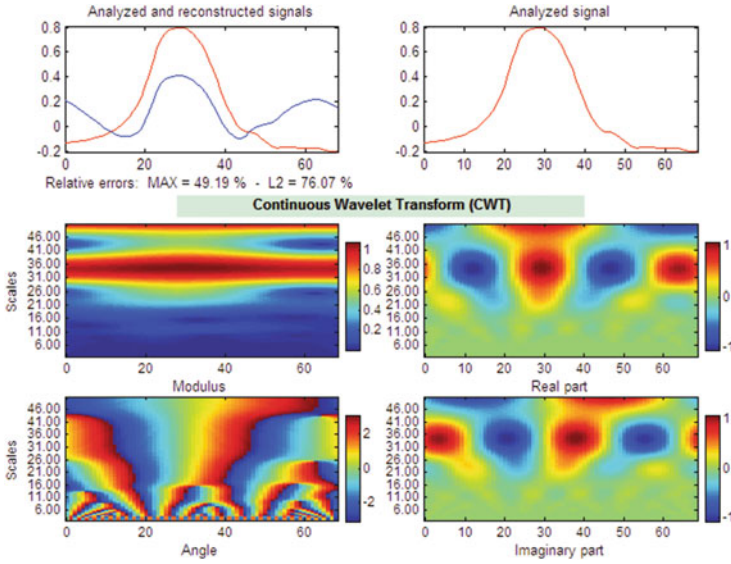
Fig. 4 CWTFT of QRS complex for MITDB data #203
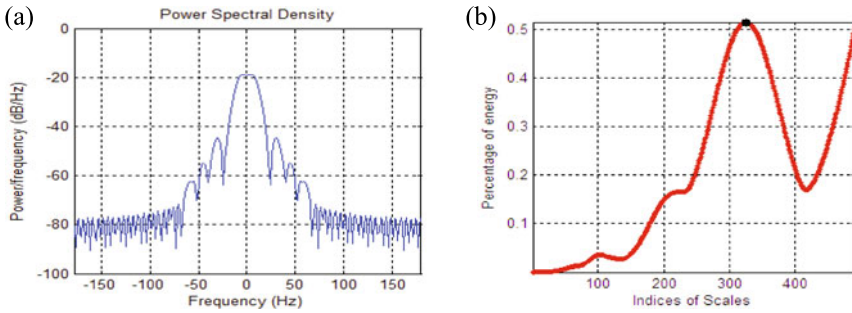
Fig. 5 **a** Max power, **b** energy estimation

$$C(x, y; g(t), \psi(t)) = \frac{1}{2\Pi} \int_{\infty}^{-\infty} \hat{g}(\omega)\sqrt{x}\hat{\psi} * (x\omega)e^{j\omega y}d\omega \tag{9}$$

Using period-gram (PSD) and CWTFT power energy distribution (Fig. 4) of each signal of both the databases is computed in Table 1 where the power (Fig. 5a) (frequency and time domain) and energy (maximum percentage (Fig. 5b) and scale maximum) are tabulated.

Table 2 Analysis of variance of QRS complex for frequency domain power of FANTASIA and MITDB

Source	DF	Adj SS	Adj MS	F-value	p-value
Factor	1	0.7491	0.74910	21.91	0.000
Error	58	1.9827	0.03418		
Total	59	2.7318			

Table 3 Analysis of variance of QRS complex for maximum energy percentage of FANTASIA and MITDB

Source	DF	Adj SS	Adj MS	F-value	p-value
Factor	1	3.105	3.10451	35.97	0.000
Error	58	5.006	0.08630		
Total	59	8.110			

3.2 One-Way ANOVA

One-way ANOVA is implemented to determine whether there are statistically significant differences among the means of several populations (i.e., power and energy of healthy and arrhythmic QRS complex).

Table 2 shows that there is a significant difference (i.e., p-value = 0.000) in between the frequency domain power of QRS complex of healthy and arrhythmic signals. In the frequency domain power of FANTASIA and MITDB, the intervals (at 95% CI) for the means of these two data sets do not overlap which suggests that the population means for these levels are different.

Table 3 shows that there is a significant difference (i.e., p-value = 0.000) in between the maximum energy percentage of QRS complex of healthy and arrhythmic signals. In between the maximum energy percentage of FANTASIA and MITDB, the intervals (at 95% CI) for the means of these two data sets do not overlap which suggests that the population means for these levels are different. The same results are obtained with the time domain power and maximum scale energy parameters.

4 Conclusions

The PSD and CWTFT compute the average power and distribution of energy power at different scales for normal and arrhythmic QRS complex. The mean energy power spectral parameter of QRS complexes of healthy subjects (FANTASIA) and arrhythmic (MITDB) signals proves to be lying in a different domain, which are validated with one-way ANOVA (p-value < 0.05). The proposed methodology also considers the critical role of a detection method for QRS complex where the failed detection percentage is found to be $F_d(\%) = 0$ for FANTASIA and $F_d(\%) = 1.85\%$ for MITDB databases. Moreover, the high-frequency noise components are eliminated using SWT (i.e., 50 Hz power line interference embedded with FANTASIA). This proposed approach for analyzing QRS complex certainly formulates a new way

of considering research possibility to detect arrhythmia segregating from healthy signals. The further research lies in testing this methodology with different erratic signals and extending the possibility of finding changes in ischemic heart disease (IHD) by analyzing QRS, ST-segment, and T wave.

References

1. Kohler BU, Henning C, Orglmeister R (2002) The principles of software QRS detection. IEEE Eng Med Biol 21(1):42–57
2. Rangayyan RM (2002) Biomedical signal analysis. IEEE press, New York
3. Yun-Chi Y, Wen-June W (2008) QRS complexes detection for ECG signal: the difference operation method. Comput Methods Progr Biomed 91:245–254
4. Li C, Zheng C, Tai C (1995) Detection of ECG characteristic points using wavelet transforms. IEEE Trans Biomed Eng 42(4):21–28
5. Abibullaev B, Don SH (2011) A new QRS detection method using wavelets and artificial neural-networks. Springer J Med Syst 35:683–691
6. Ruchita G, Sharma AK (2010) Detection of QRS complex of ECG recording based on wavelet transform using Matlab. Int J Eng Sci 2(7):3038–3034
7. Dinh AN, Kumar DK, Pah ND, Burton P (2002) Wavelet for QRS detection. In: Engineering in medicine and biology society, 23rd conference of IEEE, pp 7803–7211
8. Zidelmal Z, Amirou A, Adnane M, Belouchrani A (2012) QRS detection using wavelet coefficients. Comput Method Progr Biomed 107(3):490–496
9. Chen SW, Chen CH, Chan HL (2006) A real-time QRS method based on moving-averaging incorporating with wavelet denoising. Comput Method Progr Biomed 82(3):187–195
10. Mehta SS, Ligayat NS (2007) Comparative study of QRS detection in single lead and 12-lead ECG based on entropy and combined entropy criteria using support vector machine. J Theor Appl Inform Technol 3(2):8–18
11. Poli R, Cagnoni S, Valli G (1995) Genetic design of optimum linear and nonlinear QRS detectors. IEEE Trans Biomed Eng 42(11):1137–1141
12. Coast DA, Stern RM, Cano GG, Briller SA (1990) An approach tocardiac arrhythmia analysis using hidden Markov models. IEEE Trans Biomed Eng 37(9):826–836
13. Saini I, Singh D, Khosla A (2013) QRS detection using K-nearest neighbour algorithm (KNN) and evaluation on standard ECG databases. J Adv Res 4(4):331–344
14. Meyer C, Gavela JF, Harris M (2006) Combining algorithms in automatic detection of QRS complexes in ECG signals. IEEE Trans Inform Technol B 10(3):468–475
15. Moraes JCTB, Seixas M, Vilani FN, Costa EV (2002) A QRS complex detection algorithm using electrocardiogram leads. Comput Cardiol 29:205–208
16. Pan J, Tompkins WJ (1985) A real-time QRS detection algorithm. IEEE Trans Biomed Eng BME 32(3):230–236
17. Chia-Hung L, Yi-Chun D (2010) Fractal QRS-complexes pattern recognition for imperative cardiac arrhythmias. Dig Signal Process 20:1274–1285
18. Lin C-C (2010) Analysis of unpredictable components within QRS complex using a finite-impulse-response prediction model for the diagnosis of patients with ventricular tachycardia. Comput Biol Med 40:643–649
19. Soroor B, Jafarnia DN (2011) Detection of QRS complexes in the ECG signal using multiresolution wavelet and thresholding method. Comput Cardiol 38:805–808
20. Sasan Y, Jean-Marc V (2014) Adaptive mathematical morphology for QRS fiducial points detection in the ECG. Comput Cardiol 41:725–728
21. Atiyeh K, Reza HM (2014) Real-time electrocardiogram P-QRS-T detection—delineation algorithm based on quality-supported analysis of characteristic templates. Comput Biol Med 52:153–165

22. Iyengar N, Peng C-K, Morin R, Goldberger AL, Lipsitz LA (1996) Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. Am J Physiol 271:1078–1084
23. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
24. Manpreet K, Seema SB (2011) Comparison of different approaches for removal of baseline wander from ECG signal. In: 2nd international conference and workshop on emerging trends in technology (ICWET) proceedings published by international journal of computer applications © (IJCA)
25. Mallat A (1998) Wavelet tour of signal processing. Academic Press, San Diego, USA

Radiometric Calibration of RISAT-1 SAR Image Using Integral Method

T. R. Manjula, T. N. Praveen, M. Vinod Raju, B. D. Vishwanath, P. Meghana and G. Raju

Abstract Synthetic aperture radar plays a key role in remote sensing applications ranging from mapping earth's surface to disaster monitoring applications. Accuracy or correctness of SAR data or images is commendable in all these applications. Component aging and gain fluctuations in receiver and transmitter and atmospheric factors cause deviations in SAR performance which necessitates frequent SAR calibration. Validating the measured value against the standard or reference value is known as SAR calibration. SAR calibration is also essential as to compare the measured data with similar other sensors. This paper brings out several aspects of RISAT-1 FRS mode SAR image calibration using standard point targets such as corner reflectors. Integral method of calibration is demonstrated on RISAT-1 SAR sample image and the calibration factor is determined and validated.

Keywords Synthetic aperture radar · Calibration · Corner reflector · Integral method · Radar cross section · Point targets

1 Introduction

Synthetic aperture radar (SAR), a remote sensing sensor used for a multitude of applications ranging from biomass estimation, change detection, military surveillance to disaster control applications. Radar imaging satellite 1 (RISAT-1) is microwave remote sensing satellite [1] carrying SAR sensor operates in C band (5.35 GHz) enables mapping of Earth's surface providing global coverage. Capability of imaging in co, cross, and circular polarizations such as HH, VV, HV, VH, and circular polarization has extended the range of application to agricultural field for monitoring crop growth and study of soil moisture [2].

Data acquired by sensors are influenced by a number of factors such as scattering, low illumination, aging, and other environmental factors which will alter the reflected

T. R. Manjula (✉) · T. N. Praveen · M. Vinod Raju · B. D. Vishwanath · P. Meghana · G. Raju
Electronics and Communication Engineering, SET, Jain University, Bengaluru, India
e-mail: bhindyaa@gmail.com

intensity. Deviations in SAR specifications from designed specifications will result in data or image not being a true representation of the terrain mapped and requires validating the measured value against the standard or reference. Moreover, SAR, apart from being a sensor is a measurement instrument which records the terrain reflectivity hence like any other instrument, SAR calibration is essential to ensure consistency in the derived backscatter values.

SAR calibration is also a technique to evaluate the performance of a chosen space-borne SAR (RISAT-1) in terms of accuracy and quality of measurement [3]. SAR images acquired from various sensors, under different temporal or spatial is comparative only after calibration. Calibration provide SAR imagery in which the pixel values are directly related to the backscatter coefficients of the illuminated area and thus significant in providing reliable data products for scientific and commercial communities.

SAR images are subjected to two types of calibration: geometric and radiometric. In radiometric calibration, the magnitude of each pixel is validated against the standard value; corrections are estimated and applied to ensure pixel intensity represents the true terrain reflectivity. Geometric calibration, on the other hand, ensures SAR images are geo-located with respect to reference coordinate system.

Two widely used calibration techniques are based on using transponders [4] and calibration targets such as corner reflectors (CR). The former method though provides accurate calibration, is complex in procedure and requires maintenance, corner reflector, on the other hand, offer simplest and cost-effective calibration strategy that provides comparatively accurate calibration if the surrounding interference is accounted. Naturally distributed homogeneous targets such as Amazon forest with exactly known backscatter coefficients is used as distributed target calibration, known as vicarious calibration.
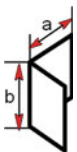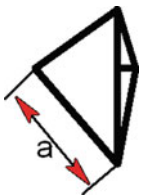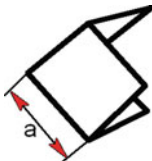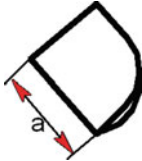
In this paper, radiometric calibration of RISAT-1, FRS mode SAR image is demonstrated. Fine resolution mode (FRS), single look complex image (SLC) SAR image of Ahmadabad region is selected. The SAR image with point targets representing the corner reflectors is taken as a sample image to determine the calibration factor. The determined calibration factor is verified.

The paper is organized as follows: A brief discussion on commonly used corner reflectors for calibration, their merits and demerits are discussed giving the brief summary on the selection of CR for calibration. Following next is the brief discussion on calibration field and alignment of a corner reflector. The methodology of calibration based on integral method is discussed next followed by calibration results and discussion.

2 Background Study

Corner reflector is basically a reflector which reflects signal back towards the source with high reflectivity and this property is utilized in SAR calibration. It is made up of two or three mutually perpendicular surfaces, the radar cross section (RCS) of

Table 1 Types of SAR calibration targets

Calibration target	RCS	Merits and demerits
 Diplane	$\sigma = \frac{8\pi a^2 b^2}{\lambda}$	• Supports cross polarization (HV and VH) and Like polarizations (HH and VV) • High RCS • Low stability not suitable for adverse environmental conditions
 Triangular trihedral	$\sigma = \frac{4\pi a^4}{3\lambda^2}$	• Supports only like polarizations (HH and VV) • High RCS • Less sensitive to deviations in incidence angle • Stable suitable for adverse environmental conditions
 Square trihedral	$\sigma = \frac{12\pi a^4}{\lambda^2}$	• Supports only like polarizations (HH and VV) • Highest RCS • Sensitive to deviations in incidence angle • Low stability not suitable for adverse environmental conditions
 Circular trihedral	$\sigma = \frac{0.507\pi^3 a^4}{\lambda^2}$	• Supports only like polarizations (HH and VV) • High RCS • Sensitive to deviations in incidence angle • Low stability not suitable for adverse drastic environmental conditions

which is theoretically computed based on side dimensions and signal wavelength. Table 1 shows different types of corner reflectors and their RCS commonly used in SAR calibration.

Corner reflectors are widely used as calibration targets because of their high reflectivity and known RCS. The CR must have reflectivity much higher than the surrounding scatterers (clutter) and hence reflector deployment should be done in low clutter area. The side dimension of the CR is ideally selected larger than 10 times the signal wavelength, and yet not too large to saturate [5] the SAR. For example, for C band SAR calibration at a wavelength of 56 cm, the corner reflectors of side dimension between 56 and 100 cm are selected.

From the table, square trihedral offers highest RCS and in the second position is the circular trihedral. Further, constructing square trihedral is simpler than circular one and is also intolerant to deviations in incidence angles. From stability point of view, triangular trihedral is mechanically more stable than square one. Furthermore triangular trihedral offers an advantage of maintaining RCS over wide deviation in incidence angle. Since only square dihedral CR is sensitive to polarization and hence suitable for cross polarization calibration (HV/VH). To facilitate the comparison of the performance of square and triangular trihedral CR in calibration, a SAR sample image consisting of point targets representing both square and triangular trihedral corner reflector is chosen for this presentation.

2.1 Calibration Field and Alignment

Among the various factors that significantly impact the calibration accuracy of SAR images is the surrounding clutter [6]. The clutter can arise from innumerable objects on a ground surface such as building, metal fence, rocks, trees, etc. Hence to achieve good calibration accuracy, corner reflectors are deployed in the area of low interference from the surroundings. The clutter level is determined by computing signal to clutter ratio (SCR) from the impulse response of the point target [5]. One of the parameters that will also influence the calibration accuracy is the misalignment of CR. CR should be installed on a flat surface area free of elevations or slopes and is aligned in the LOS of the satellite. Alignment of CR involves azimuth and elevation angle setup of CR in the direction of look angle of the satellite. Any deviation from LOS direction results in a decline in the RCS value of the point targets in the acquired SAR image.

Elevation angle of CR $= 90° -$ (incidence angle $-$ CR default angle (35.26°))

CR default angle is the angle between the ground and the line passing through the vertex. Azimuth angle = Satellite azimuth angle $\pm 90°$ (negative for right look direction and positive for left look direction).

2.2 Radar Cross Section

The amount of energy reflected by the target towards the source when it is illuminated by EM energy is referred to as RCS or backscattering radar cross section. Mathematically it is defined as the ratio of the power density of the reflected signal at the source to a power density of the incident signal on the target taking into account the range distance R. RCS is given by

$$\text{RCS} = \sigma = 4\pi R^2 \left(\frac{S_r}{S_i}\right) m^2 \tag{1}$$

where S_r and S_i are the density of the power reflected and incident of the target respectively.

The amount of backscattering (σ) depends on a multitude of parameters of the target such as shape, dielectric properties, orientation, and roughness. These properties are likely to vary for different observation angles, frequencies, and polarizations. The normalized RCS is given by $\sigma_0 = \sigma/A$ which is referred to sigma naught which implies scattering cross section per unit area of the surface and is a measure of the property of the target.

3 Methodology of SAR Calibration

Corner reflector of the required RCS are designed and deployed in calibration site, aligned in the LOS of the satellite, are imaged as point targets in the SAR image. These targets appear as point sources or targets in the SAR image. Impulse response function (IRF) of the point target is characterized as focused energy distribution in the azimuth and elevation directions of SAR image. Impulse response measurement is useful in determining the calibration factor by comparing against theoretical RCS of a point target.

SAR calibration procedure involves deployment of CR of known RCS and aligning it in the direction of the satellite. SAR acquisitions over the deployed area result in an image which consists of spectacular bright points representing the CR response, known as point targets. Point target's response in a SAR image is a finite energy signal. The received signal is given by the radar equation

$$P_r = \frac{P_t G \sigma \lambda^2}{(4\pi)^3 R^4} \tag{2}$$

σ corresponds to the RCS of the target. From the received signal, P_r, the reflected intensity of the target that is σ (RCS) is extracted. By comparing the energy of the point targets against reference one (σ_{ref}) calibration factor "k" is determined.

There are two methods to determine the backscattered intensity of the point targets: peak method [7, 8] and integral method. The integral method is appealing as all parameters can be derived from the SAR image.

4 Integral Method of Calibration

Integral method [3] of calibration involves estimating the intensity backscattered from a point target by analyzing IRF and integrating energy over an area around the point target. This is referred to the total backscattering intensity of the area which includes the contribution from point target as well as the background clutter as shown in the Fig. 1.

The total intensity of the target area is given by
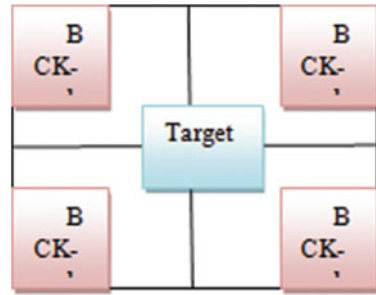
$$I_T = \sum_{i=1}^{p} \sum_{j=1}^{q} i_T(i, j) \tag{3}$$

I_t is the intensity of the pixel (i, j) in the target area. $p \times q$ are the pixels in range and azimuth direction respectively.

The intensity backscattered from the background called as clutter is given by

$$I_B = \frac{1}{N_B} \sum_{i=1}^{m} \sum_{j=1}^{n} i_B(i, j) \tag{4}$$

I_b is the intensity of the pixel (i, j) in the background area. N_B is the total number of background pixels.

The intensity backscattered from the point target itself is estimated as

$$I_p = I_T - I_B \tag{5}$$

Based on the type of CR and its dimension. The theoretical RCS is computed is denoted as (σ_{ref}). For example for trihedral triangular CR, RCS is given in Table 1.

The calibration constant k at the local incidence angle θ_i is estimated as [9]

$$k = \frac{I_p \delta_a \delta_r}{\sigma_{ref} \sin(\theta_i)} \tag{6}$$

where I_p is the intensity backscattered from the point target δ_a and δ_r, the pixel spacing along azimuth and range respectively.

Energy is extracted from the Impulse response function (IRF) of the point target directly or from the interpolated IRF. For small dimensions of the image, the IRF obtained is too skewed to accurately estimate the intensity of the point target and would affect the calibration accuracy. IRF is interpolated to obtain a smooth function. The interpolation is performed in the frequency domain using FFT routines in both range and azimuth, for interpolation factor of 16. Point target energy is estimated by integrating over an area of two times the resolution cell to obtain the main lobe energy I_p. The calibration factor k for interpolated IRF is given by

Fig. 2 Google earth image
of Ahmadabad
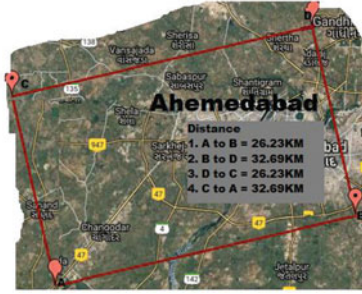
Fig. 3 Corner reflector
deployment area

$$K = \frac{I_p \delta_a \delta_r}{\sigma_{reff_{\mathrm{int}}^2} \sin(\theta_i)} \tag{7}$$

where f_{int} is interpolation factor.

5 Study Area and Data Set

Figure 2 shows Google Earth image of Ahmadabad region and Fig. 3 illustrates the
deployment area where triangular and square trihedral corner reflector was deployed
on February 2, 2013 prior to satellite pass. The products used for calibration determi-
nation are level-1 SLC image in FRS-1 mode of linear (HH) and circular polarization
(RH and RV) (Table 2).

6 Results and Discussion

Algorithm

- Extraction of 16×16 pixel area of point targets
- Conversion of amplitude image to an intensity image

Table 2 RISAT-1 SAR data set

DOP	Imaging node	Beam no/look direction	Polarization
2nd Feb 2013	Ascending	30/R	HH
1st Oct 2013	Descending	40/R	RV + RH

R Right, *H* Horizontal, *V* Vertical
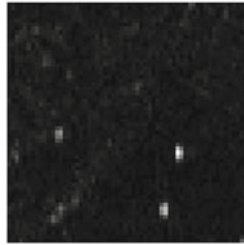
Fig. 4 60 × 60 pixel area surrounding the 3 point targets and their locations
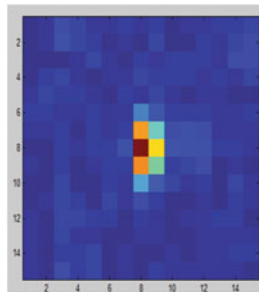
Fig. 5 Spatial image of point target of 16 × 16 pixel area

- Background clutter extraction
- Clutter corrected image
- Image interpolation
- Computation of Calibration factor

Figure 4 shows the three-point targets corresponding to three corner reflectors two of which are square trihedral CR of side dimension 60 cm, the other one is triangular CR with 90 cm dimension. A 16 × 16 pixel, area of sub-image centering the point target is extracted and is converted into intensity image as shown in Fig. 5, the corresponding 3D impulse response function in HH polarization is shown in the Fig. 6.

The background noise or interference is estimated by integrating intensity in the four corner windows of size 4 × 4 as shown in Fig. 7. Average background intensity
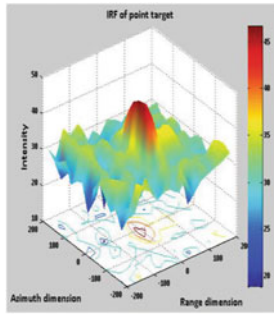
Fig. 6 3D point target response of the reflectivity of CR
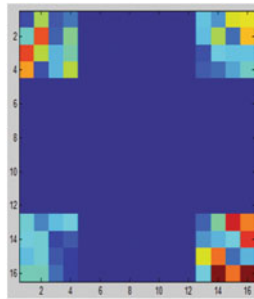
Fig. 7 Back ground clutter extraction
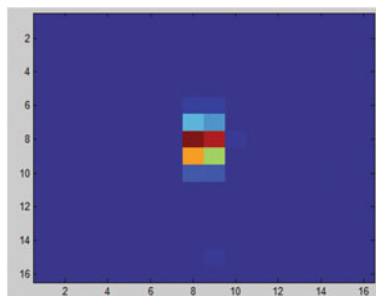
Fig. 8 Clutter corrected point target

is estimated as given by Eq. 2. Clutter corrected intensity image is generated by eliminating the average background interference from the target response as shown in Fig. 8 as given by Eq. 5. The energy of the point target is estimated by integrating the intensity of IRF, two times the resolution cell around the peak of the point target.

Table 3 shows the calibration result obtained for all the three point targets, corresponding to three corner reflectors (TT: Triangular trihedral, ST: Square trihedral).

Table 3 Calibration targets

Data products	Corner reflector type	Polarization	Calibration factor
Level 1 SLC FRS-1 RISAT-1 SAR image	TT	HH	78.1
	ST	HH	77.4
		HV[a]	–
	TT	RH	74.48
	ST		75.3283
	TT	RV	71.7187
	ST		72.344

[a]Point target analysis for HV is not carried as the type of corner reflectors (trihedral) is not effective for cross polarization (HV)

7 Conclusion

SLC of level 1 in FRS-1 mode of RISAT-1 SAR image is evaluated for calibration and calibration factor is determined for HH, RH, and RV polarization using the integral method. The results are verified with the header file. Variation in calibration factor in the range of 0.3–0.6 is observed between and square and triangular corner reflector and between two triangular reflectors and is attributed to the difference in spatial locations of the CRs and clutter levels.

References

1. Tapan Misra SS, Rana NM, Desai DB, Dave1 Rajeevjyoti1 RK, Arora1 CVN, Rao1 BV, Bakori1 R, Neelakantan, Vachchani JG (2013) Synthetic aperture radar payload on-board RISAT-1: configuration, technology and performance. Current Sci 104:446–461
2. Ganesan PG, Jagdhuber T (2016) Soil moisture estimation using hybrid polarimetric SAR data of RISAT-1. IEEE Trans Geosci Remote Sens 54:2033–2049
3. Gupta M, Sharma A, Karthikeyan B (2014) SAR image quality assessment of RISAT-1 SAR using trihedral corner reflectors in different beams. J Geomat 8:130–139
4. Touzi R, Hawkins RK, Cote S (2013) High-precision assessment and calibration of polarimetric RADARSAT-2 SAR using transponder measurements. IEEE Trans Geosci Remote Sens 51:487–503
5. Lacomme P, Marchais J-C, Hardange J-P, Normant E (2001) Air and spaceborne radar systems: an introduction, 1st edn. William Andrew Publishing, LLC, ISBN-13:978-1891121135
6. Manjula TR Calibration site selection. Technical report: JU/ISRO/RESPOND/2017/TR-03, Jain University Bangalore
7. Ulander LMH (1991) Accuracy of using point targets for SAR calibration. IEEE Trans Aerosp Electron Syst 27:139–148

8. Praveen TN, Vinod Raju M, Vishwanath BD, Meghana P, Manjula TR, Raju G (2018) Absolute radiometric calibration of SAR image using peak method. In: Submitted to 3rd IEEE international conference on recent trends in electronics, information and communication technology, (RTEICT-2018)
9. Mishra MD, Patel P, Srivastava HS et al (2015) Absolute radiometric calibration of SAR data using standard targets. Int J Remote Sens Geosci (IJRSG) 4:28–32

Optimization of Microstrip Patch Array Antenna for Gain Enhancement

Rohan Singh, Arun Kumar Singh, Rabindranath Bera and Bansibadan Maji

Abstract This paper aims on designing a compact, cheap, low profile patch antenna for X-Band (i.e., 8–12 GHz). The substrate of the antenna is chosen to be FR4 epoxy substrate which has a thickness of 1.6 mm and dielectric constant of 4.4. The design is simulated of HFSS software by which return loss, gain, radiation pattern, and frequency versus gain plot are computed.

Keywords Micro-strip patch antenna · Return loss · Gain · Dielectrics

1 Introduction

An antenna, generally, is a metallized tool used for transmitting and/or receiving radio waves. A micro-strip patch antenna has a ground plane on one side of the dielectric substrate and it has an emitting patch on the other side. Micro-strip antennas have gained enormous significance recently because of their small size, light weight, ease of fabrication, and integration. The demand for wireless communication is system has undergone an exponential increase in the past few years and is continuing to grow now also.

R. Singh · A. K. Singh (✉) · R. Bera · B. Maji
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India
e-mail: arunsingh.smit@gmail.com

R. Bera
e-mail: rbera@smit.smu.edu.in; rbera50@gmail.com

B. Maji
e-mail: bansibadan.maji@ece.nitdgp.ac.in

B. Maji
Department of Electronics and Communication Engineering, National Institute of Technology, Duragpur, Duragpur, West Bengal, India

Fractal based Micro-strip Patch Antennas (MPAs) are widely used in a wireless system which demands higher bandwidth, i.e., ultra wideband applications. MPAs built on fractal type is the current trend of research these days [1].

The definition of fractal is that some of its parts have the identical shape as the whole object but in a dissimilar scale or size [2, 3]. The main benefit of this antenna is that these antennas are being used for multi-band frequency operations [4, 5].

Wearable biomedical systems which use a MPA also has an advantage from wireless links. Since the wearable system is connected to a network, a runner can constantly monitor his/her heart rate, pulse, his/her current speed, and even the current position also using the tracking service like GPS [6].

But unfortunately, there are some major drawbacks also such as less gain, efficiency [7]. These factors limit the interpretation of MPAs. Umbers of methods are researched and implemented in order to overcome this drawback. Few of the techniques are a change of feeding point, impedance matching, creating slots in the antenna, changing the height of the substrate, and many more [8–11].

A high gain patch with about 36 dB gain can also replace the DTH antenna.

A design for a single feed, single layer, low cost and compact micro-strip patch antenna for multi-band wireless applications. The antenna should find use for C-band, X-band, Wi-MAX, and ISM band radio, high-speed wireless LANs and RADAR applications. Since it resonates at five frequencies it provides multi-band operation with return loss less than −10 dB [12].

H-shape meandered micro-strip patch antenna is for multi-band applications. The design has a maundered patch structure, which provides multi-band, low profile, high gain, and compact antenna element. The patch has a neat and compact dimension. It has applications in X-band, C-band, S-band, and L-band communications [13].

Dual-band micro-strip patch antenna was placed on a FR-4 substrate to solve the challenges of multi-frequency applications. This antenna has used FR-4 substrate having dielectric constant of 4.54 and the thickness of the substrate is 1.6 mm. This antenna gave dual-band operation by resonating at 6.07 GHz and 9.15 GHz, respectively, and it can be used in applications related to radio navigation and satellite communication at low power. The return loss of antenna is lower than −10 dB [14].

A multi-band fractal antenna is I-shaped: Three iterations were applied one after the other on a square-shaped patch that resulted in fractal antenna that resonated at 4.7 GHz, 6.5 GHz, 7.7 GHz, and 8.5 GHz, respectively. This antenna has applications in secure communication and defense [15].

2 Antenna Parameters

To design a micro-strip patch antenna, the calculation of width (W) and length (L) of patch, as well as the substrate, is required. Following parameters such as resonant frequency (f_0), dielectric constant (ε_r), and height (h) are considered for calculations:

A patch whose dimensions are calculated according to Eqs. 1–7 as described in [16–19]. Practical width is given by

$$W = \frac{c}{2 f_0 \sqrt{\frac{\epsilon_r + 1}{2}}} \tag{1}$$

Mathematics for the Effective Dielectric Constant. This is based on the dielectric constant of the dielectric, width, and height of the patch antenna.

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12 \frac{h}{w} \right]^{-\frac{1}{2}} \tag{2}$$

Mathematics for the effective length

$$L_{eff} = \frac{c}{2 f_0 \sqrt{\epsilon_{eff}}} \tag{3}$$

Mathematics for the length extension ΔL

$$\Delta L = 0.412 h \frac{\left(\epsilon_{eff} + 0.3 \right) \left(\frac{W}{h} + 0.264 \right)}{\left(\epsilon_{eff} - 0.258 \right) \left(\frac{W}{h} + 0.8 \right)} \tag{4}$$

Mathematics for actual length of the patch

$$L = L_{\text{eff}} - \Delta L \tag{5}$$

Mathematics for the length of substrate

$$L_s = L + 6h \tag{6}$$

Mathematics for the width of the substrate

$$W_s = W + 6h \tag{7}$$

2.1 Antenna Design

The geometry of proposed antenna is shown in Fig. 1. FR4 epoxy is the substrate that has been used it has a thickness of 1.6 mm and dielectric constant (ϵ_r) of 4.4.

Fig. 1 2 × 2 patch antenna
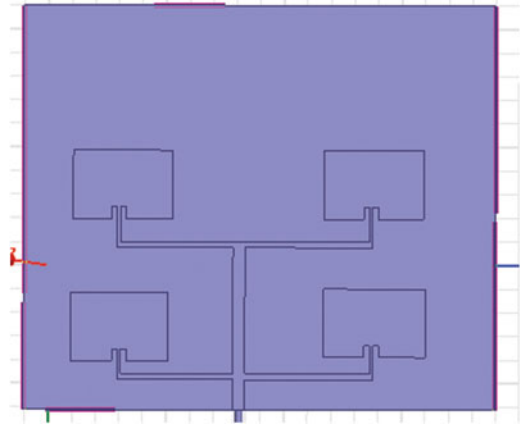
Fig. 2 Return loss from 2 × 2 patch antenna
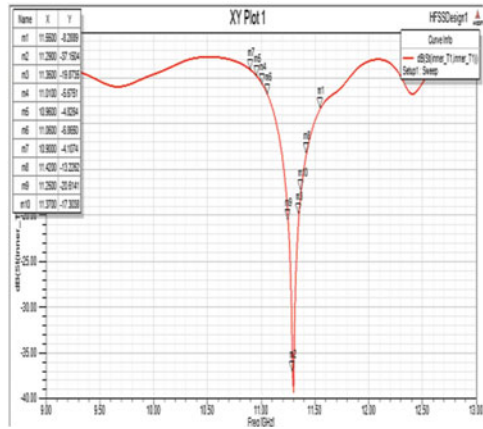
2.2 2 × 2 Patch Antenna and Its Return Loss

A 2 × 2 patch antenna is designed where the length of the patches, L = 8.3 mm (Patch 1), 8.6 mm (Patch 2), 8.4 mm (Patch 3), 8.5 mm (Patch 4), and width of the patch, W = 5.8 mm.

The measurements of substrate was $40 \times 34.25 \times 1.6$ mm^3.

Figure 2 shows −39 dB return loss at 11.3 GHz (Fig. 3).

The gain obtained is 8.8130 dB at 10.41 GHz and 6.0452 at 11 GHz for 2 × 2 patch antenna as shown in Fig. 4.

Fig. 3 Radiation pattern of
2 × 2 patch antenna
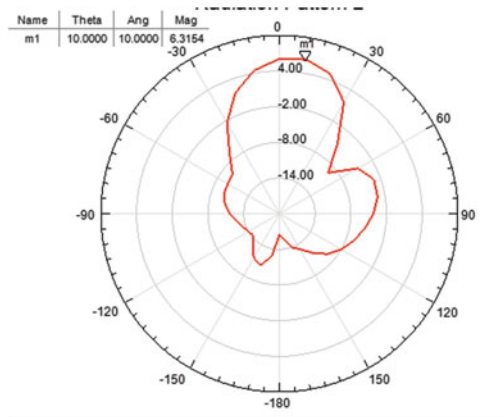
Fig. 4 Frequency versus
gain plot for 2 × 2 patch
antenna
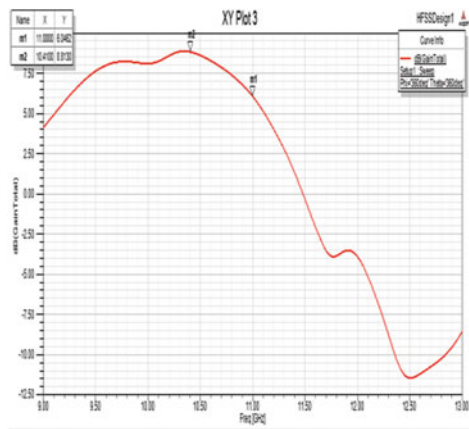
2.3   2 × 2 Patch Antenna with Fractal

A 2 × 2 patch antenna width fractal is designed where the length of the patches, L =
8.3 mm (Patch 1), 8.8 mm (Patch 2), 8.6 mm (Patch 3), 8.7 mm (Patch 4) and width
of the patch, W = 5.8 mm.

The measurement of the substrate was 40 × 33.25 × 1.6 mm^3 (Fig. 5).

Figure 6 shows −65.01 dB return loss at 10.77 GHz (Fig. 7).

The gain obtained is 7.6931 dB at 11 GHz for 2 × 2 patch antenna as shown in
Fig. 8.

Fig. 5 2 × 2 patch antenna with fractals
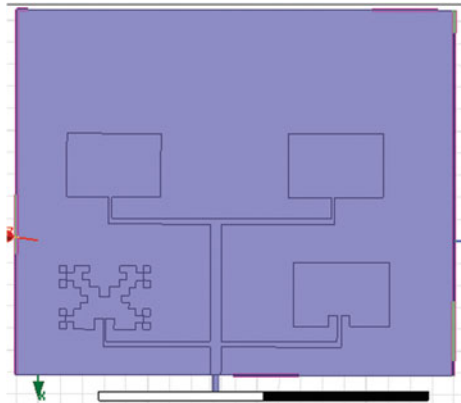
Fig. 6 Return loss from 2 × 2 patch antenna with fractal
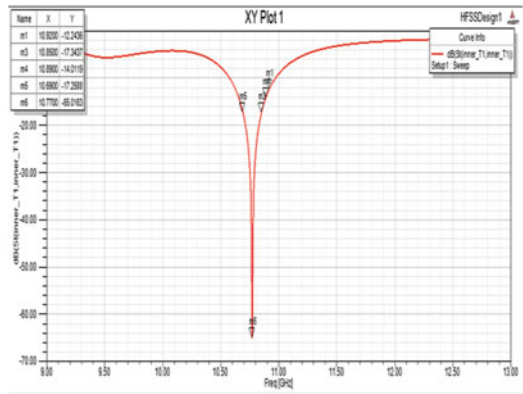
Fig. 7 Radiation pattern of 2 × 2 patch antenna with fractal
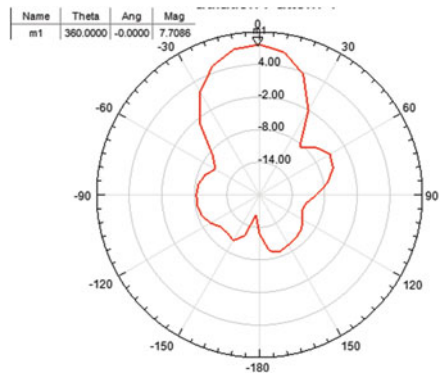
Fig. 8 Frequency versus gain plot for 2 × 2 patch antenna with fractal
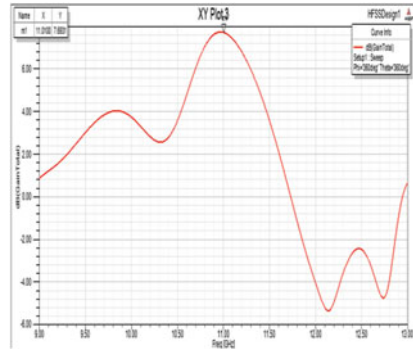
Fig. 9 4 × 4 patch antenna
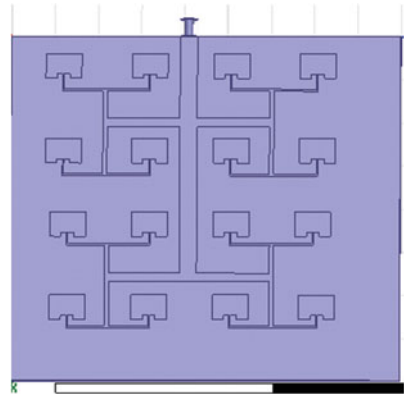
2.4 4 × 4 Patch Antenna

A 4 × 4 patch antenna is designed where length of the patches, L = 8.3 mm and width of the patches, W = 5.75 mm

The measurement of the substrate was 90 × 80 × 1.6 mm³ (Figs. 9, 10, 11, 12 and 13).

Simulations were performed using HFSS™ (High-Frequency Structure Simulator). It is a simulation tool for three-dimensional full-wave electromagnetic field simulation that is applicable in industries. HFSS provides H and E-fields, S-parameters, far and near emitted field results and currents. It combines simulations, solid modeling, visualization, and automation.

Fig. 10 Return loss from 4
× 4 patch antenna
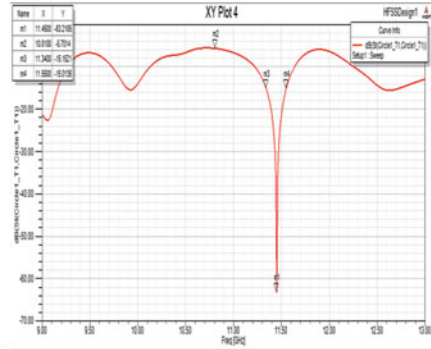
Fig. 11 Radiation pattern of
4 × 4 patch
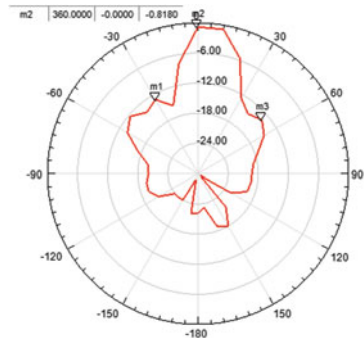
Fig. 12 Frequency versus
gain plot for 4 × 4 patch
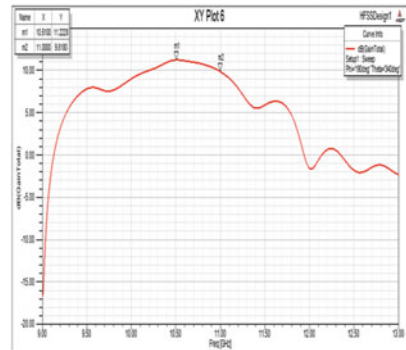
3   Conclusions

Thus a 2 × 2, 2 × 2 with Fractals and a 4 × 4 patches were designed. The return
loss seen in the 2 × 2 patch was −39 dB at 11.3 GHz with a gain of 8.8130 dB.
The return loss seen in the 2 × 2 patch with fractals was −65.01 dB at 10.77 GHz
with a gain of 7.6 dB whereas in the case of 4 × 4 the return loss was −63.21 dB at
11.45 GHz and with a gain of 11.22 dB.

Fig. 13 Antenna pattern
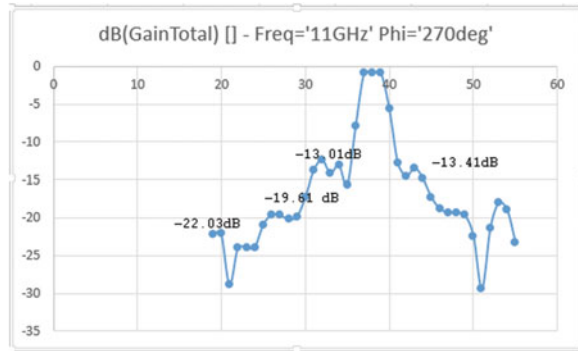
References

1. Meshram VP, Sangara P, Wanjari P, Chintawar I (2013) Design and fabrication of wideband fractal antenna for commercial applications. In: International conference on machine intelligence and research advancement, pp 150–154
2. Peitgen HO, Jurgens H, Saupe D (1990) Chaos and fractals. Springer, New York
3. Jones DE, Reeve H, Saupe D (1990) Fractals and chaos, Crilly AJ, Earnshaw RA, Jones H (eds), Springer, New York
4. Vinoy KJ (2002) Fractal shaped antenna elements for wide- and multiband wireless applications. A thesis in engineering science and mechanics, The Pennsylvania State University, The Graduate School College of Engineering
5. Yadava RL, Ram M, Das S (2010) Multiband triangular fractal antenna for mobile communications. Int J Eng Sci Technol 2(11):6335–6348
6. Jovanov E, Milenkovic A, Otto C, de Groen P (2005) A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. J Neuro Eng Rehabil 2(6)
7. Liu Y, Si L-M, Wei M, Yan P, Yang P, Lu H, Zheng C, Yuan Y, Mou J, Lv X, Sun H (2012) Some recent developments of microstrip antenna. Int J Antennas Propag (Article ID 428284), pp 1–10
8. Kumar G, Ray KP (2003) Broadband microstrip antennas. Artech House antennas and propagation library
9. Pozar DM, Kaufman B (1987) Increasing the bandwidth of a microstrip antenna by proximity coupling. Electronic Lett 23(8):368–369
10. Pues HF, Van de Capelle AR (1989) Impedance-matching technique for increasing the bandwidth of microstrip antennas. IEEE Trans Antennas Propag 37(11):1345–1354
11. Jadon AS, Sharma J, Prajapat A, Bhadauria A (2013) Coplanar rectangular patch antenna for X band applications using inset fed technique. Int J Electron Commun Eng Technol 95–102
12. Singh C, Gangwar RPS (2011) Design and analysis of a compact low cost patch antenna for different wireless operations. In: International conference on emerging trends in networks and computer communications (ETNCC), pp 18–22
13. Patel H, Patel SK, Bhalani J, Kosta Y (2012) Design of meandered H-shaped square microstrip patch antenna. In: Nirma University international conference on engineering (NUiCONE), pp 1–4
14. Sharma R, Kumar M (2013) Dual band microstrip antenna for C- and X band wireless applications. In: International conference on multimedia, signal processing and communication technologies (IMPACT), pp 154–158
15. Kohli S, Dhillon SS, Marwaha A (2013) Design and optimization of multiband fractal microstrip patch antenna for wireless applications. In: International conference on computational intelligence and communication networks (CICN), pp 32–36

16. Balanis C (2003) Antenna theory and design, 3rd edn. McGraw Hill publishing company, New Delhi
17. Pharwaha APS, Singh J, Kamal TS (2010) Estimation of feed position of rectangular microstrip patch antenna. IE J-ET 91
18. Sivia JS, Pharwaha APS, Kamal TS (2013) Design of sierpinski carpet fractal antenna using artifical neural network. Int J Comput Appl 68(8):5–10
19. Sivia JS, Bhatia SS (2015) Design of fractal based microstrip rectangular patch antenna for multiband applications. In: IEEE International Advance Computing Conference (IACC), pp 712–715

Design of a Dual-Band Low-Profile Single-Arm Spiral Antenna for Vehicular Applications

Pranjal Borah and Tulshi Bezboruah

Abstract In this work, a low-profile wired single-arm spiral antenna is designed with a ground plane for its possible use in vehicular applications. Measurements are carried out in the planar configuration for two different dielectrics, namely air and silicone gel substrate placed in between the antenna and the ground plane. Further, the antenna is conformed into a semi-hemispherical geometry without changing the dimensions and measurements are repeated. The result shows a dual-band behavior with an operation mainly in the L- and S-band rather than a wide-band performance observed for commonly reported spiral antenna.

Keywords Planar · Spiral · Silicone gel · Semi-hemispherical · Vehicular

1 Introduction

With the development in the recent technology, an automobile is becoming smarter day by day, where almost 24 different kinds of antennas are mounted on it. The antenna design for the vehicle is mostly determined by the parameters like limit inclination of a vehicle, antenna environment, angular velocity, speed of the vehicle as well as the operating temperature, available height, etc. [1]. In most of the cases, a low-profile antenna is preferred as it can be easily mounted on the vehicle surface consuming less power and space. For such requirement, the antenna designers prefer mostly the spiral antennas which can be transformed easily to a low-profile configuration. Spiral antennas are most popular because of its wideband characteristics in terms of impedance and axial ratio and its polarization [2]. Generally, spiral antennas consist of two arms which are to feed via a well-matched balun that offers a wide

P. Borah (✉) · T. Bezboruah
Department of Electronics and Communication Technology, Gauhati University, Guwahati 781014, Assam, India
e-mail: pranworld@gmail.com

T. Bezboruah
e-mail: zbt_gu@yahoo.co.in

range of bandwidth. However, the designers have to face difficulties while designing the feeding balun due to various complexities. This shifts the trend of designing two-arm spiral (TAS) antenna to single-arm spiral (SAS) [3] where the complexities of designing the balun is overcome by using only a single point of a co-axial feed. Works are reported on single-arm spiral antennas with some planar and non-planar configuration in various geometries such as: (a) rectangular, (b) hexagonal and (c) circular, etc., for their use in various communication systems [4–9]. As reported, the SAS shows comparable characteristics to that of TAS or multi-arm spiral antenna in terms of bandwidth, polarization, etc. [10].

In this work, a low-profile planar wired SAS antenna with a ground plane is designed and fabricated for its possible use in vehicular applications. The antenna can be confronted easily in the headlights of a vehicle and can also be mounted on the rooftop of the vehicle. The design is first made planar and measurements are carried out for air and flexible silicone gel (commercially available) dielectric placed in between the ground plane and the antenna. Further, measurements are carried out by modifying the antenna geometry to a semi-hemispherical form without changing the previous dimensions of the antenna.

2   Antenna Design

A spiral antenna operates mainly in three different principles namely: (i) traveling wave that gives the broadband characteristics, (ii) fast wave which is due to the mutual coupling between the arms and (iii) leaky wave that helps in the radiation phenomena [11].

In this work, a low-profile single-arm spiral antenna is designed and fabricated from copper wire whereas the ground plane is made up of AgHT-8, a transparent conducting polymer sheet. Commercially available silicone gel is used as one of the dielectric substrates due to some of its advantages such as flexibility, resistive to UV radiation and water, etc. which also acts as the supporting stand for the SAS preventing it from mechanical deformation. The silicone gel which is to be used as a dielectric substrate is first put into a masking holder and allow it to solidify before it is used. Figure 1a, b shows the schematic diagram of the antennas in planar and semi-hemispherical geometry. The design parameters of the SAS antenna are given in Table 1. Some of the fabricated antennas are shown in Fig. 2a–d.
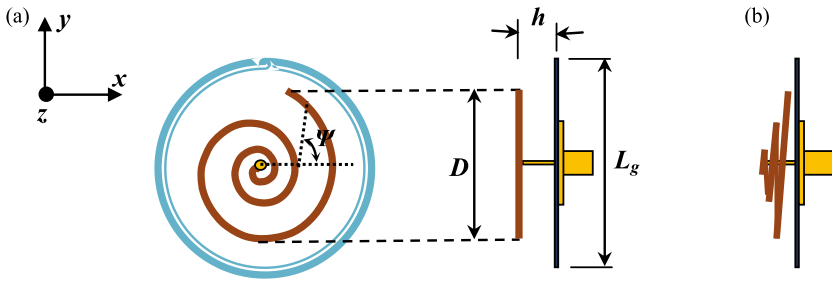
Fig. 1 a Schematic diagram of the antenna in planar configuration. **b** Schematic diagram of the antenna semi-hemispherical configuration
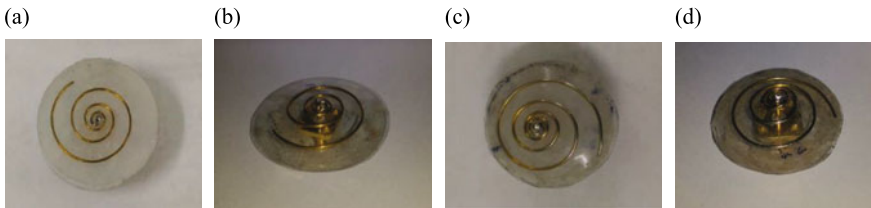
Fig. 2 a, **b** SAS planar antenna with silicone gel and air dielectric. **c**, **d** SAS semi-hemispherical antenna with silicone gel and air dielectric

3 Experimental Results

3.1 Return Loss Measurements

Measurements are carried out for observing the matching performance of the SAS antennas in the vector network analyzer (Rohde and Schwarz—ZNB 20). Figure 3 shows the measured return loss plot with air and silicone gel dielectric placed in between the spiral and the ground plane. Table 2 shows the measured return loss along with their corresponding resonant frequencies.

Table 1 Design parameters of the SAS antenna

Parameters	Values
D	30 mm
h	4 mm
L_g	40 mm
Ψ	70°
Silicone gel permittivity	2
Length of the spiral arm	150 mm
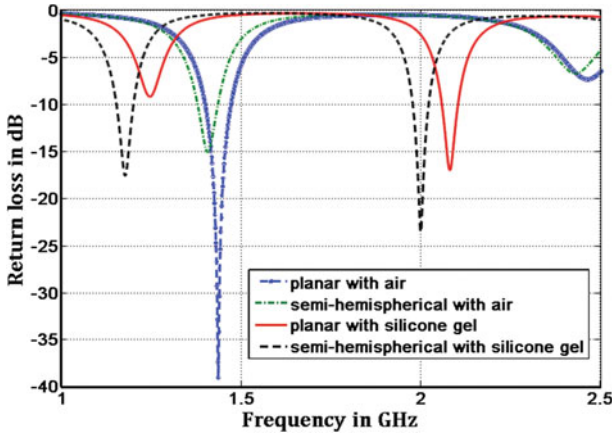
Fig. 3 Measured return loss plots for the planar and semi-hemispherical antenna configurations with different dielectric

Table 2 Peak return loss along with their corresponding resonant frequencies

Antenna configuration		Lower resonant frequency f_1 (GHz)	Return loss at f_1 (dB)	Upper resonant frequency f_2 (GHz)	Return loss at f_2 (dB)
Planar	Air	1.47	−39.03	2.46	−7.43
	Silicone gel	1.29	−9.23	2.08	−17.09
Semi-hemispherical	Air	1.41	−15.18	2.43	−6.78
	Silicone gel	1.17	−17.68	1.99	−23.59

From the results, it is observed that the antenna shows a dual-band behavior in the L- and S-band. With the introduction of the silicone gel dielectric, the resonant frequencies of both the antenna configuration shift towards the lower side of the RF spectrum. Moreover, a shift in the resonant frequency to the lower side is observed for the antenna with semi-hemispherical configuration as compared to the planar configuration in the presence of both air and silicone gel dielectric. This may be due to the increase in the overall area of the antenna while modifying from planar to semi-hemispherical configuration.

3.2 Radiation Pattern Measurements

Free-space radiation pattern measurements are carried out for the two antenna configuration by using DAMS antenna measurement system by calibrating with vector network analyzer. The measurements are carried out for the two principal plane (x-z

(a)

(b)

(c)

(d)
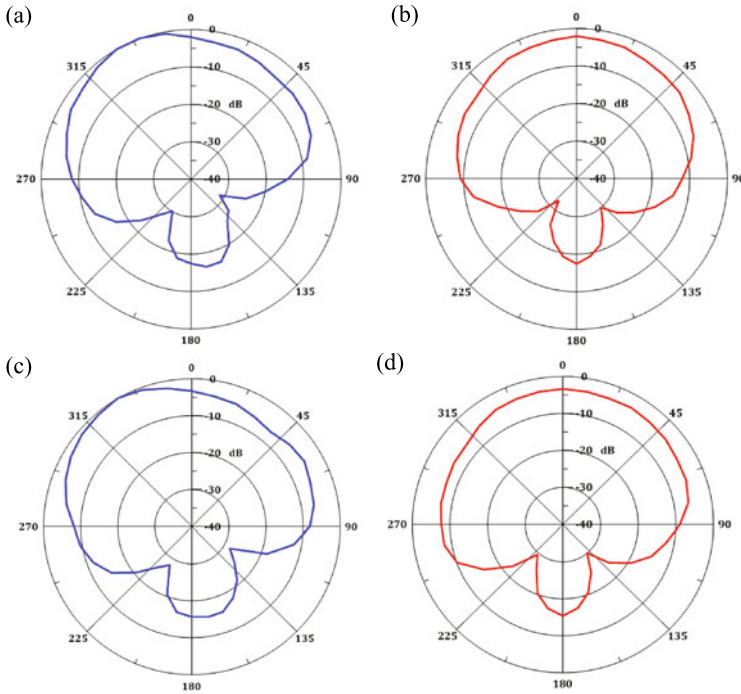
Fig. 4 **a**, **b** Measured x-z and y-z plane of the planar configuration @ 1.29 GHz. **c**, **d** Measured x-z and y-z plane of the semi-hemispherical configuration @ 1.17 GHz

& y-z) as shown in the orientation of Fig. 1. The measured results show similar patterns for the antenna configuration with air and silicone gel dielectric introduced in between the spiral and the ground plane. The measured plots of the planar and semi-hemispherical antenna configuration with silicone gel as the dielectric substrate for the lower resonating frequencies (f_1) are shown in Fig. 4a–d.

Measured radiation pattern plots of the planar and semi-hemispherical antenna configuration with silicone gel as the dielectric substrate for the upper resonating frequencies (f_2) are shown in Fig. 5a–d.

From the measured radiation pattern, it is observed that the main-lobe pattern in the x-z plane tilts away from the broadside direction for both lower and upper resonant frequencies of the antenna configurations which may be due to the extension of the single-arm spiral in one side. However, the measured pattern in the y-z plane maintains a good broadside radiation with its main-lobe at 0°. From the measured radiation pattern, it can be concluded that the main-lobe pattern in the x-z plane tilts away from the broadside direction for both lower and upper resonant frequencies of the antenna configurations which may be due to the extension of the single arm spiral in one side. However, the measured pattern in the y-z plane maintains a good broadside radiation with its main-lobe at 0°.
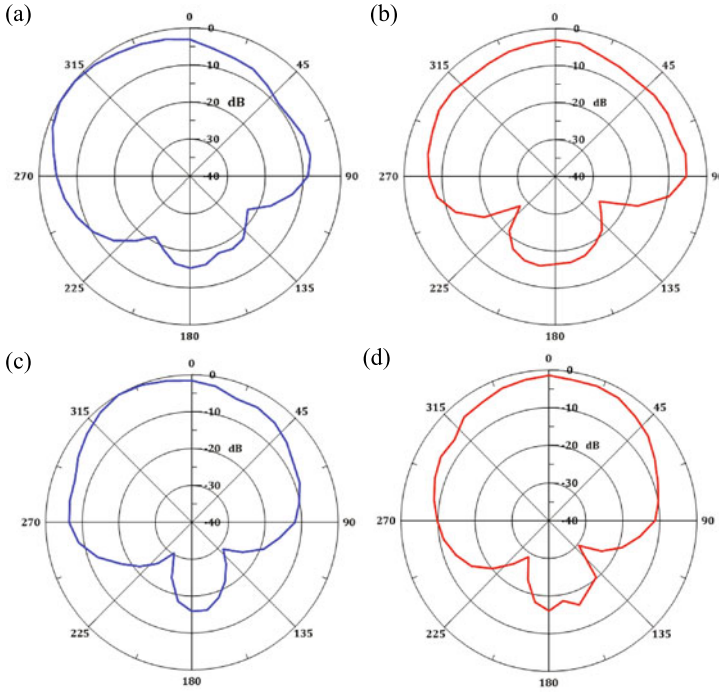
Fig. 5 **a**, **b** Measured x-z and y-z plane of the planar configuration @ 2.08 GHz. **c**, **d** Measured x-z and y-z plane of the semi-hemispherical configuration @ 1.99 GHz

4 Discussion and Conclusion

This paper investigates the idea and design consideration of a low profile planar single arm spiral antenna with a minimum number of turns which can be used for vehicular applications. The antenna shows a potential of dual-band operation in L- and S-band. First, the planar antenna configuration is tested with air and silicone gel dielectric placed in between the spiral arm and the ground plane. The results for this antenna configuration shows a better matching performance up to a return loss of −39.03 dB at 1.43 GHz for the lower resonant frequency with an air dielectric, whereas the upper frequency resonates at 2.46 GHz with a return loss of −7.43 dB. With the introduction of the silicone gel dielectric the lower resonant frequency degrades to a return loss of −9.23 dB at 1.29 GHz, on the other hand, the upper resonant frequency shows an improvement in the matching performance with a return loss of −17.09 dB resonating at 2.08 GHz. Second, the planar configuration is modified to a semi-hemispherical geometry without changing the dimensions and measurements are carried out for both air and silicone gel dielectric placed in between the spiral arm and the ground plane. For the semi-hemispherical configuration with the silicone gel dielectric, the antenna shows a better dual band operation with a return loss of −17.68 dB and −

23.59 dB resonating at 1.17 GHz and 1.99 GHz respectively. The radiation pattern measurement for both the antenna configurations shows a good broadside radiation pattern in the y-z plane, whereas a tilt in the radiation pattern with the main-lobe away from the broadside is observed in the x-z plane. From the experimental results and structural advantages of the antenna like low profile and ability to confirm it can be used as a potential element for vehicular dual-band applications by mounting it on the vehicle rooftop, headlights, etc. Analysis and optimization of the antenna parameter can be carried out by using simulation solvers for achieving better results in the future.

References

1. Shishlov AV (2011) Vehicular antennas for satellite communications. In: VIII International conference on antenna theory and techniques, Kyiv, pp 34–39
2. Nakano H, Shinma Y, Yamauchi J (1997) A monofilar spiral antenna and its array above a ground plane-formation of a circularly polarized tilted fan beam. IEEE Trans Antennas Propag 45(10):1506–1511
3. Nakano H, Satake R, Takeuchi K, Yamauchi J (2011) Roof-mounted vehicle single-arm spiral antenna without a balun circuit. In: XXXth URSI general assembly and scientific symposium, Istanbul, pp 1–4
4. Mehta A, Mirshekar-Syahkal D, Nakano H (2006) Beam adaptive single arm rectangular spiral antenna with switches. IEE Proc Microw Antennas Propag 153(1):13–18
5. Barton RJ, Collins PJ, Crittenden PE, Havrilla MJ, Terzuoli AJ (2007) A compact passive broadband hexagonal spiral antenna array for VHF remote sensing. In: IEEE international geoscience and remote sensing symposium, Barcelona, pp 593–595
6. Oh Y, Nam S (2017) Design of a low-profile 2 to 6 GHz circular polarized single arm hexagonal spiral array antenna. In: International applied computational electromagnetics society symposium, Italy, pp 1–2
7. Chuang G, Junjian B, Zhiliang T, Zhaoxiang M (2014) Analysis and design of conical equiangular spiral antenna using MPIE and QGA technique. In: IEEE 17th international conference on computational science and engineering, Chengdu, pp 1810–1813
8. Mohamad S, Cahill R (2017) Spiral antenna with reconfigurable HIS using liquid crystals for monopulse radar application. In: IEEE conference on antenna measurements & applications, Tsukuba, pp 55–58
9. Hirose K, Nakano H (1996) A single-arm spiral antenna with a tilted beam. In: Symposium on antenna technology and applied electromagnetics, Canada, pp 179–182
10. Nakano H, Igarashi T, Oyanagi H, Iitsuka Y, Yamauchi J (2009) Unbalanced-Mode spiral antenna backed by an extremely shallow cavity. IEEE Trans Antennas Propag 57(6):1625–1633
11. Balalnis CA Antenna theory and design, 3rd edn, Wiley

Application of Defected Ground Structure for Stable Gain with Ultrawide Bandwidth

Pawar Umesh Ankush, Abhijyoti Ghosh, L. Lolit Kumar Singh and Sudipta Chattopadhyay

Abstract A rectangular microstrip patch antenna integrated with ring-shaped defected ground structure has been studied for the improvement of impedance bandwidth. The use of defected ground structure (DGS) for the improvement of impedance bandwidth is a very new concept. Parametric studies have been carried out to find the optimum size of the defect. Around 46.5% impedance bandwidth (-10 dB) with 5.7 dBi peak gain is achieved with the optimum defect size. The proposed structure is very simple and easy to fabricate. The proposed structure is very much useful for the high-speed communication systems where wide impedance bandwidth along with stable radiation pattern is the primary requirement.

Keywords Microstrip patch antenna · Defected ground structure · Bandwidth · Gain · Voltage standing wave ratio

1 Introduction

In modern wireless communication systems microstrip patch antenna is the most common and popular candidate due to its simple design, low profile, thin and conformal properties. But high-speed wireless communication demands wide bandwidth with stable radiation pattern where conventional microstrip patch antenna lags as it provides narrow bandwidth (typically 2–3%) and low gain (1–3 dB) [1,

P. U. Ankush
Indian Army Unit, Delhi, India

A. Ghosh (✉) · L. Lolit Kumar Singh · S. Chattopadhyay
Department of Electronics and Communication Engineering, Mizoram University, Aizawl, Mizoram, India
e-mail: abhijyoti_engineer@yahoo.co.in

2]. Different techniques have been adopted by the antenna researcher community to improve the bandwidth of the conventional patch antenna like use of composite substrate, feed structure modification, aperture coupled dual polarization, etc. The impedance bandwidth (-10 dB) reported in [3] is only 7% with the use of composite substrates. 24% impedance bandwidth (-10 dB) is reported with aperture coupled dual polarization structure in [4, 5]. The above structures are very complex and the fabrication process needs much efforts. Modification of the feed structures [6–10] and the shape of the patch [11–16] are also reported to improve impedance bandwidth of the conventional patch antenna. 21.5% and 20% impedance bandwidth (-10 dB) is reported by using a very complex structure with L-probe fed inverted EE-H shape slotted rectangular patch antenna [6] and suspended probe feeding [7]. Stacked patch [8] and two-layer shorted square patch [9] structure is used to obtain impedance bandwidth (-10 dB) around 16% and 11% respectively. Impedance bandwidth of 10% by the dual polarized stacked patch is reported in [10]. However, all the above structure are complex and bulky. Impedance bandwidth (-10 dB) of 30%, 27%, and 54% is reported by E shape [11], U shape [12], and ψ shape [13] patch respectively. Shorted patch has been reported in [14, 15] to attain a impedance bandwidth (-10 dB) of 22%, 11% respectively whereas 25% impedance bandwidth (-10 dB) is reported with shorted patch with defected patch surface [16].

Defected ground structure based on electronic band gap (EBG) or photonic band gap (PBG) is well-known techniques and is used to improve different radiation characteristics of the microstrip patch antenna. Z-shaped DGS is reported in [17], to achieve impedance bandwidth (-10 dB) of 12.2%. Around 6–7% impedance bandwidth (-10 dB) is reported with Dumbelled [18] and slot type [19] DGS. Compare to these, [20, 21] show 25 and 22% impedance bandwidth (-10 dB) using cross and arc-shaped DGS. However, the design of optimum cross and arc defect need much more efforts. All the above-reported article improves the impedance bandwidth by increasing the size of the defect which increases the back radiation as a results antenna performance degrades.

In the present investigation, a simple single-layer ring-shaped defected ground structure (RDGS) integrated (Fig. 1) has been studied to improve the impedance bandwidth at entire X band frequency with stable radiation pattern and gain in the entire operating bandwidth. The proposed structure improves the impedance bandwidth by decreasing the size of the defect which in turn reduces the back radiation. Around 46% impedance bandwidth with stable gain above 4 dB is obtained from the proposed structure.

(a) (b)
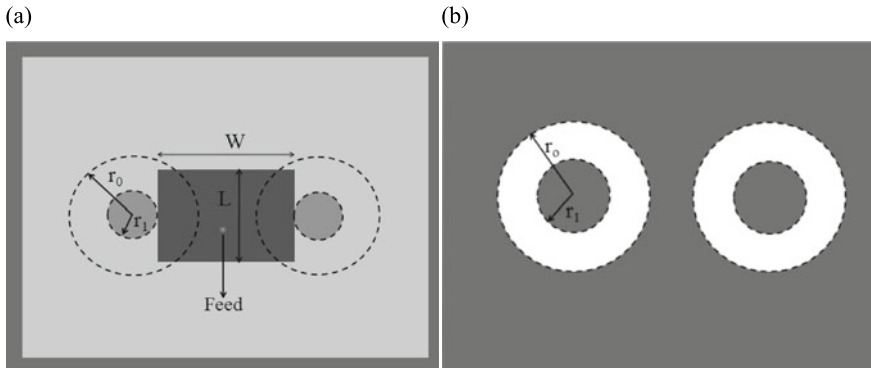
Fig. 1 Schematic representation of the proposed ring defected ground structure integrated rectangular microstrip patch antenna **a** top view, **b** bottom view

2 Theory, Parametric Studies and Proposed Structure

2.1 Theory, Parametric Studies

In conventional microstrip patch antenna patch (conductor) and ground plane (conductor) are placed in the either side of the dielectric substrate where electric walls (PEC) are formed by top (patch) and bottom (ground plane) conductors while four open boundaries are considered as the magnetic walls (PMC). Placing a defect in the ground plane modifies the electric fields between patch and the ground plane which finally alters the input and radiation characteristics of patch antenna. Now when the defect is introduced in the ground plane losses in the cavity model of the antenna increases which in turn increases the bandwidth of the antenna. Optimum performance of the patch in terms of impedance bandwidth depends on the geometry of the defect so the defect needs to be chosen very carefully.

Therefore, to increase the bandwidth of the patch antenna a pair of ring-shaped defect is placed on the ground plane near the non-radiating edges of the patch. The outer radius of the ring is kept fixed at 6 mm (i.e., $r_o = 6$ mm) while the inner radius (r_i) of the ring is varying to change the overall defect structure.

The impedance bandwidth, quality factor, and loss of an antenna are related as

$$\text{Bandwidth (BW)} = 1/Q_T \tag{i}$$

$$1/Q_T = 1/Q_r + 1/Q_d + 1/Q_c \tag{ii}$$

$$Q_T \cong 1/\text{Loss} \tag{iii}$$

where, Q_T, Q_r, Q_d, Q_c are total quality factor, quality factor due to radiation, quality factor due to dielectric and quality factor due to conductor respectively.

Fig. 2 Reflection coefficient profile for conventional and proposed antenna for different values of inner radius (r_i) keeping outer radius fixed ($r_o = 6$ mm)
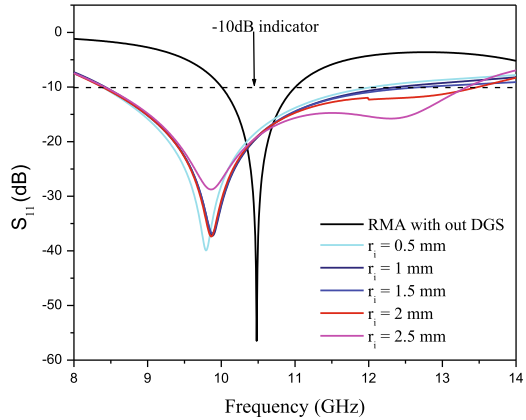
Fig. 3 Variation of impedance bandwidth (-10 dB) as a function of radius of the inner circle (r_i) keeping outer radius of the ring fixed ($r_o = 6$ mm)
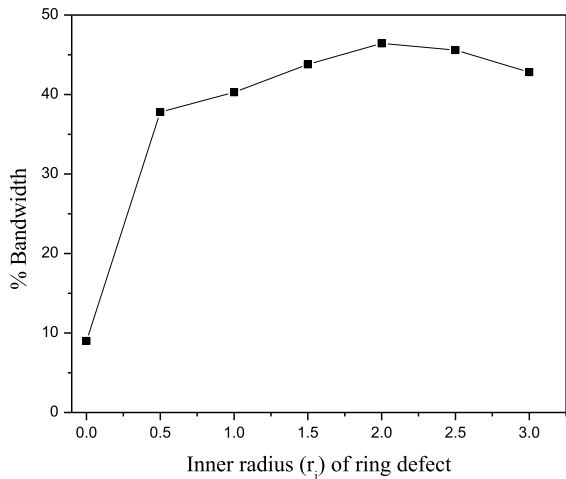
Introduction of the ring defect ($r_o = 6$ mm and $r_i = 0.5$ mm) suddenly increases the loss due to the radiation from the corner of the radiating edges of the patch which results sudden improvement of impedance bandwidth characteristics of the proposed antenna compare to the conventional patch antenna. The impedance bandwidth of the conventional patch antenna is only 9% whereas the same with ring defect ($r_o = 6$ mm and $r_i = 0.5$ mm) becomes almost 38% which is clear from Figs. 2 and 3. Introduction of the defect increases the effective permittivity of the antenna due to which resonant frequency of the proposed antenna shifts towards the lower side of the spectrum [22, 23].

The concentration of the magnetic field near the non-radiating edges of the patch is more. As the value of inner radius (r_i) of the ring increases the amount of conduction region due to the presence of the copper increases. The increment of the conduction region increases the losses due to the conduction which in turn decreases the value

Table 1 Details parameters of the proposed ring DGS integrated rectangular microstrip patch antenna on 70 mm × 70 mm ground plane

L (mm)	W (mm)	h (mm)	r_o (mm)	r_i (mm)	ε_r
8	12	1.575	6	2	2.33

of Q_c and finally the value of Q. As the substrate (dielectric) is kept unmodified so the loss due to dielectric remains same in both the cases. With $r_i = 2$ mm the inner radius of the ring reaches just below the non-radiating edge of the patch where most of the magnetic fields from that area terminates in the copper ground plane and results maximum conduction loss. The maximum conduction loss will produce minimum quality factor which results maximum impedance bandwidth (47%). The further increment of the inner radius decreases the bandwidth which is evident from Fig. 3. So the optimum size of the ring defect is $r_o = 6$ mm and $r_i = 2$ mm.

2.2 Proposed Structure

Initially, a rectangular patch antenna with ground plane 70 mm × 70 mm, patch size 8 mm × 12 mm with PTFE material as substrate (thinness (h) = 1.575) has been considered. A ring-type defected with inner circle radius "r_i" and outer circle radius "r_0" has been placed at the ground plane just below the non-radiating edges of the patch (Fig. 1). The different parameters of the proposed structure are presented in Table 1.

3 Results and Discussions

The results [24] obtained from the proposed ring DGS integrated rectangular microstrip patch antenna is discussed in this section. The reflection coefficient profile of conventional RMA and proposed RDGS-integrated RMA with optimum values of r_o and r_i is presented in Fig. 4. The conventional RMA resonates at 10.48 GHz and the proposed antenna resonates at 9.86 GHz. The improvement of the impedance bandwidth is clearly visible from Fig. 4. The impedance bandwidth (-10 dB) almost covers the full X band (8–12 GHz) of frequency.

Therefore,

$$\frac{2(f_H - f_L)}{(f_H + f_L)} = 0.5 > 0.2 \tag{iv}$$

and [25], the present proposed structure can be considered as Ultrawide Band (UWB) patch antenna.
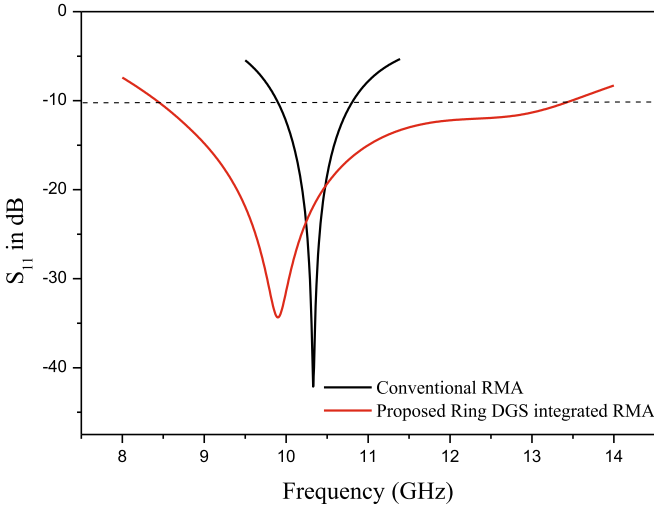
Fig. 4 Comparison of reflection coefficient profile of the convention RMA and proposed ring DGS integrated RMA
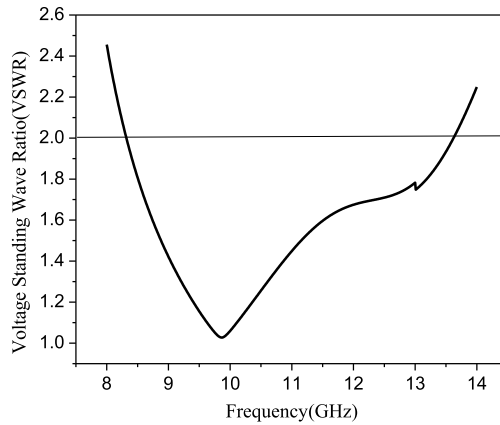
Fig. 5 VSWR profile for proposed antenna for $r_i = 2$ mm and $r_o = 6$ mm

The VSWR profile of the proposed antenna is shown in Fig. 5. The VSWR is reported to be below 2 which further confirm the wide bandwidth of the proposed antenna.

The gain of the proposed structure is 5.7 dBi which is quite good compared to the gain of the conventional patch antenna (2–3 dBi). Figure 6 shows the variation of the gain of the proposed structure over its whole impedance bandwidth. Gain maintains almost stable pattern within the operating bandwidth of the proposed structure except at the higher end of the spectrum. This may be due to more loss at higher frequency side.
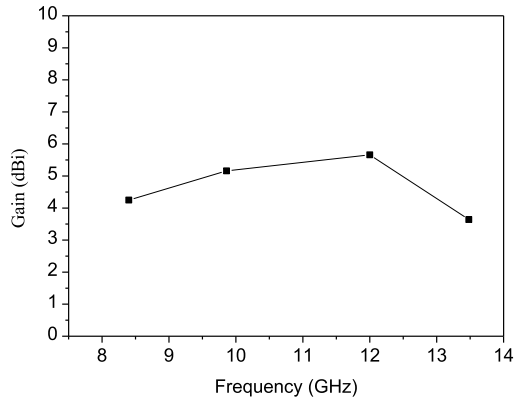
Fig. 6 Variation of gain as a function of frequency for the proposed structure with $r_i = 2$ mm and $r_o = 6$ mm

Figure 7 shows the complete radiation characteristics of the proposed RDGS in the entire operating band. It is clear from the figure that the E-plane and H-plane co polarization pattern is quite stable in the whole operating spectrum from 8.38 to 13.49 GHz.

4   Conclusion

A simple single element ring DGS integrated rectangular microstrip patch antenna is propose to achieve wide bandwidth with a stable gain in entire operating bandwidth. Wide bandwidth is always desirable from microstrip planer structure. Wide bandwidth through defected ground structure is a very new concept. Around 47% impedance bandwidth (-10 dB) is obtained from the proposed structure. In the proposed structure the bandwidth is increased by decreasing the size of the defect. The proposed structure covers the whole X band frequency. The proposed structure is simple and can be utilized for the applications where wide bandwidth is the primary requirement.
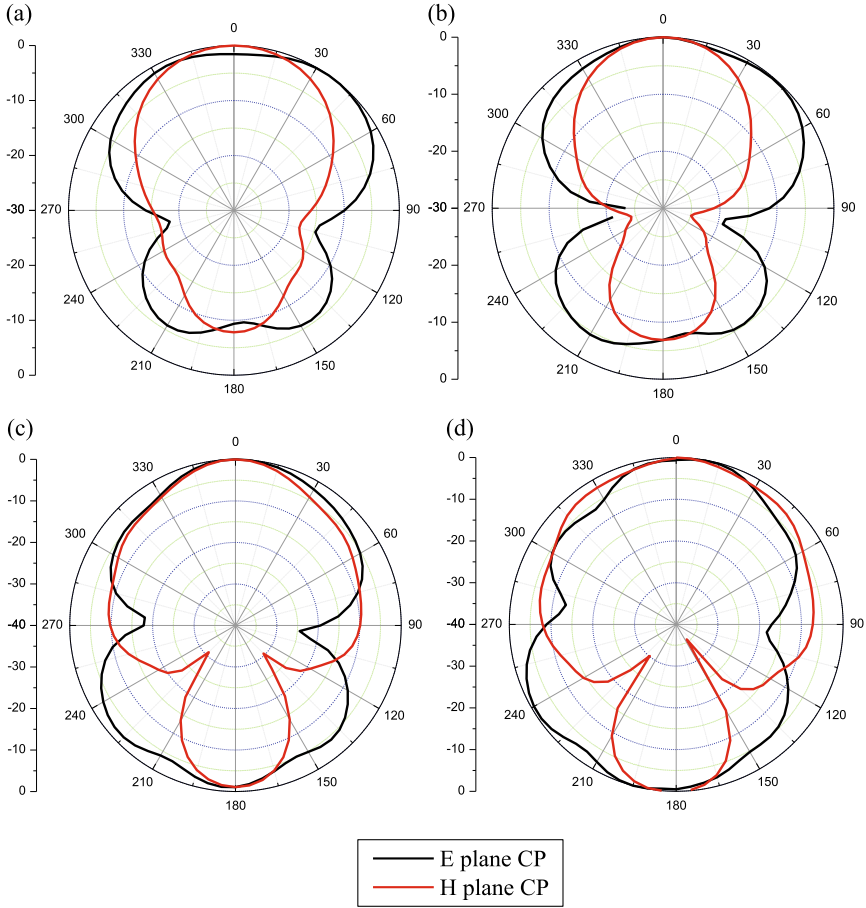
Fig. 7 Radiation pattern for E-plane and H-plane with proposed RDGS-integrated RMA at different frequencies **a** 8.38 GHz, **b** 9.86 GHz, **c** 12.75 GHz, **d** 13.49 GHz

References

1. Garg R, Bhartia P, Bahl I, Ittipiboon A (2001) Microstrip antenna design handbook. Artech House, Norwood
2. Guha D, Antar YMM (eds) (2011) Microstrip and printed antennas—new trends. Techniques and Applications. John Wiley, U.K
3. Chattopadhyay S, Siddiqui JY, Guha D (2009) Rectangular microstrip patch on a composite dielectric substrate for highgain wide-beam radiation patterns. IEEE Trans Antennas Propag 57:3324–3327
4. Gao S, Li LW, Leong MS, Yeo TS (2003) A broad-band dual-polarized microstrip patch antenna with aperture coupling. IEEE Trans Antenna Propag 51:898–900
5. Lai CH, Han TY, Chen TR (2008) Broadband aperture coupled microstrip antennas with low cross polarization and back radiation. Prog Electromag Res Lett 5:187–197

6. Islam MT, Shakib MN, Misran N (2009) Design analysis of high gain wideband L-probe fed microstrip patch antenna. Prog Electromag Res 95:397–407
7. Chen ZN, Chia MYW (2003) Broad-band suspended probe-fed plate antenna with low cross-polarization level. IEEE Trans Antennas Propag 51:345–346
8. Loffler D, Wiesbeck W (1999) Low-cost X-polarised broadband PCS antenna with low cross-polarisation. Electron Lett 35:1689–1691
9. Baligar JS, Revankar UK, Acharya KV (2001) Broadband two-layer shorted patch antenna with low cross-polarisation. Electron Lett 37:547–548
10. Granholm J, Woelders K (2001) Dual polarization stacked microstrip patch antenna array with very low cross-polarization. IEEE Trans Antennas Propag 49:1393–1402
11. Yang F, Zhang X, Ye X, Samii, YR (2001) Wide-band E-shaped patch antennas for wireless communications. IEEE Trans Antenna Propag 49:1094–1100
12. Tong KF, Luk KM, Lee KF, Lee RQ (2000) A broad-band U-slot rectangular patch antenna on a microwave substrate. IEEE Trans Antennas Propag 48:1148–1152
13. Sharma SK, Shafai L (2009) Performance of a novel Ψ-shape microstrip patch antenna with wide bandwidth. IEEE Antenna Wirel Prop Lett 8:468–471
14. Ghosh A, Chakraborty S, Ghosh SK, Singh LLK, Chattopadhyay S, Basu B (2017) Slot loaded microstrip antenna: a possible solution for wide banding and attaining low cross polarization. In: International conference on intelligent computing communication & devices, advances in intelligent computing & system, Springer
15. Poddar R, Chakraborty S, Chattopadhyay S (2016) Improved cross polarization and broad impedance bandwidth from simple single element shorted rectangular microstrip patch: theory and experiment. Frequenz 70:1–9
16. Ghosh A, Chattopadhyay S, Singh LLK, Basu B (2017) Wide bandwidth microstrip antenna with defected patch surface for low cross polarization applications. Int J RF Microw Comput Aided Eng 27:1–10
17. Kandwal A, Sharma R, Khah Sunil K (2013) Bandwidth enhancement using Z-shaped defected ground structure for a microstrip antenna. Microw Opt Technol Lett 55:2251–2254
18. Ghosh A, Chakraborty S, Chattopadhyay S, Nandi A, Basu B (2016) Rectangular microstrip antenna with dumbbell shaped defected ground structure for improved cross polarised radiation in wideelevation angle and its theoretical analysis. IET Microw Antennas Propag 10:68–78
19. Ghosh A, Ghosh D, Chattopadhyay S, Singh LLK (2015) Rectangular Microstrip antenna on slot type defected ground for reduced cross polarized radiation. IEEE Antennas Wirel Propag Lett 14:324–328
20. Ghosh A, Chattopadhyay S, Chakraborty S, Basu B (2017) Cross type defected ground structure integrated microstrip antenna: a novel perspective for broad banding and augmenting polarization purity. J Electromag Waves Appl 31:461–476
21. Chakraborty S, Chattopadhyay S (2017) Arc cornered microstrip antenna with defected ground structure for broadbanding and improved cross polarization suppression over whole skew planes. Int J Microw Wirel Technol 9:437–446
22. Oraizi H et al (2008) Miniaturization of Wilkinson power dividers by using DGS. Prog Electromagn Res Lett 4:113–120
23. Boutezdar et al (2011) An improvement of DGS lowpass/bandpass filters using H slot resonators and coupling matrix method. J Microw Opt Electromagn Appl 10:295–307
24. HFSS, High frequency structure simulator, Ver. 14, Ansoft Corp. USA
25. Wiesbeck W, Adamiuk G, Sturm C (2009) Basic properties and design principles of UWB antennas. Proc IEEE 97:372–385

Novel Composite Dielectric Resonator Antenna for 5G Applications

Manash Pratim Barman and Bidisha Dasgupta

Abstract In this paper, a novel compact dielectric resonator antenna (DRA) is reported. It operates at 28 GHz with a bandwidth of 2.8 GHz. It is a composite structure, consisting of two-quarters of hemispherical DRA (qHDRA) and one rectangular DRA (RDRA). A passive microstrip slot aperture (MSA) coupled feeding technique is designed to enable a suitable magnitude and phase excitation of the DRA for radiation. The results for reflection coefficient, bandwidth, gain, radiation pattern, and E-field strength are analyzed using Ansoft High-Frequency Structural Simulator (HFSS) [1]. The peak gain of the proposed antenna is about 6 dBi. Also an impedance bandwidth (IBW) of over 10% and an efficiency of 99.5% are observed. The proposed antenna structure is a good candidate for future 5G applications.

Keywords Dielectric resonator antenna (DRA) · Quarter hemispherical DRA (qHDRA) · Rectangular DRA (RDRA) · 5G · Microstrip slot aperture (MSA) coupled

1 Introduction

Since its inception, wireless technology is evolving at an astounding pace. Most of the current communication systems use a cellular spectrum which is below 3 GHz. The frequency spectrum that is currently in use, is getting overcrowded with an increase in the number of devices. Thus, millimeter-wave frequency, which is relatively unused, has become a better choice for the future 5G system [2]. Millimeter-wave bands are 26, 28, 38, and 60 GHz. Among these bands, the popular frequency candidates for the 5G systems are 28 and 38 GHz as they are not much affected by atmospheric absorption [2–4]. Since the frequency is very high, the size of the antenna will

M. P. Barman (✉) · B. Dasgupta
Indian Institute of Information Technology, Guwahati, India
e-mail: manashpratim2209@gmail.com

B. Dasgupta
e-mail: bidisha@iiitg.ac.in

be very small. High manufacturing precision and capabilities will be required to manufacture and implement these complicated structures [5]. Conductor losses will be high at these frequencies.

In this paper, Dielectric Resonator (DR) has been chosen to be designed for 5G system instead of using other conventional metallic printed antennas as they lack metal parts, which become lossy at high frequencies, dissipating energy. The DRAs have been widely reviewed in literature [6]. It consists of a dielectric volumetric structure that is made of a non-conductive, dielectric material with low loss and is excited via a coupling feed line. In past, various shapes of DRAs have been proposed in the literature [6] for a variety of applications including hemispherical, cylindrical, rectangular, etc., and the design procedures for all of these types are well documented [6]. Various composite structures of DRA are also well known in the literature [6]. In [7], composite cylindrical DRAs have been investigated. A new geometry of half-hemispherical DRA has been studied in [8]. A hemisphere is structurally suitable to radiate uniform oblique radiation surrounding it. In [9], the eightfold symmetry of a hemisphere is reported first time as a quarter of HDRA (qHDRA). The composite structure helps in producing enhanced gain, larger bandwidth and much-improved radiation pattern maintaining the 3-D symmetry.

DRAs can be fed using many different excitation techniques including microstrip. Each feeding approach can excite one or more radiation modes [10], and thus, each has its own advantages and disadvantages. Very limited research works have been carried out in past on qHDRA. The present study aims at exploring the potentiality of this geometry at 28 GHz frequency range and aims to fill the void to an extent.

In this paper, one novel compact antenna structure has been proposed which consists of two qHDRAs and one RDRA for short-range mm-wave communications at 28 GHz. Microstrip slot aperture feeding technique is used to feed the antenna. The design can be fitted on the edge of any modern handheld device to provide high-speed 5G communication links.

2 Antenna Configuration

Please refer to Figs. 1, 2 and 3 for the design of the antenna. The antenna consists of a composite DR structure consisting of two qHDRAs and a RDRA mounted on a ground plane of size, $a = 10$ mm and $b = 9$ mm in "x" and "y" axes respectively as shown in Fig. 1. A slot of dimensions, $c = 0.3$ mm and $d = 2.75$ mm is cut on the ground plane. A microstrip feed line of dimensions, $p = 5.9$ mm and $q = 0.3$ mm is used to couple the antenna structure electromagnetically and it is printed on the bottom of the Duroid substrate (dielectric constant 2.2) of thickness $t = 0.254$ mm. A source connects the feed line at the bottom of the plane to the ground which is at the top of the plane. The qHDRA are cut from a sphere of radius, $r = 2$ mm. The RDRA has a length, $l = 2$ mm, width, $w = 0.8$ mm and a height, $h = 2.24$ mm. Rogers RT with a permittivity of 10.2 was used as the material of the antenna.
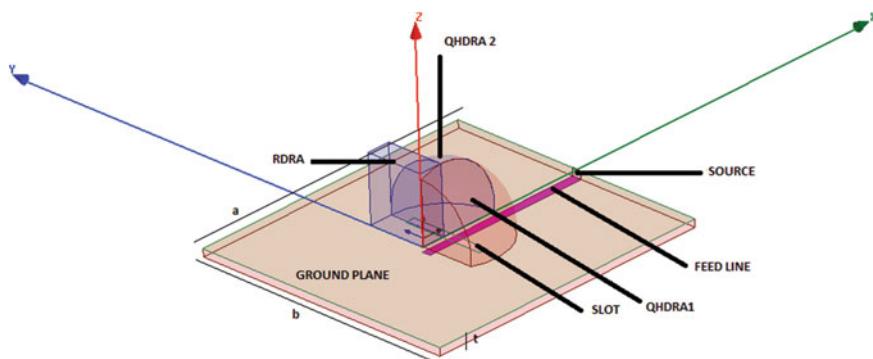
Fig. 1 Perspective view of the antenna
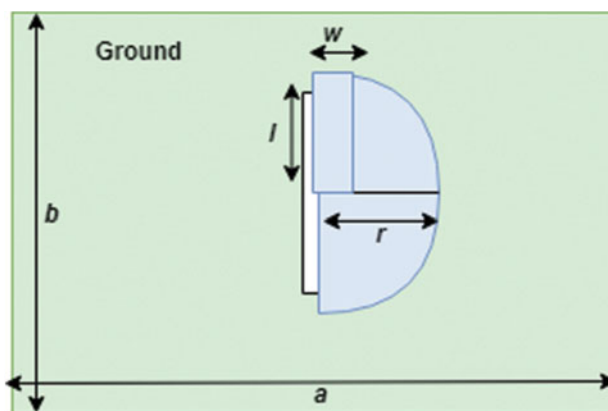
Fig. 2 View of the top plane
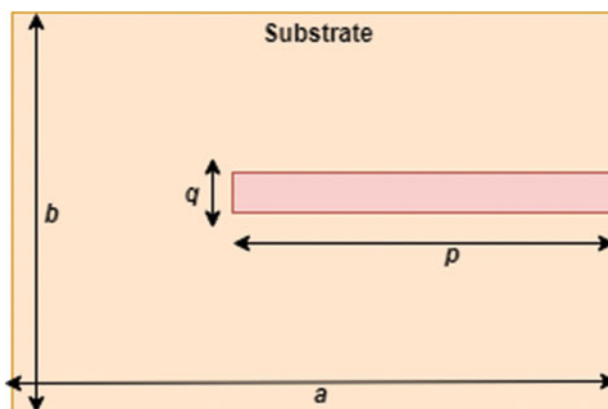
Fig. 3 View of the bottom plane

Fig. 4 Design chronology

QHDRA 1   QHDRA 2

Combined QHDRA   Proposed Structure
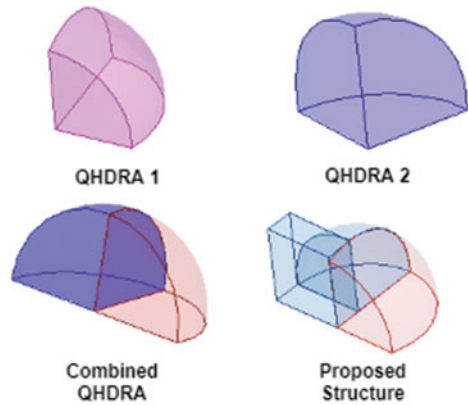
3 Results

The qHDRA 1 was designed by slicing it from a sphere of radius, r = 2 mm. Then, qHDRA 2 was designed by slicing it from another sphere of the same radius. However, it is cut in such a way that it is smaller than qHDRA 1. Then, the two QHDRAs were placed side by side and the combined QHDRA structure was formed. Eventually, qHDRA 1 and qHDRA 2 were joined and a RDRA was introduced, thereby arriving at the proposed structure. The chronological changes starting from a single qHDRA to the proposed design are shown in Fig. 4. The MSA feeding technique is used to feed the structures in Fig. 4.

A comparison of the return loss characteristics of the structures in Fig. 4 is shown in Fig. 5. The proposed structure provides a dip of around −45 dB. The resonance frequency is centered at 28 GHz with an operating bandwidth of 2.8 GHz or 10% IBW. A peak gain of 6.17 dB can be observed from the 3D gain plot shown in Fig. 6. The radiation pattern reveals the fact that the proposed antenna is providing a broadside radiation pattern throughout the band. The radiation patterns in principal planes are furnished in Figs. 7 and 8 corresponding to S11 minima. In Fig. 9, the radiation pattern of total gain is shown. Figure 10 shows the efficiency of the proposed design over the frequency range. A peak efficiency of 0.995(99.5%) is observed at 28 GHz. The electric field distribution in Fig. 11 shows HEM like mode. Figure 12 shows good impedance matching with 50 Ω connector.

4 Conclusion

A novel composite antenna structure consisting of two QHDRAs and one RDRA is proposed for future 5G applications. A maximum gain of 6.17 dBi is obtained and an operating bandwidth of 2.8 GHz is achieved. The design provides sufficient

Fig. 5 Return loss characteristics of the different structures

Fig. 6 3D plot of the proposed design at 28 GHz
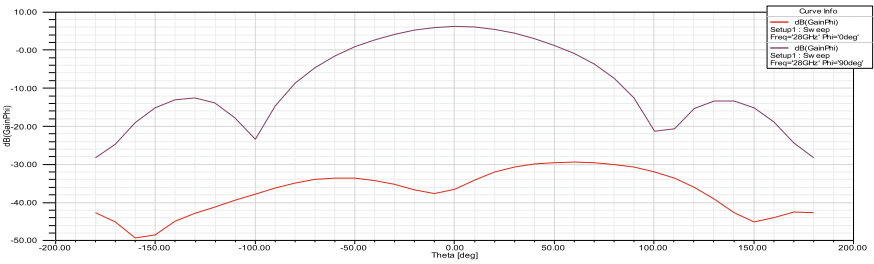
Fig. 7 Variation of gain phi in two principal planes
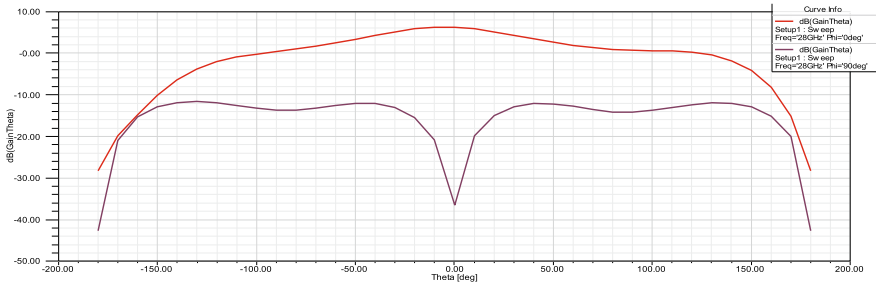
Fig. 8 Variation of gain theta in two principal planes
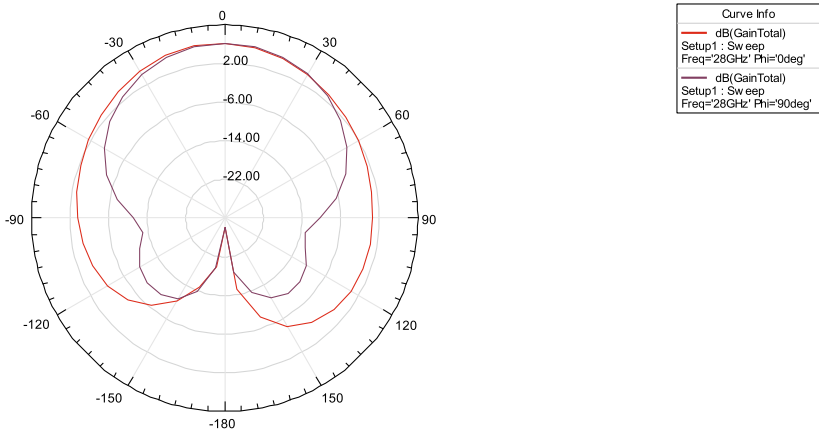
Fig. 9 Radiation pattern of total gain
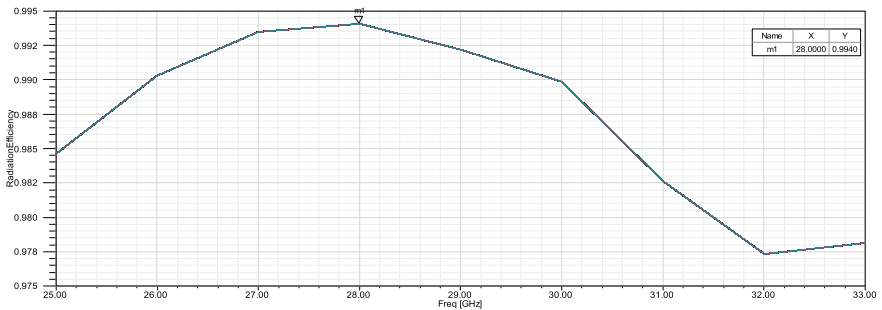
Fig. 10 Efficiency over the frequency band of the proposed antenna geometry
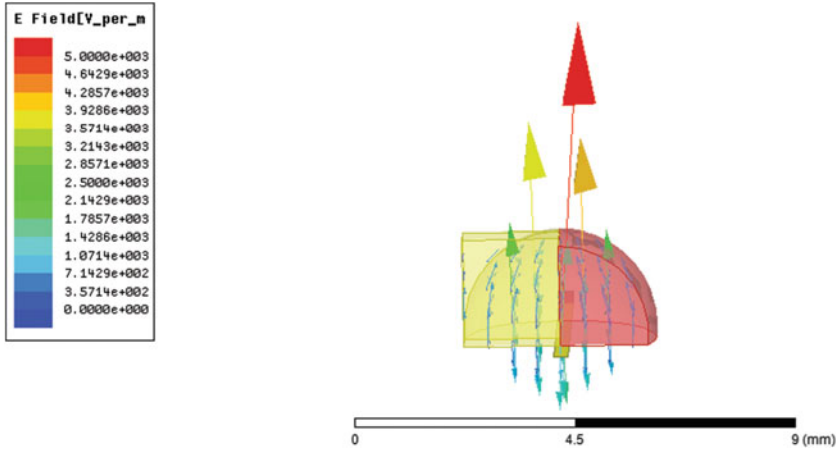
Fig. 11 Electric field distribution at 28 GHz
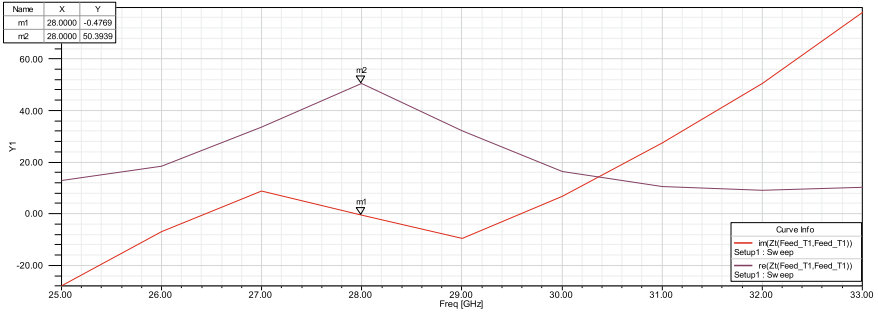
Fig. 12 Z characteristics of the proposed antenna

bandwidth at 28 GHz for intended applications. The proposed antenna also has acceptable radiation performance. In addition, the design is compact in size and can fit on the edge of any commercial handheld device for short range, high-speed data communication links. Future work would include designing a massive MIMO antenna with the design proposed in this paper.

References

1. High Frequency Structural Simulator
2. Rappaport TS et al (2013) Millimeter-wave mobile communications for 5G cellular: it will work! IEEE Access 1:335–349
3. Wei L et al (2014) Key elements to enable millimeter wave communications for 5G wireless systems. IEEE Wirel Commun 21(6):136–143

4. Sulyman AI et al (2014) Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands. IEEE Commun Mag 52(9):78–86
5. Liu D et al (2013) Guest editorial for the special issue on antennas and propagation at MM- and sub MM-waves. IEEE Trans Antennas Propag 61(4):1502–1507
6. Petosa A (2007) Dielectric Resonator Antenna Handbook, Artech House
7. Guha D, Antar YMM (2006) Four-element cylindrical dielectric resonator antenna for wide-band monopole-like radiation. IEEE Trans Antennas Propag 54(9):2657–2662
8. Guha D, Antar YMM (2006) New half-hemispherical dielectric resonator antenna for broad-band monopole-type radiation. IEEE Trans Antennas Propag 54(12):3621–3628
9. Guha D, Gupta B, Antar YMM Quarter of hemispherical dielectric resonator: new geometry explored to design a wideband monopole-type antenna. XXIXth URSI General Assembly, 7–16 Aug, 2008, Chicago, USA
10. Birand MT, Gelsthorpe RV Experimental millimetric array using dielectric resonator antennas. In: IEEE international symposium on antennas and propagation, vol 2. Montreal, Canada, pp 690–693, July 1997

Design of a Dual-Band Rectangular Microstrip Patch Antenna with Diagonal Slit for Its Use in L- and S-Band Communications

Tilak Sarmah, Pranjal Borah and Tulshi Bezboruah

Abstract In this paper, an investigation is carried out on the effect of diagonal slit introduced on the rectangular patch antenna to obtain dual-band operation by simulating on High-Frequency Structure Simulator. Initially, a rectangular patch by microstrip line feeding technique is designed with the help of standard equations at the resonant frequency of 2.4 GHz. Later on, a slit of width 80 µm is introduced diagonally, with an inclination of 45° along the length of the patch and simulations has been carried out for different slit locations by varying along the width of the patch to study the effect of the slit on the antenna characteristics. The simulation result shows that the simple rectangular microstrip patch antenna resonates at a single frequency of 2.43 GHz with a return loss of −22.61 dB. With the introduction of the slit, the antenna shows a dual-band behavior with resonating frequencies lower than the simple rectangular patch antenna. The best dual-band performance is obtained with a return loss of −15.93 dB and −21.53 dB resonating at 1.88 GHz and 2.17 GHz, respectively.

Keywords RMSA · Dual-band · Slit · Inclined

1 Introduction

Microstrip patch antenna (MPA) finds its applications in various technologies because of its advantages like (i) low profile, (ii) lightweight, and (iii) compact in size. However, the main disadvantage of microstrip antenna is its narrow bandwidth (BW) and poor impedance matching capacity. This disadvantage can be overcome by making

T. Sarmah (✉) · P. Borah · T. Bezboruah
Department of Electronics and Communication Technology, Gauhati University, Guwahati 781014, Assam, India
e-mail: tilaksarmah1@gmail.com

P. Borah
e-mail: pranworld@gmail.com

T. Bezboruah
e-mail: zbt_gu@yahoo.co.in

the antenna to resonate at multiple frequencies at the required operating bands. It is a major challenge for the designer to design an antenna of compact in size and operating at multiple frequencies. As reported, the dual-band operation of a patch antenna can be obtained by some techniques such as (a) by using shorting post [1–3], (b) by using stubs [4–6], and (c) by cutting slits [7–9] etc.

In the proposed work the dual-band operation for a rectangular microstrip patch antenna (RMSA) is obtained by cutting a slit diagonally along the length of the designed rectangular patch with an inclination of 45°. Simulations have been carried out for different slit locations by varying along the width of the patch. The results are compared with the simple RMSA. Simulated return loss performances show a good dual-band behavior in the L- and S-band of the radio frequency (RF) spectrum. This is followed by the simulation to observe the radiation pattern which shows a good pattern in the broadside direction. During investigation, the slit is also introduced horizontally and vertically along the length of the patch which shows single resonance of the structure within the range of 1–2.75 GHz.

2 The Design Consideration

The design of the RMSA is carried out by using the standard equations as in [10].

2.1 Calculation for Width (W)

The width of the antenna is expressed by Eq. (1) [10]:

$$w = \frac{1}{2 f_r \sqrt{\mu_0 \in_0}} \sqrt{\frac{2}{\in_r +1}} = \frac{c}{2 f_r} \sqrt{\frac{2}{\in_r +1}} \tag{1}$$

where

c free space velocity of light.
$\in_r$ Dielectric constant of substrate.

2.2 Calculation of Effective Dielectric Constant (ε_{eff})

The effect of dielectric constant is given in Eq. (2) [10]:

$$\in_{eff} = \frac{\in_r +1}{2} + \frac{\in_r -1}{2} \left(\frac{1}{\sqrt{1 + \frac{12h}{w}}} \right) \tag{2}$$

2.3 Calculation of Actual Length of the Patch (L_{eff})

The actual length of the patch is given in Eq. (3) [10]:

$$L_{eff} = \frac{c}{2f_r\sqrt{\epsilon_{eff}}}\qquad(3)$$

3 The Antenna Configuration

In the proposed investigation, we have designed and analyzed the antenna by using ANSOFT High-Frequency Structure Simulator (HFSS) version: 13.0.0. Here, FR4 ($\varepsilon_r = 4.4$) is used as the dielectric substrate. Both the antennas are fed with microstrip feed line having characteristic impedance of 50 Ω. Figure 1a, b show the schematic diagram of the designed RMSA without and with the introduction of the slit respectively. The dimensions of the antenna are given in Table 1.
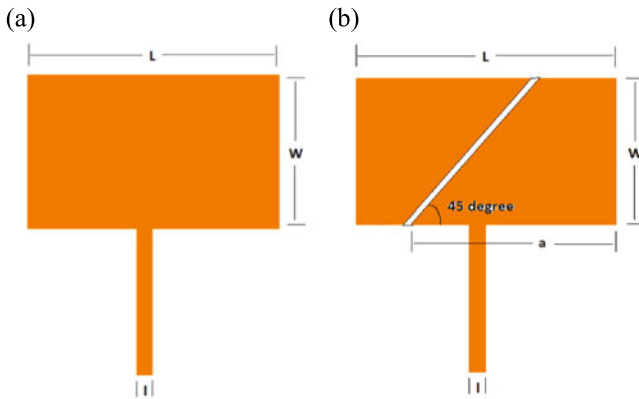
Fig. 1 a, b Schematic diagram of the microstrip patch antenna without and with diagonal slit

Table 1 Parameters of the antenna

Parameters	Simulation values
Resonant frequency	2.4 GHz
Dielectric constant	4.4
Substrate thickness	1.5 mm
Patch width	28.2 mm
Patch length	47 mm
Width of the strip line	1.8 mm
Width of the slit	80 μm
Slit variable **a**	0, … 46 mm

4   The Return Loss Measurement

The simulated return loss measurements for different values of slit variable "**a**" are given in Table 2, whereas the return loss plots of the RMSA without and with slit (for a = 40 mm) is shown in Fig. 2a, b. From the simulation results show that the RMSA without the slit resonates only at 2.43 GHz with a return loss of −22.60 dB. With the introduction of the diagonal slit, the patch shows a dual-band behavior by resonating at two different frequencies. For the diagonal slit cut over the patch, a better dual band is observed for the slit variable **a** = 40 mm with return losses of − 15.93 dB and −21.53 dB at 1.88 GHz and 2.17 GHz, respectively.

4.1   The Radiation Pattern

To observe the radiation patterns of the RMSA, they are simulated at their respective resonant frequencies without and with diagonal slits. The result shows that for the RMSA without the slit has a good broadside radiation pattern for both the E (x-z)

Table 2 Values of resonant frequencies and return losses of the antenna with respect to the slit variable (**a**) from the right edge of the patch

Sl. no.	Slit variable "a" (mm)	Lower resonant frequency f_1 (GHz)	Return loss for f_1 (dB)	Upper resonant frequency f_2 (GHz)	Return loss for f_2 (dB)
1	0 (solid)	2.43	−22.60	–	–
2	05	2.43	−24.22	–	–
3	08	2.45	−20.61	–	–
4	12	2.47	−16.69	–	–
5	20	2.52	−12.33	–	–
6	25	1.63	−2.00	2.54	−1.93
7	30	1.75	−2.62	2.41	−3.19
8	35	1.87	−5.47	2.21	−4.87
9	37	1.91	−10.78	2.17	−7.03
10	38	1.89	−12.38	2.15	−8.85
11	39	1.89	−13.61	2.15	−11.60
12	40	1.88	−15.93	2.17	−21.53
13	41	1.85	−15.25	2.13	−18.93
14	42	1.83	−15.46	2.21	−4.56
15	45	1.67	−5.77	2.60	−12.86
16	46	1.63	−4.43	2.25	−3.34

(a) (b)
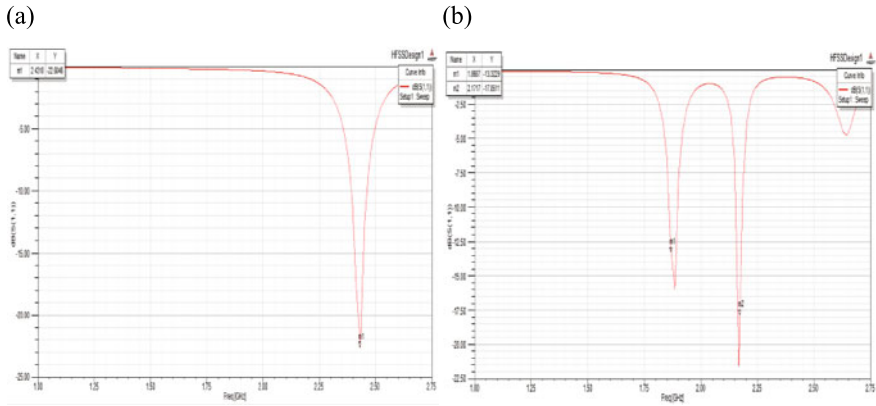
Fig. 2 a, b Simulated return loss plots of the microstrip antenna without and with slits (for **a** = 40 mm)

plane and the H (y-z) plane. For the RMSA with diagonal slit, the radiation pattern in the lower frequency shows a good broadside pattern, whereas a tilt in the pattern is observed for both the principal planes in the upper resonant frequency. The simulated radiation patterns in 2D are shown in Figs. 3a, b and 4a–d. The 3D radiation patterns are shown in Fig. 5a–c.

(a) (b)
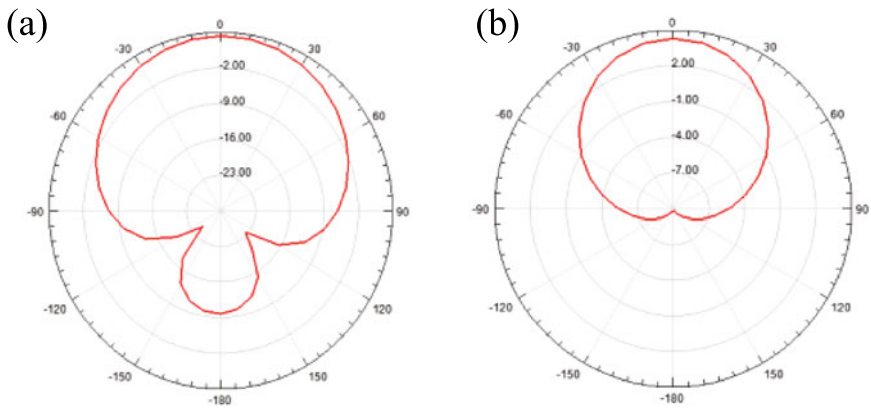
Fig. 3 a 2D E(x-z) plane pattern for RMSA without slit at 2.43 GHz. **b** 2D H(y-z) plane pattern for RMSA without slit at 2.43 GHz

Fig. 4 a 2D E(x-z) plane pattern for RMSA with slit at 1.88 GHz. **b** 2D H(y-z) plane pattern for RMSA with slit at 1.88 GHz. **c** 2D E(x-z) plane pattern for RMSA with slit at 2.17 GHz. **d** 2D H(y-z) plane pattern for RMSA with slit at 2.17 GHz
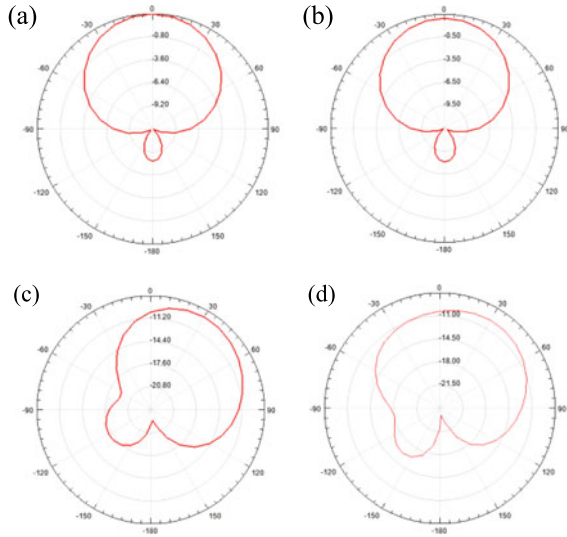
Fig. 5 a 3D Radiation pattern for RMSA without slit at resonant frequency 2.43 GHz. **b** 3D radiation pattern for RMSA with slit at resonant frequency 1.88 GHz. **c** 3D radiation pattern for RMSA without slit at resonant frequency 2.17 GHz
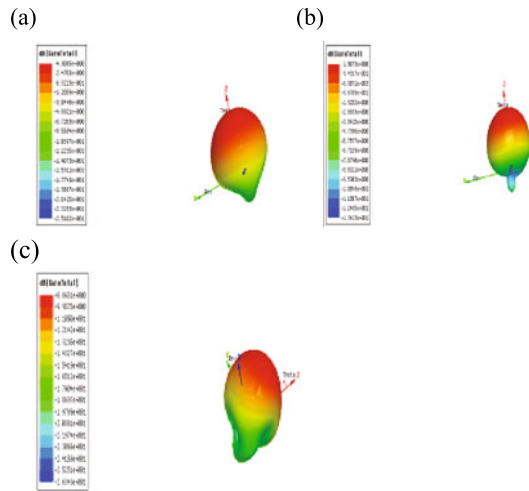
5 Discussion and Conclusion

From the comparative study of the simulation results of the two RMSA, it can be concluded that with the same dimensions, by introducing an inclined slit on the pure RMSA, one can transform it into a dual-band antenna and can also improve the antenna performances. With the introduction of the slit on the pure RMSA, the electrical path length increases which holds the reason for a dual-band performance with a shift in the resonant frequency to the lower side of the RF spectrum than the RMSA without the slit.

References

1. Borah P, Bhattacharyya S (2015) Design of a dual band v-shaped patch antenna using shorting posts. Microw Opt Technol Lett 58(2)
2. Sanad M (1994) Effect of the shorting posts on short circuit microstrip antennas. IEEE 794
3. Kumar P, Singh G Microstrip antennas loaded with shorting post. SciRes (http://www.SciRP.org/journal/eng/). Accessed June 2009
4. Schrank H, du Plessis M, Cloete J (1994) Tuning stubs for microstrip-patch antennas. IEEE Antennas Propag Mag 36(6)
5. Deshmukh AA, Baxi P, Kamdar C, Vora B, Ray KP Analysis of stub loaded rectangular microstrip antenna ©2012 IEEE
6. Vedaprabhu B, Vinoy KJ A double u-slot patch antenna with dual wideband characteristics, fre IEEE 2010
7. Singh J, Singh T, Sohi BS Design of slit loaded rectangular microstrip patch antenna. In: Proceedings of 2015 RAECS UIET Panjab University Chandigarh, 21–22nd Dec 2015
8. Sharma A, Gupta SK, Kanaujia BK, Pandey GP Analysis of superstrate loaded slit cut circular patch antenna for triple band operations. In: 2012 fourth international conference on computational intelligence and communication networks
9. Borah P, Bordoloi AK, Bhattacharyya NS, Bhattacharyya S (2012) Bridged 'V'-shaped patch antenna for dual band communication. Electron Lett 48(8)
10. Balanis CA (2005) Antenna theory analysis and design, Wiley-Interscience, pp 810–817

Modeling and Analysis of Metamaterial-Based Antenna for Wi-Fi and WLAN Applications

Amrithaa Seshadri and Navneet Gupta

Abstract In this work, a 4 × 4 Rectangular Split Resonator (SRR) array with resonance at 5.8 GHz is proposed. By optimizing the geometry through genetic algorithm, sub-wavelength resonance at the desired frequency was obtained, while maintaining the antenna's radiation efficiency at broadside. The total size of the antenna is 30 mm × 30 mm, thus ensuring compactness. The design was compared with a Circular SRR and an alternating sequence of Rectangular and Circular SRRs. Further, the permittivity and permeability of the structure was plotted, which was negative at the resonance frequency.

Keywords Metamaterials · Split-ring resonators · Antenna array · Miniaturization · Radiation performance

1 Introduction

Extensive research is being carried out in antenna engineering to optimize the gain of microstrip patch antennas. Due to its low profile, planar configuration, and ease in fabrication [1], patch antennas are preferred in mobile handsets, Wi-Fi, and stealth technology applications. Since the invention of metamaterials in 2002, they are being specially engineered to increase the performance of electrically small antenna systems [2–5]. Its novel structure stores and re-radiates energy, thus behaving as if it were much larger than its real size. This enables the construction of miniaturized antennas without major performance degradation.

A. Seshadri
Department of Electronics and Communication Engineering, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai, UAE
e-mail: amrithaa3011@gmail.com

N. Gupta (✉)
Department of Electrical & Electronics Engineering, Birla Institute of Technology and Science, Pilani, Pilani Campus, Pilani, Rajasthan, India
e-mail: ngupta@pilani.bits-pilani.ac.in

Antenna arrays are widely used to achieve superior performance as compared to a single element antenna since the Signal to Noise Ratio (SNR) is maximized and diversity reception is also enabled, in case of arrays. In a recent study by Nordin et al. [6], a 1 × 3 SRR array was fed by a trident feed and slotted ground plane was used to achieve UWB transmission parameters with acceptable levels of gain throughout the frequency band. In another study by Joshi et al. [7] probed into the effect of metallic via holes in the ground plane, on the SRR array. It was concluded that metallic holes increased the bandwidth by the factor of 8.20, and higher gain was also achieved.

In this paper, we have proposed a novel design for a 4 × 4 SRR array. Three configurations, viz., Rectangular (R-SRR), Circular (C-SRR), and their combination (M-SRR) were considered for simulation. The design parameters were optimized using Genetic Algorithm (GA) [8]. The SRR arrays were designed for primary resonance in the 5.3–5.8 GHz frequency band, which is the second most favored range for Wi-Fi and WLAN applications according to the newer IEEE 802.11ac technology as it provides faster data rates and less interference as compared to the conventional 2.4–2.5 GHz frequency.

2 Design of the 4 × 4 SRR Array

The design and analysis of the structure have been carried out using full-wave simulation software CST MICROWAVE STUDIO that uses Finite Integral Technique for computation.

A rectangular and circular SRR with splits etched on both ends (as shown in Fig. 1), were used to create the 4 × 4 SRR array. The dimensions used for designing SRR are given in Table 1.

The design of R-SRR and C-SRR are used to construct a 4 × 4 array, where the patches are connected to each other using thin feed strips, as shown in Fig. 2a, b. The whole array is connected to the feed line through a common horizontal strip that also
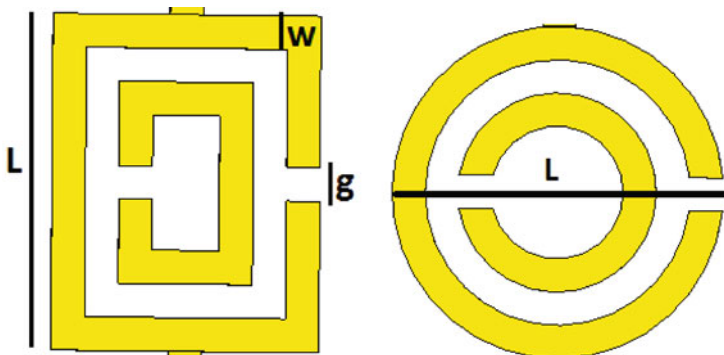
Fig. 1 (left) Rectangular, and (right) circular SRR patch with dimensions

Table 1 Dimensions for the SRR patch

Parameter	Dimension (in mm)
Length of patch (L)	4
Gap width (g)	0.4
Width of SRR (w)	0.4
Thickness (t)	0.01

acts as a capacitive element in the circuit analysis. This arrangement is etched on an FR-4 substrate of thickness 1.6 mm and dimensions 30 mm × 30 mm with ground plane in the back, as shown in Fig. 2.

Electric and magnetic boundaries are implemented along the x and z axes, while waveguide transmission port is applied to the feed on the y-axis. For the third design as shown in Fig. 2c, R-SRR and C-SRR elements are placed alternatively.

Here, GA was used in CST to optimize the inter-element spacing (d) and length of the capacitive strip (x). The optimized values of "*d*" and "*x*" are calculated to be 6.633 and 25.2 mm respectively.

The S-parameter analysis is done for all the three arrays and the plots obtained are shown in Fig. 3. It is observed that all the three arrays show almost similar trend. In the frequency range 2–12 GHz, there are four peaks around 3.8, 4.5, 5.83, and 6.5 GHz, where 5.83 GHz is the resonant frequency of operation.

Further, the gain of the three antenna configurations was analyzed and is shown in Fig. 4. R-SRR shows a good gain of 6.9 dBi at broadside, while M-SRR has a gain of 4.73 dBi, and C-SRR shows a very low gain of 1.11 dBi. Thus, the R-SRR is chosen as our primary design, and its EM analysis is discussed in the next section.

3 Analysis of the Metamaterial Structure

To analyze the response of the structure as a metamaterial, the S-parameter retrieval approach [9] is used to calculate the permittivity and permeability as per Eqs. (1)–(4):

$$n = \frac{1}{kd} \cos^{-1}\left[\frac{1}{2S_{21}}(1 - S_{11}^2 + S_{21}^2)\right] \tag{1}$$

$$z = \frac{\sqrt{(1 + S_{11})^2 - S_{21}^2}}{\sqrt{(1 - S_{11})^2 - S_{21}^2}} \tag{2}$$

$$ep(j) = \frac{n(j)}{z(j)} \tag{3}$$

$$mu(j) = n(j) * z(j) \tag{4}$$

Fig. 2 Front view of the
a R-SRR, **b** C-SRR, and
c M-SRR arrays

(a)

(b)

(c)
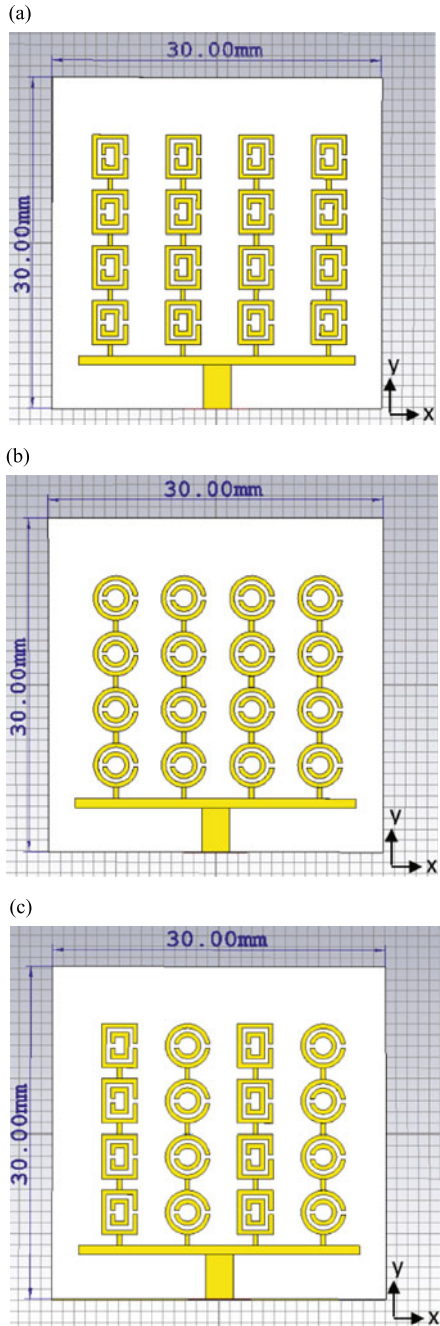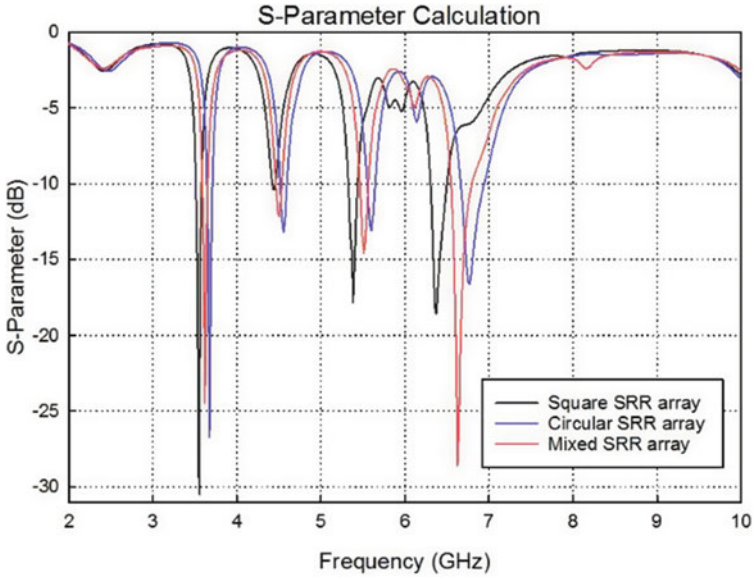
Fig. 3 The return loss S_{11} of the R-SRR, C-SRR and M-SRR array
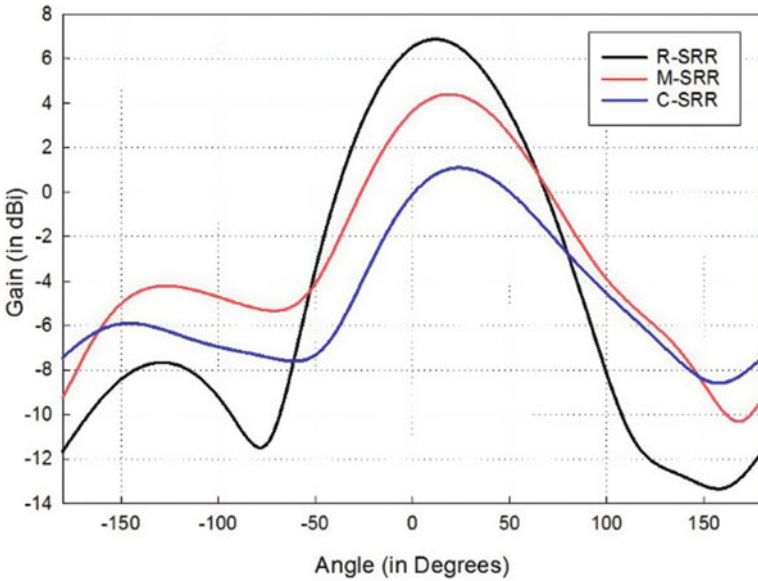
Fig. 4 The measured radiation performance of the R-SRR, C-SRR, and M-SRR arrays in the xy plane
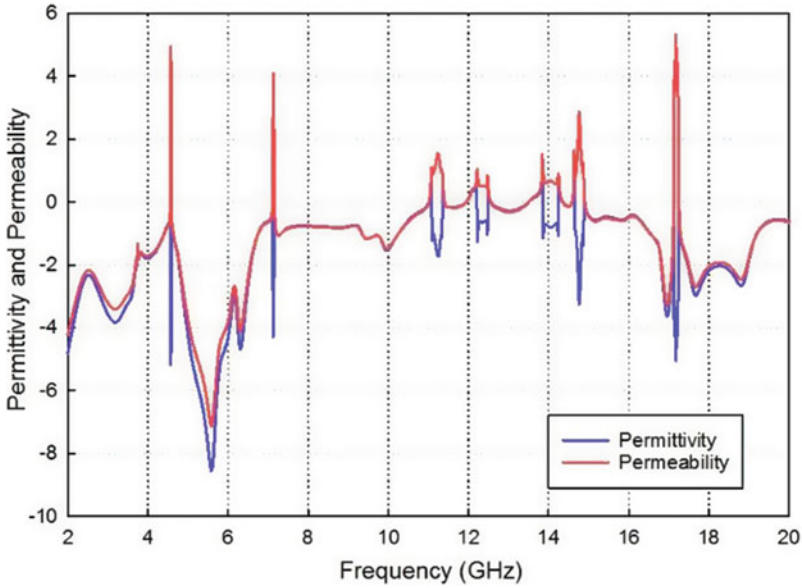
Fig. 5 Permittivity and Permeability plot for the R-SRR structure

where "*n*" is the refractive index, "*z*" is the wave impedance (η) "*ep*" is the permittivity (F/m), "*mu*" is the permeability (H/m), "*k*" is the wavenumber (ω/c), and "*d*" is the spacing between the two loops. The reflection and transmission coefficients S_{11} and S_{21} were exported from CST to MATLAB, and the results are shown in Fig. 5. The negative value of the permittivity at the resonant frequency of 5.8 GHz, confirms that the structure is acting as a metamaterial.

4   Conclusion

In this paper, three different configurations of a 4×4 SRR array were studied, viz., Rectangular, Circular, and an Alternating Sequence of the former. The structure was designed to operate at around 5.8 GHz, to support Wi-Fi and WLAN applications. It was observed that at the resonant frequency, the Rectangular SRR array exhibited the maximum gain of 6.9 dBi at broadside. Further, the NRZ approach was used to plot the negative permittivity and permeability of the structure at the resonant frequency, thus confirming its performance as a Left-Handed Metamaterial (LHM). Due to its high gain for broadside and compactness of the structure (30×30 mm^2), this metamaterial (SRR) antenna can be implemented in mobile handsets, and other wireless applications.

References

1. Singh I, Tripathi VS (2011) Micro strip patch antenna and its applications: a survey. Int J Comp Tech Appl 2(5):1595–1599
2. Ouedraogo RO et al (2012) Miniaturization of patch antennas using a metamaterial-inspired technique. IEEE Trans Antennas Propag 60(5):2175–2182
3. Naser-Moghadasi M et al (2014) Compact EBG structures for reduction of mutual coupling in patch antenna MIMO arrays. Prog Electromagnet Res 53:145–154
4. Majid HA, Abd Rahim MK, Masri T (2009) Microstrip antenna's gain enhancement using left-handed metamaterial structure. Prog Electromagnet Res 8:235–247
5. Tang J et al (2014) A continuously tunable miniaturized patch antenna. IEEE Antennas Wirel Propag Lett 13:1080–1083
6. Nordin MAW, Islam MT, Misran N (2013) Design of a compact ultrawideband metamaterial antenna based on the modified split-ring resonator and capacitively loaded strips unit cell. Prog Electromagnet Res 136:157–173
7. Joshi JG, Pattnaik SS, Devi S (2011) Metamaterial loaded dual band microstrip patch antenna. Antenna Week (IAW), 2011 Indian, IEEE
8. Ge Y, Esselle KP, Hao Y (2007) Design of low-profile high-gain EBG resonator antennas using a genetic algorithm. IEEE Antennas Wirel Propag Lett 6:480–483
9. Ziolkowski RW (2003) Design, fabrication, and testing of double negative metamaterials. IEEE Trans Antennas Propag 51(7):1516–1529

Microstrip Line Fed Two Element E-Plane HDRA Antenna Array

Goffar Ali Sarkar and Susanta Kumar Parui

Abstract Antenna array is very much needed for high gain, beam shaping wireless applications. In this paper, a two-element hemispherical dielectric resonator antenna (HDRA) for E-plane beam shaping has been studied. Microstrip line has been used for excitation of the antenna array element. The antenna offers impedance bandwidth of 6.25% and peak gain of 9.3 dBi.

Keywords HDRA · Antenna array · Peak gain · Impedance bandwidth · Beamwidth

1 Introduction

Since the 1960s, after the invention of low-loss ceramic materials, dielectric resonator (DR) has been used as a high Q-element in microwave circuits such as oscillators, filters, and so forth [1, 2] where DR is made of high permittivity material. After the first work published by Stuart et al. in 1983, the researcher started an investigation on DR as an antenna element [3]. A wide investigation is carried out and vivid update is given in two books [4, 5]. DRA gets immense attention by the researchers due to its attractive features such as low loss, low profile, wide impedance bandwidth, ease of fabrication, ease of excitation, high radiation efficiency, etc. DRA can be excited by all the feeding methods used in conventional patch antenna such as aperture coupled [6, 7], probe coupled [8, 9], microstrip line coupled [10, 11], conformal strip coupled [12], co-planar waveguide feed, and so on.

In many applications with single-element antenna, we cannot achieve desired specifications such as high gain, directional radiation patterns. To meet those demands antenna array is needed. A number of works are carried out on two-element antenna arrays [13–17]. All the studies were carried out is on cylindrical, rectangular, and

G. A. Sarkar (✉) · S. K. Parui
Department of Electronics and Telecommunication Engineering, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, West Bengal, India
e-mail: goffar.ali@gmail.com

triangular shape DRA element with H-plane beam shaping except [13], where E-plane beam shaping and H-plane beam shaping have been studied. Since then, as far as the authors knowledge, no one has investigated on two-element antenna arrays for E-plane beam shaping. In this paper, we have investigated a two-element antenna array on hemispherical shape geometry. Each HDRA antenna elements are excited using microstrip line and a 3 dB power divider is used for equal power distributions. The antenna structure is designed and simulated using Ansoft High-Frequency Structure Simulator based on the FEM method.

2 Antenna Array Design

2.1 Resonant Frequency

A HDRA can support *TE* and *TM* mode. For *TE* mode, E_r and $E_\emptyset$ vanish at $\theta = 90°$, whereas for *TM* mode E_r and $E_\emptyset$ vanish at $\theta = 0°$. For a hemispherical cavity, the resonant frequency of lowest order TE_{111} mode can be calculating using [18] and is represented as

$$f_r = \frac{4.775 \times 10^7 Re(kR)}{\sqrt{\epsilon_r}R} \tag{1}$$

2.2 Study of Coupling

Study of mutual coupling is very essential before designing an antenna array as it has a great impact on resonance frequency, impedance bandwidth, and radiation patterns. Owing to coupling effects, the resonant frequency can be shifted from its desired value, also radiation patterns may be degraded. If the distance between the antenna array element increases isolation increased, results in more directive radiation with increased sidelobe levels and vice versa. Here coupling is studied for a separation gap between two elements of $\lambda_0/2$ (λ_0 free space wavelength). The simulated coupling and reflection coefficient is depicted in Fig. 1.

2.3 Antenna Array Geometry

The authors' have investigated on single-element HDRA fed by microstrip line in [10] and avoided the discussions here. Instead, they have used the same single antenna element for designing a two element E-plane HDRA array and depicted in Fig. 2. The array is designed on Arlon AD270 (dielectric constant 2.7 and loss tangent 0.0023)

Fig. 1 Simulated S_{11} and S_{21}
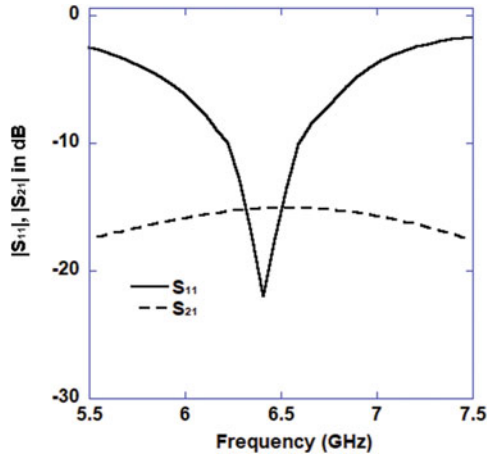
substrate of area L × W mm^2. The radiating element is made of Eccostock HiK materials. It has dielectric constant 20 and loss tangent 0.002. The design dimensions of the array are given in Table 1.

Fig. 2 Schematic diagram of DRA array
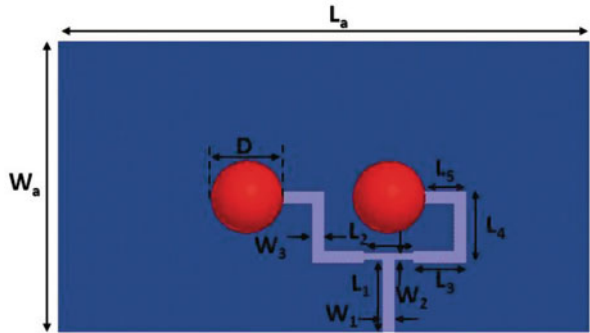
Table 1 Design parameters of array

Parameters	L_a	L_1	L_2	L_3	L_4
Value (mm)	89	5.935	7.77	8.68	18.055
Parameters	L_5	W_a	W_1	W_2	W_3
Value (mm)	5.7	48	2.13	1.2	2.13

2.4 Study of Coupling

The simulated reflection coefficient is presented in Fig. 3. From the figure, it has been observed that S_{11} minima occurred at 6.4 GHz. At S_{11} minima, the array offers 10 dB impedance bandwidth of 6.25% covering the frequency range of 6.2–6.6 GHz. The antenna array radiates in broadside direction as array elements are excited with fundamental mode TE_{111}. The electrical field distributions of fundamental modes are shown in Fig. 4. In Fig. 5a, b, simulated normalized radiation patterns are depicted for E-plane and H-plane, respectively. 3 dB beamwidth for E-plane and H-plane has been found as 50° and 78°, respectively. Peak gain in broadside direction has been observed as 9.3 dBi. A comparative study is carried out in a tabular format and is given in Table 2. From the table, it is clear that our proposed designed have comparable performance with others but DRA element consumes the smallest volume among all.

Fig. 3 Simulated S_{11} of the proposed array
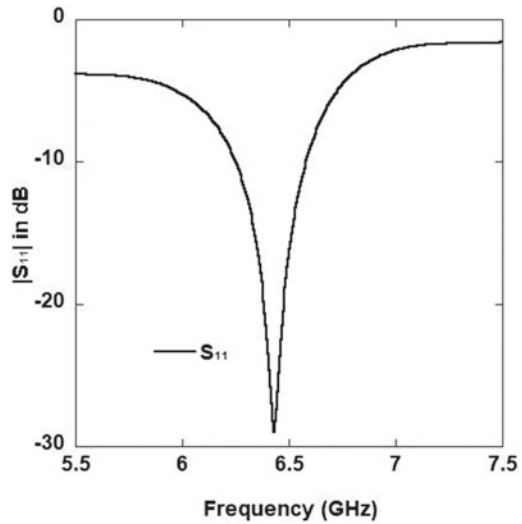
Fig. 4 Electric fields distributions for fundamental TE_{111}

(a)

(b)
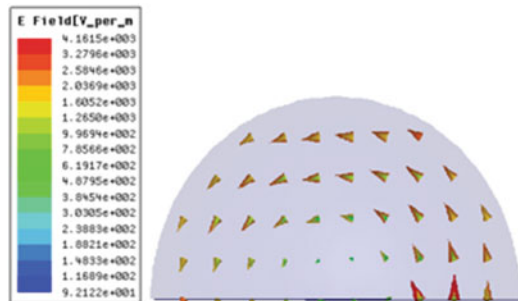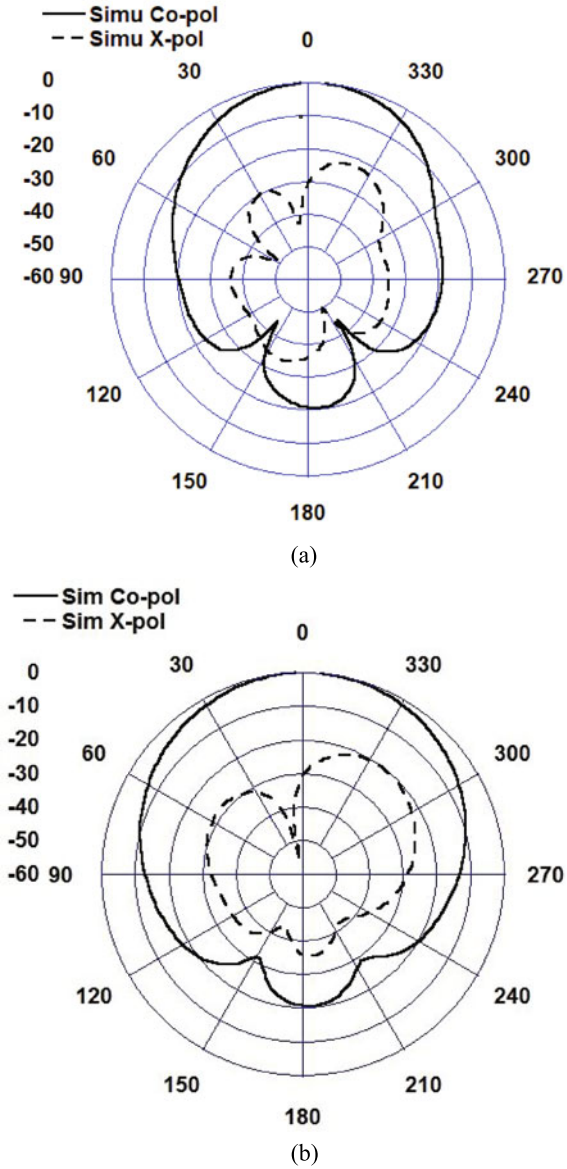
Fig. 5 Simulated normalized radiation patterns **a** for E-plane and **b** for H-plane

Table 2 Comparison with previous works

Sl. no.	Shape of DRA	Sing. element vol. (mm^3)	ε_r	BW (%)	Gain (dBi)	Beam shaping	Reference
1	Cylindrical	1662	9.5	–	8.9	E-plane and H-plane	[13]
2	Cylindrical	850	14.2	5.2	5.7	H-plane	[14]
3	Annular	4642	34.73	4.688	–	H-plane	[15]
4	Triangular	2382	9.8	11.7	4.18	H-plane	[16]
5	Hemispherical	409	20	6.25	9.3	E-plane	This work

3 Conclusion

A two-element linear array in E-plane for C-band wireless application has been investigated. The proposed array consists of two hemispherical shapes of DRA excited with microstrip line underneath the DRA elements. The simulated result illustrates that the antenna resonates at 6.4 GHz and provide 6.25% impedance bandwidth. The offers peak gain of 9.3 dBi. It has a penalty of application for a higher number of antenna array elements.

References

1. Cohn SB (1968) Microwave bandpass filters containing high-Q dielectric resonators. IEEE Trans Microw Theory Tech 16(4):218–227
2. Abe H, Takayama Y, Higashisaka A, Takamizawa H (1978) A highly stabilized low-noise GaAs FET integrated oscillator with a dielectric resonator in the C band. IEEE Trans Microw Theory Tech 26(3):156–162
3. Long S, McAllister M, Shen L (1983) The resonant cylindrical dielectric cavity antenna. IEEE Trans Antennas Propag 31(3):406–412
4. Luk KM, Leung KW (2003) Dielectric resonator antennas. Research Studies, Baldock
5. Petosa A (2007) Dielectric resonator antennas handbook. Artech House, Boston, London
6. Martin JS, Antar YM, Kishk AA, Ittipiboon A, Cuhaci M (1990) Dielectric resonator antenna using aperture coupling. Electron Lett 26(24):2015–2016
7. Sarkar GA, Rana B, Parui SK (2017) Dual polarized dielectric resonator antenna with high port isolation. In: Proceedings of 12th international conference on microwaves, antenna propagation and remote sensing (ICMARS 2017). Jodhpur, India, Feb 2017, pp 93–95
8. Sarkar GA, Rana B, Parui SK, Rahaman SE Port isolation enhancement of a dual polarized rectangular dielectric resonator antenna. In: Proceedings of 1st international conference on electronics, materials engineering and nano-technology (IEMENTech 2017). Kolkata, India, April 2017, pp 1–4
9. Maity S, Gupta B (2016) Experimental investigations on wideband triangular dielectric resonator antenna. IEEE Trans Antennas Propag 64(12):5483–5486

10. Sarkar GA, Rana B, Parui SK Port isolation enhancement of a dual polarized hemispherical dielectric resonator antenna. In: Proceedings of international symposium on antennas and propagation (APSYM 2016). Cochin, India, Dec 2016, pp 282–285
11. Sarkar GA, Rana B, Parui SK A direct microstrip line feed hemispherical dielectric resonator antenna array. In: Proceedings of 1st international conference on electronics, materials engineering and nano-technology (IEMENTech 2017). Kolkata, India, April 2017, pp 1–3
12. Huang CY, Chiou TW, Wong KL (2001) Dual-polarized dielectric resonator antennas. Microw Opt Technol Lett 31(3):222–223
13. Chow KY, Leung KW, Luk KM, Yung EK (1995) Cylindrical dielectric resonator antenna array. Electron Lett 31(18):1536–1537
14. Drossos G, Wu Z, Davis LE (1996) Two-element broadside arrays using cylindrical dielectric resonator antennas. Microw Opt Technol Lett 11(6):342–345
15. Aras MS, Rahim MK, Rasin Z, Aziz MA An array of dielectric resonator antenna for wireless application. In: Proceedings of international RF and microwave conference 2008 (RFM 2008). Kuala Lumpur, Malaysia, Dec 2008, pp 459–463
16. Kumari R, Behera SK A compact dual resonance dielectric resonator antenna array with partial ground plane. In: Proceedings of 1st international conference on recent advances in information technology (RAIT 2012). Dhanbad, India, March 2012, pp 809–812
17. Abdullah AS, Majeed AH, Sayidmarie KH, Abd-Alhameed RA (2014) Two elements elliptical slot CDRA array with corporate feeding for X-band applications. Iraq J Electr Electron Eng 10(1):48–54
18. McAllister MW, Long SA (1984) Resonant hemispherical dielectric antenna. Electron Lett 20(16):657–659

Design and Development of MIMO Radar

Prachi and Soumyasree Bera

Abstract RADAR works on the principle of reflection and need not have clear visibility to LOS, by using radio waves position, range, angle, or velocity can be determined based on received echo signal. Loss of signals and interference is more with SISO system. Therefore, MIMO system is used to minimize errors, improve resolution, and signal-to-noise ratio as they provide substantially improved immunity to interference and detection probability of the targets. This paper deals with the development of a MIMO radar system of the proposed design specification in LabVIEW (software) and with that, detection and characterization of the target are done and further comparison with the SISO radar system.

Keywords MIMO · SISO · LOS · RADAR · LabVIEW

1 Introduction

In this paper, it describes the range of waveform, required signal processing techniques, and antenna array configuration. It describes how multiple orthogonal signals from the transmitter elements are coherently combined on reception of a phased array to form multiple beams. Advantages and limitations of MIMO radar are also discussed [1].

In this paper, signal processing techniques for target detection and finding of different target parameters is described. It also describes the factors, which lead to the RCS spatial diversity and compares the architecture of MIMO radar with multiple-element radar. It describes signal modeling techniques with mathematical derivation, spatial decorrelation, multiple-antenna radar like multi-static radar, conventional phased array radar and adaptive radar and about static and moving target detection and finds the direction of the target [2].

Prachi · S. Bera (✉)
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: soumyasree.bera@gmail.com

In this paper, by using statistical MIMO radar, it describes the improvement in radar performances by creating spacing between the antennas at the transmitter and receiver side so that target angular spread is observed which exploits the spatial diversity of target scatterers and shows that how MIMO radar outperforms the conventional phased array radar whenever the detection probability is above 0.8. In MIMO radar, it is not restricted to detection only but the application of MIMO radar with angular diversity with multiple targets through coherent processing and rank analysis of channel matrix and signal processing modeling in different radars like phased array radar and MISO radar with their mathematical analysis [3].

In this paper, the adaptive techniques are introduced for MIMO radar imaging, waveform constraints are analyzed for ground moving target indication (GMTI) by using wide area surveillance MIMO radar, limitations on adaptive spatial clutter mitigation is observed in MIMO radar, velocity of the target is estimated and antenna placement is done for MIMO radar by using widely separated antennas, MIMO radar using compressive sampling and imaging of moving targets with multi-static SAR is performed and analyzed [4].

In this paper, the problem of MIMO radar is being considered by employing space–time coding (STC) to achieve diversity. So for that at the transmitter side, two criteria are implemented for code construction to establish the optimum achievable performance for MIMO radar systems. STC is used to establish spatial diversity and also contrasts the effect of fading [5].

In this paper, simultaneously transmission and reception of multiple coded waveforms are observed from multiple phase centers and to process them different signal processing is used which initially appears as spatial samples, which correspond to the convolution of the transmitted and received aperture phase centers. The samples provide the ability to measure the channel, which consists of the transmitted and received path, propagation, and the target. It describes the adaptive coherent beam formation technique, which is applied to the received signal to control transmit beam shape and improve track-while-scan accuracy [6].

In this paper, for target detection and localization space–time coding is used in which transmitted signals are space–time code, which allows to digitally steer the beam pattern in the transmitted signal in addition to the received signal, which helps to avoid beam shape loss. Virtual sensors are used to form narrower beams having lower sidelobes, which provide high performance in target detection, angular estimation, and resolution [7].

2 System Modeling

Here, an abstract model of a system is being developed in which each block represents different aspects of the system, which help to understand the working of the system (Fig. 1).

Here, in this proposed model, the polyphase code (P4 code) is developed in the software. Then, pulsed P4 code is developed at the baseband level of 75 kHz whose

Tx:

Rx:
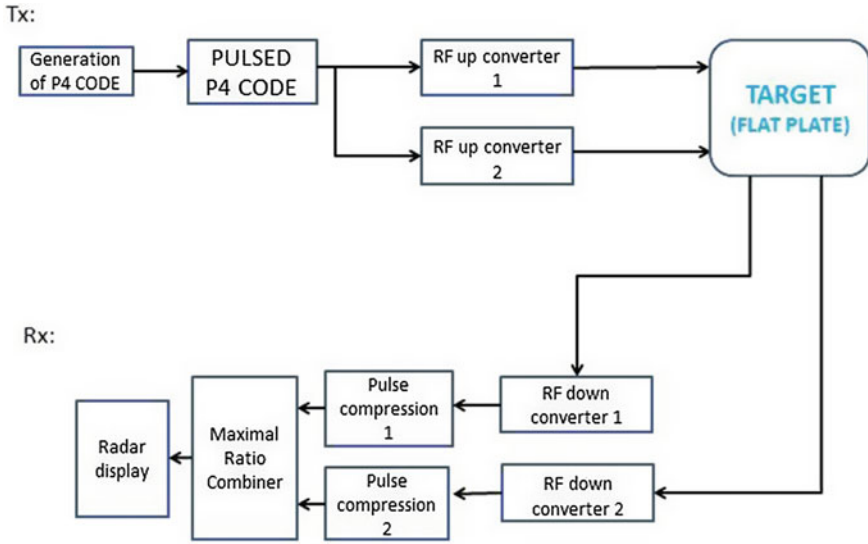
Fig. 1 Proposed block diagram of MIMO radar system

on-time bits and off-time bits are calculated, and the signal is up-converted and down-converted at RF frequency of 2.4 GHz. The upconverted signal is fed into the channel and inside the channel noise and delay is provided and target modeling is done which is dependent on the distance, which is provided by realizing the radar range equation. The output from the channel is fed to the down-converter and correlation in SISO and MRC (maximal ratio combiner) in MIMO is performed on the received signal with that of the reference signal. Then, target detection and characterization using a flat plate are observed and analyzed (Table 1).

2.1 Transmitter

Baseband Generation:

Polyphase code is used as a baseband signal that uses harmonically related phases based on phase increment [8]. The phase sequence of the P4 signal is given by

$$\Phi_i = \frac{\pi}{N}(i-1)(i-N-1) \tag{1}$$

where,

i = 1, 2, 3, …, N.
N = Compression ratio.

2.2 Channel

Channel consists of

(a) **Radar Range Equation** [9]:

$$Received\ power,\ P_r = \frac{P_t \cdot G_t \cdot G_r \cdot \lambda^2 \cdot \sigma}{(4\pi)^3 \cdot R^4} \tag{2}$$

where,

P_t = Transmit power
G_t = Transmit antenna gain
λ = Transmit wavelength
σ = Target radar cross section
G_r = Received antenna gain
R = Distance.

(b) **Target Modeling**:

Horizontal polarization of the target is given by [10]

$$\sigma 1H = \cos(k \cdot a \cdot sin\,\theta) + j\,\frac{\sin(k \cdot a \cdot sin\,\theta)}{sin\,\theta} \tag{3}$$

$$\sigma 2H = \frac{4 \cdot e^{j\left(k.a+\frac{\pi}{4}\right)}}{\sqrt{2\pi ka}} \tag{4}$$

$$\sigma 3H = \frac{e^{-j(k \cdot a \cdot sin\,\theta)}}{1 - sin\,\theta} \tag{5}$$

Table 1 Design specification of the MIMO radar

Sl. no.	Parameters	Specification
1	RF	2.4 GHz
2	Radar waveform	Spread spectrum
3	Unambiguous range	2 km
4	PRF	75 kHz
5	PRI	13.33 µs
6	Duty cycle	45%
7	Bandwidth	5 MHz
8	Length of P4 code	30 bits
9	Antenna configuration	2 × 2
10	Target type	Flat plate
11	Data acquisition	Using LabVIEW

$$\sigma 4H = \frac{e^{j(k \cdot a \cdot \sin \theta)}}{1 + \sin \theta} \tag{6}$$

$$\sigma 5H = 1 - \frac{e^{j\left(2 \cdot k \cdot a + \frac{\pi}{2}\right)}}{2\pi ka} \tag{7}$$

$$\sigma H = \frac{b^2}{\pi} \left| \sigma 1H - \sigma 2H \left[\frac{1}{\cos \theta} - \frac{\sigma 2H}{4} (\sigma 3H + \sigma 4H) \right] \sigma 5H^{-1} \right|^2$$

where $k = 2\pi/\lambda$, a = width of the target, b = height of the target.

2.3 Receiver Signal Processing

On the ith receive antenna, the received signal is

$$y_i = h_i x + n \tag{9}$$

where

y_i is the received symbol on the ith receive antenna,
h_i is the channel on the ith receive antenna,
x is the transmitted symbol and
n_i is the noise on ith receive antenna.

(a) **Correlator**:
Correlation is a measure of similarity of two series as a function of the displacement of one relative to the other [10].

$$X(\tau) = \int V_1(t) V_2(t - \tau) dt \tag{10}$$

(b) **Maximal-Ratio Combiner**:
Maximum-ratio combining (MRC) [11] is the method of combining the signal received by multiple receiver antenna to enhance signal-to-noise ratio (SNR) and thus limiting the impact of noise, interference, etc.

3 Results

Detection and characterization of the target using a flat plate is done (Fig. 2).
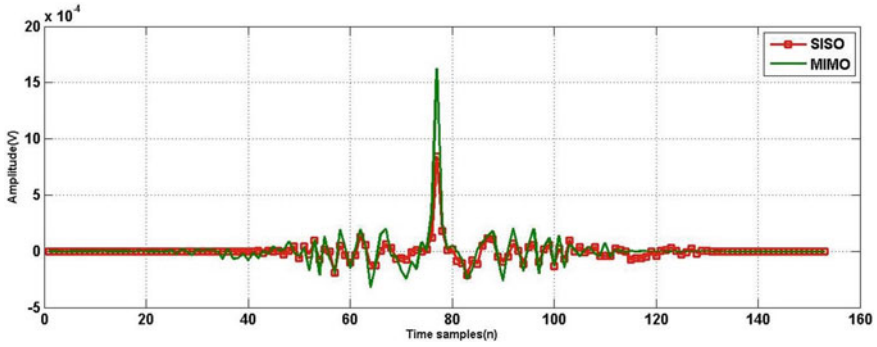
Fig. 2 Target detection
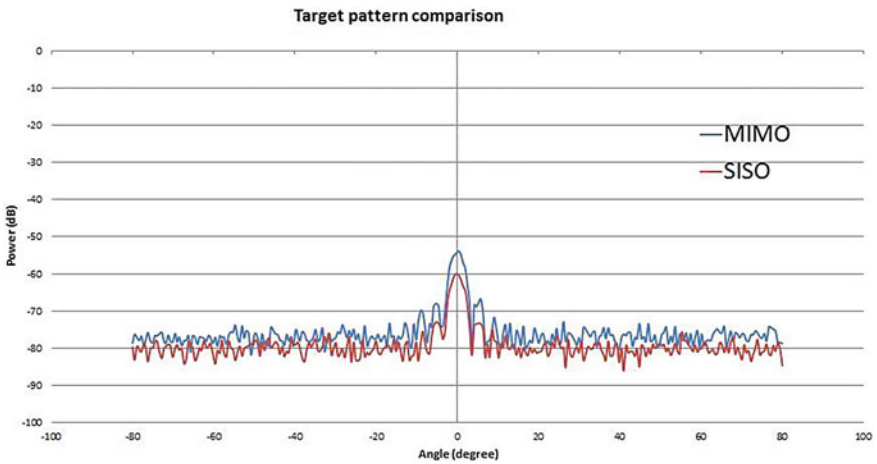
Fig. 3 Target characterization

3.1 Target Detection by Using SISO and MIMO

Here, correlation is performed on the received signal with that of the reference signal in SISO and MRC (maximal-ratio combiner) is performed in MIMO, which detects and determines the maximum similarity between the transmitted signal and the received signal (Fig. 3).

3.2 Characterization of the Target

Target characterization for the SISO as well as MIMO is shown above, which shows an increase in gain.

3.3 Conclusion

In the detection of the target by using SISO, the amplitude of the correlated signal is (8.21E−04) and by using MIMO the amplitude of the MRC is (1.65E−03). Hence, by using MIMO the amplitude is increased by (8.29E−04). Therefore, it concludes that the detection probability of the target is increased in case of MIMO.

In case of pattern characterization, the pattern of the target received by using SISO is having power of (−60.5 dB) and the pattern received by using MIMO is having power of (−54.4 dB), which means by using MIMO of 2×2 antennas 6 dB gain is achieved.

Hence, by using MIMO target detection and characterization are improved.

References

1. Donnet BJ, Longsta ID (2006) MIMO radar, techniques and opportunities. In: 3rd European radar conference, Queensland, Australia, pp 112–115
2. Haimovich AM, Blum RS, Cimini LJ (2008) MIMO radar with widely separated antennas. IEEE Signal Process Mag 25:116–129
3. Fishler E, Haimovich A, Blum R, Ciminio L, Chizhik D, Valenzuela R (2004) Performance of MIMO radar systems: advantages of angular diversity Eran. In: IEEE thirty-eighth Asilomar conference on signals,systems and computers, vol 1, pp 305–309
4. Li J, Blum RS, Stoica P, Haimovich AM, Wicks MC (2010) Introduction to the issue on MIMO radar and its applications. IEEE J Select Top-ICS Signal Process 4:2–4
5. Maio AD, Lops M (2007) Design principles of MIMO radar detectors. IEEE Trans Aerosp Electron Syst 43(3):886–898
6. Robey FC, Coutts S, Weikle D, McHarg JC, Cuomo K (2004) MIMO radar theory and experimental results. In: Proceedings of the 38th Asilomar conference on signals, systems and computers, vol 1, pp 300–304
7. Bekkerman I, Tabrikian J (2006) Target detection and localization using MIMO radars and sonars. IEEE Trans Signal Process 54:3873–3883
8. Lewis BL, Kretschmer FF Jr (1982) Linear frequency modulation derived polyphase pulse compression codes. IEEE Trans Aerosp Electron Syst AES-18(5):637–641
9. Skolnik M (2008) Radar handbook, 3rd edn, McGraw-Hill
10. Mahafza BR (2000) Radar systems analysis and design using MATLAB. Chapman & Hall/CRC
11. Inaba T, Araki K (2004) Space-frequency maximal ratio combining for low-elevation radar target. Electron Commun Japan (Part I: Communications) 87(5):75–85

Gain Enhancement of Microstrip Patch Using Different Array Configurations

**Arun Kumar Singh, Bansibadan Maji, Rabindranath Bera
and Riwas Gurung**

Abstract This paper discusses the design of X band patch antenna. The aim of the work is to enhance the gain and minimize the return loss as much as possible. Initially, the design was focused toward single-patch antenna, but later it was modified with an array of different configurations. For selection of substrate, the trial was made on different substrates and finally epoxy of dielectric constant 4.4 was chosen due to its readiness in the availability. All the antennas were made to resonate in X band. The motive behind choosing this frequency band is to replace the existing dish antenna used in DTH service with the proposed one. Dish used in DTH services are having a high gain of around 30–35 dB and to match this with patch antenna requires an array of larger size. However, in this paper, an array of size to 16 is shown and the work is still going on in order to increase the array size. The performance of the antenna is measured in terms of return loss, directivity, main lobe-to-side lobe ratio and radiation pattern.

Keywords Microstrip patch antenna · Return loss · Gain · Directivity · Dielectrics · Array configuration

A. K. Singh (✉) · R. Bera · R. Gurung
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of
Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: arunsingh.smit@gmail.com

R. Bera
e-mail: rbera50@gmail.com

R. Gurung
e-mail: riwasgurung1994@gmail.com

B. Maji
National Institute of Technology, Durgapur, Durgapur, West Bengal, India
e-mail: bansibadan.maji@ece.nitdgp.ac.in

1   Introduction

Wireless Communication is getting advanced day by day and the size of the communication device is getting reduced [1]. With the current device, there is a demand for multiple operations such as cellular operation, Bluetooth and WLAN of different variations within a single device. Since the size of the device should be small with less power consumption, the only antenna which matches with this characterization is patch antenna because of its low profile, easy to fabricate and cost-effectiveness [2–5]. However, there are disadvantages also associated with the patch antennas such as narrow operation bandwidth and low gain. Gain enhancement can be done with an increase in the number of arrays and the enhancement of bandwidth can be done by changing the patch shape, choosing proper dielectric constant and the optimizing the thickness of the substrate.

This paper deals with the concept of replacement of the dish antenna with an array of the patch antenna. The gain shown in the paper is around 14 dB.

Therefore, it is an attempt of entering into satellite communication domain without using conventional dish antenna.

A conventional patch antenna is shown in Fig. 1.
where

W   Width of the patch
L   Length of the patch
H   Height of the substrate.

2   Antenna Design

The desired antenna is designed in X band (8–12 Ghz). The design parameters are as mentioned in Table 1.

To design a patch antenna, the width and length of the patch antenna is to be calculated first. Since the antenna is of X band, therefore the dimensions can be calculated as [6]

Fig. 1 Microstrip patch antenna (*image courtesy* www.emtalk.com)
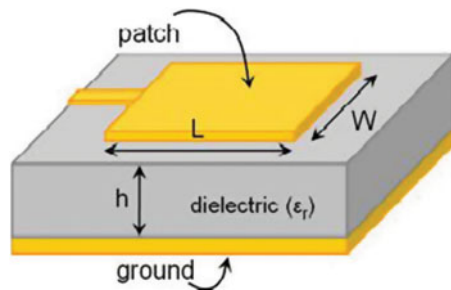
Table 1 Antenna design specifications

Variable	Value
Height of the patch (h)	1.6 mm
Effective dielectric constant of the patch (ε_{eff})	3.63
Width of patch	8.29 mm
Length of the patch	5.75 mm
Desired return loss	<-20 dB
Gain	Depends on array size
VSWR	In the range of 1–1.2

Fig. 2 Single-patch antenna
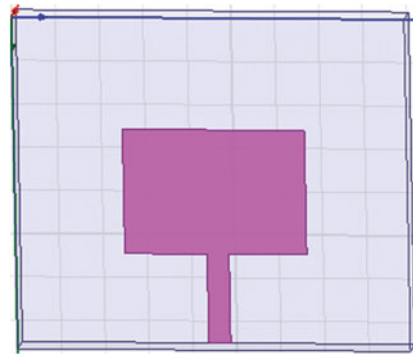
$$Width = \frac{C}{2F_r\sqrt{\left(\frac{\epsilon_r+1}{2}\right)}} \tag{1}$$

where

Fr resonating frequency.
C   speed of light.
ϵr permittivity of the substrate.

$$\epsilon_{eff} = \frac{\epsilon_r+1}{2} + \frac{\epsilon_r-1}{2}\left[1+12\frac{height}{width}\right]^{\frac{-1}{2}} \tag{2}$$

$$length = \frac{C}{2F_r\sqrt{(\epsilon_{eff})}} - 0.824\,height\left(\frac{(\epsilon_{eff}+0.3)(width/height+0.264)}{(\epsilon_{eff}-0.258)(width/height+0.8)}\right) \tag{3}$$

On calculation the dimensions were found to be 8.293 mm for width and 5.75 mm for length of the patch.

Based on the above calculation, the design of the single antenna is as in Fig. 2.
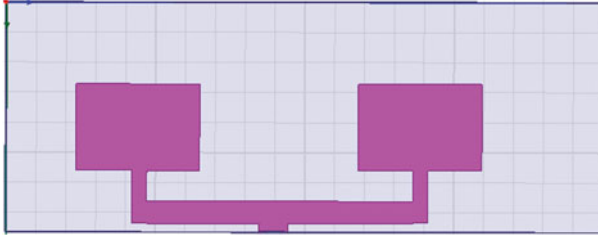
Fig. 3 Array antenna of dimension 1 × 2

(a)   **(b)**
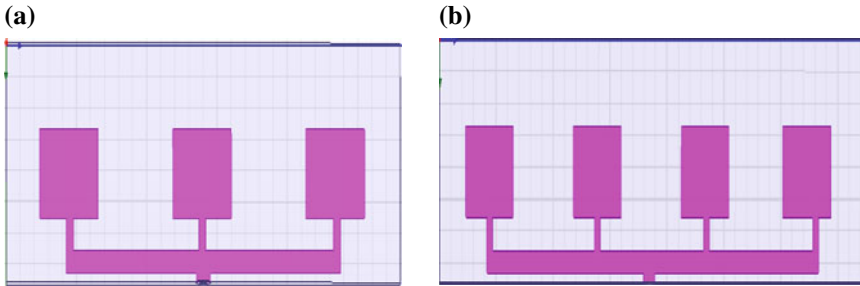
Fig. 4 **a** array antenna of dimension 1 × 3. **b** array antenna of dimension 1 × 4

Figure 1 shows the simulation model of single-patch antenna in which the length and breadth of the ground plane is calculated as

$$L_{ground} = (6 \times \text{height}) + \text{length} \tag{4}$$

$$W_{ground} = (6 \times \text{height}) + \text{Width} \tag{5}$$

And the dimension was found to be 15.35 mm for the length of the ground and 17.893 mm for the width of the ground.

In order to increase the directivity of the antenna, the single patch was modified to array design as shown in (Fig. 3).

The model shown in Fig. 2 is extended toward a linear array of size 1 × 2. The gain was increased by a considerable amount which is discussed in the later section.

Further, the model was extended toward a linear array of size 1 × 3 and 1 × 4 which is shown in Fig. 4.
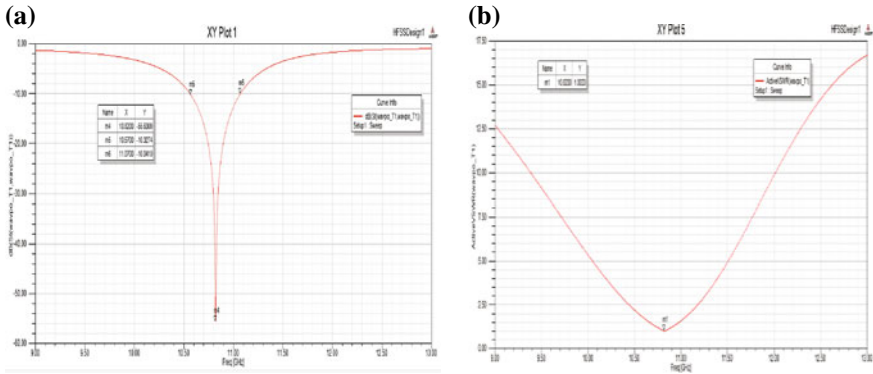
Fig. 5 a Return loss for single-patch antenna. **b** VSWR plot
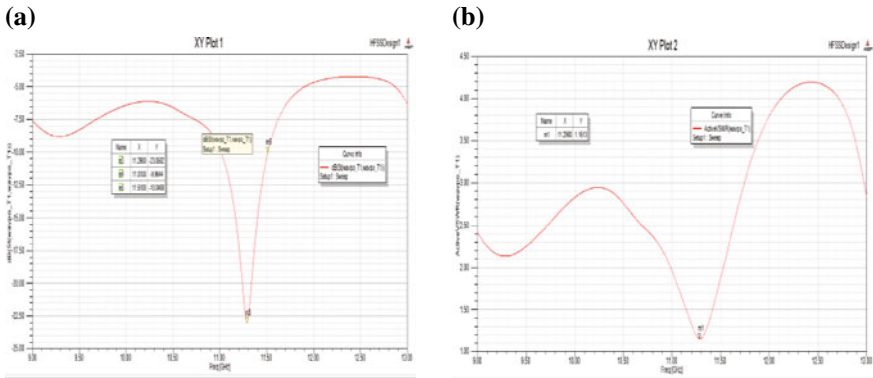
Fig. 6 a Return loss for 1 × 2 patch antenna. **b** VSWR plot

2.1   Results and Discussion

The simulated designs were analyzed in form of return loss characteristics, directivity, VSWR, and gain. More is the return loss better is the antenna.

Plots showing the results of return loss for different array configuration is shown in Fig. 5.

The obtained return loss in the simulation is around −55 dB and the VSWR was found to be 1.0033 at frequency 10.83Ghz (Fig. 6).

The obtained return loss in the simulation is around −23.05 dB and the VSWR was found to be 1.15 at frequency 11.29 Ghz (Fig. 7).

The obtained return loss in the simulation is around −26.43 dB and the VSWR was found to be 1.10 at frequency 11.30 Ghz (Fig. 8).

The obtained return loss in the simulation is around −33.07 dB and the VSWR was found to be 1.04 at frequency 11.59 Ghz (Figs. 9 and 10).

(a)

(b)
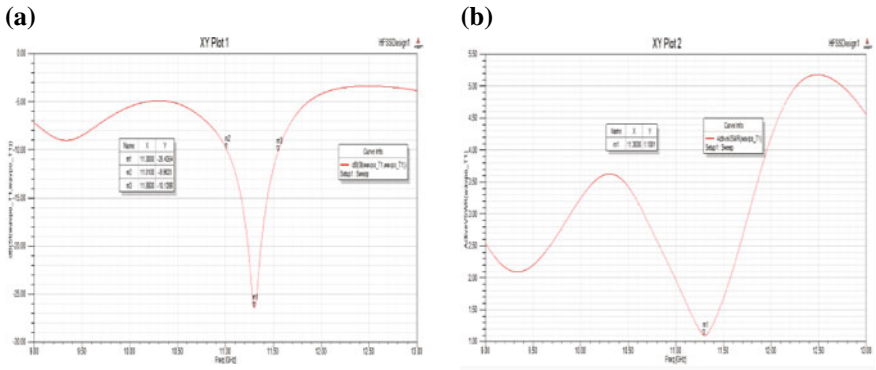
Fig. 7 **a** Return loss for 1×3 patch antenna. **b** VSWR plot

(a)

(b)
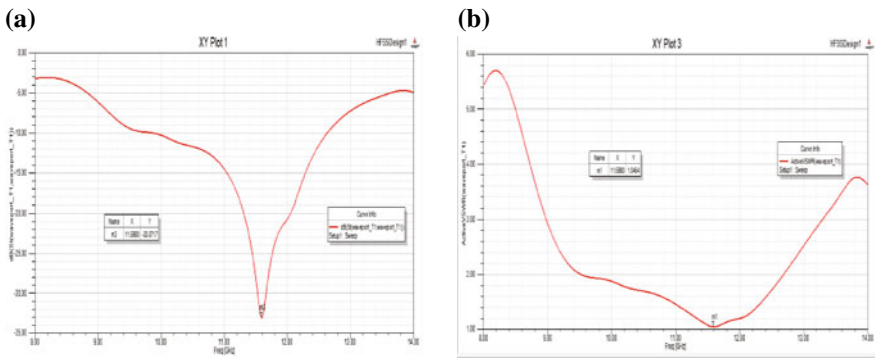
Fig. 8 **a** Return loss for 1×4 patch antenna. **b** VSWR plot

(a)

(b)
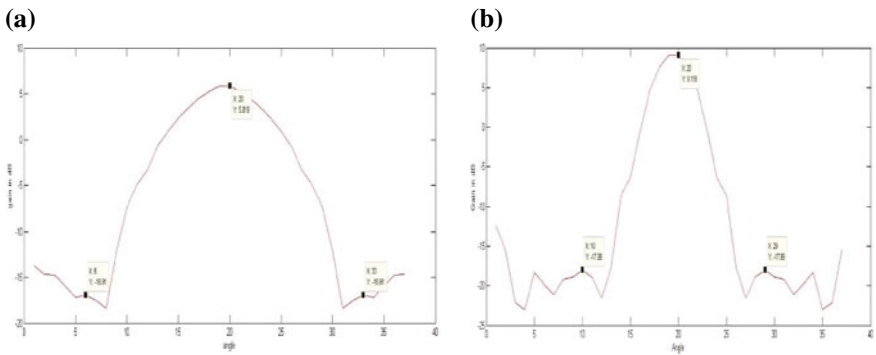
Fig. 9 **a** Antenna pattern of single patch. **b** Antenna pattern of 1×2 array
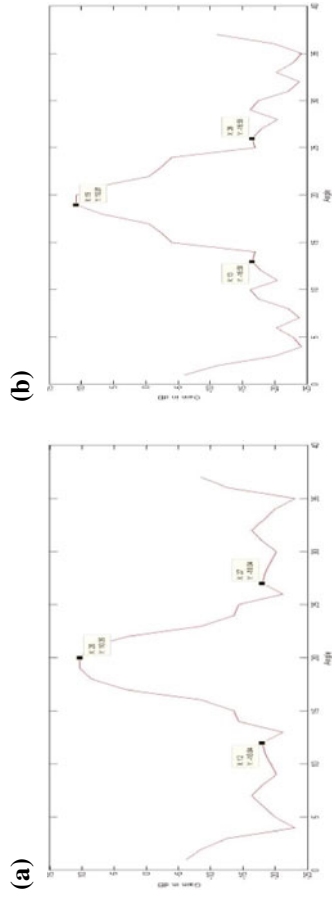
Fig. 10 **a** Antenna pattern of 1 × 3 array. **b** Antenna pattern of 1 × 4 array

Table 2 Result comparison of different array configuration

Sl. no.	Antenna configuration	Resonating frequency (GHz)	Return loss (dB)	Gain (dB)	VSWR	Main lobe to first sode lobe ratio (dB)
1	Single antenna	10.82	−55	5.809	1.0033	−16.91
2	1 × 2	11.29	−23.05	9.11	1.1513	−17.99
3	1 × 3	11.3	−26.45	10.35	1.1001	−18.04
4	1 × 4	11.59	−33.07	10.81	1.0454	−16.58

Antenna of different array sizes was designed, simulated and the results were shown in the above section. In Table 2 we can see the gain enhancement when the array size is increased. The obtained VSWR and return loss is also matching with the design specification given in Table 1. Since the antenna was designed to operate in X band, hence all the resonating frequencies were obtained in the range of that band, i.e., 8–12 Ghz.

3   Summary

The main aim of the paper was to increase the gain of the antenna. To do so work was carried out on an array antenna. As we can see in Table 2, the gain is increased in considerable amount when the array size was increased. Also, in the figures of the antenna pattern, the gain beamwidth relation is fairly observed. The beamwidth was more when the gain was less resulting in lesser directivity. However when the gain was increased beamwidth decreases and hence resulting in an increase of directivity.

References

1. Sheikh DH, Abdus KMS, Mohammed Khalid HJ, Rebeka S, Al Amin M (2016) A linearly polarized coaxial feeding dual band circular microstrip patch antenna for WLAN applications. Int J Wirel Microw Technol 3:50–60
2. George C, Cătălin M, Andrei K (2014) Design and implementation of microstrip patch antenna array. In: 2014 10th international conference on communications (COMM), Bucharest, Romania
3. Dasgupta S, Gupta B, Saha H (2010) Development of circular microstrip patch antenna array for rectenna application. In: 2010 annual IEEE India conference (INDICON), Kolkata, India, Dec 17–19
4. Chater N, Mazri T, Benbrahim M (2017) Design and simulation of microstrip patch array antenna for electronic scanning radar application. In: International conference on wireless technologies, embedded and intelligent systems (WITS), Fez, Morocco

5. Norfishah AW, Zulkifli BM, Wan NWM, Norhayati H (2010) Microstrip rectangular 4 × 1 patch array antenna at 2.5 GHz for WiMax application. In: Second international conference on computational intelligence, communication systems and networks, pp 164–168
6. Osman MAR (2006) Microstrip array antenna for automobile radar system. Master of electrical-telecommunication engineering master's thesis, Faculty of electrical engineering, Universiti Teknologi Malaysia, May

Implementation of a Feature-Adaptive Colour Image Copyright Protection Scheme

Subhadeep Koley, Manas Ranjan Nayak, Sanjeev Narayan Bal and Subir Kumar Sarkar

Abstract The increasing distribution of digital media over the Internet, and the simplicity of the digital systems have left the contents in the web vulnerable. With digital watermarking, we can impose ownership and copyright information inside digital media. In this paper, a phase congruency based digital colour image watermarking algorithm is proposed, which provides a higher degree of robustness, and excellent imperceptibility. Here, phase congruency has been used to detect the feature regions of the host image, and then the watermark has been infused into it using a novel technique called adaptive k-α blending. The coefficients of the k-α blending (k & α) are selected adaptively based on the phase congruency feature map of the host image. From rigorous testing, results indicate that our approach is robust against various geometric, non-geometric, and combined attacks.

Keywords Digital watermarking · Phase congruency · k- α blending

1 Introduction

The huge popularity of the Internet, as well as digital mobile devices such as digital cameras, smartphones, and hand-held scanners, have become an integral part of our daily lives. Documents generated by these devices and contents of the web are, in majority digital images. Therefore, it is of the utmost importance to provide own-

S. Koley (✉)
RCC Institute of Information Technology, Kolkata, West Bengal, India
e-mail: subhadeepkoley@gmail.com

M. R. Nayak · S. N. Bal · S. K. Sarkar
Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: nayakkmanass12@gmail.com

S. N. Bal
e-mail: s_n_bal@yahoo.co.in

S. K. Sarkar
e-mail: su_sircir@yahoo.co.in

ership authentication, and copyright protection scheme to a digital image. Digital watermarking is the most efficient way of implicating copyright measures to a digital image [1]. It is the process of impregnating an image with additional information that may be extracted in the future based on the unique key to establish ownership [2]. A watermarking scheme should be unprecedented in terms of robustness, and imperceptibility. However, the tradeoff between robustness, imperceptibility, payload capacity, and security is the reason for the ever emergence of research in the field of watermarking [3]. The two broad categories of watermarking are Spatial Domain watermarking, and Frequency Domain watermarking. Spatial domain watermarking is simple and fast in terms of computational complexity, whereas the frequency domain watermarking has added the advantage of excellent robustness [4]. However, nowadays Wavelet Domain techniques are gaining popularity because of the accurate Human Visual System (HVS) modeling [5]. An intensive survey of wavelet-based techniques can be found in Meerwald and Uhl [6]. We have incorporated the Integer-to-Integer Lifting Wavelet Transform (LWT) instead of conventional Discrete Wavelet Transform (DWT). Phase Congruency (PC) has been used to detect the spatial feature regions of the host image. Studies reveal that the HVS is more perceptible to phase distortions rather than those of amplitude. Kovesi [7] described that the phase congruency is a powerful, and efficient method for edge, and corner detection to combat the distortions in phase. For infusing the watermark into the cover image, we have proposed a novel method namely the *'Adaptive k-α Blending'*.

2 Watermark Embedding and Extraction

2.1 Feature Region Extraction

PC is the contrast invariant feature detection scheme, which is very significant in terms of edge, and corner detection. The problems with earlier gradient-based feature detection schemes were that they did not correctly detect or localize many of the important image features [7]. In addition, gradient-based detections are contrast, and illumination dependent [7]. The measure of PC is given by Kovesi [7] as follows

$$PC(x) = \frac{\sum_n V(x) \lfloor F_n(x) \big(cos\big(\psi_n(x) - \bar{\psi}_n(x)\big) - \big|sin\big(\psi_n(x) - \bar{\psi}_n(x)\big)\big|\big) - T \rfloor}{\sum_n F_n(x) + \delta}$$

(1)

The term $V(x)$ denotes the weights for frequency spread, as congruency over several frequencies is more fruitful than congruency over a single frequency. The symbol $\lfloor \rfloor$ denotes that the braced quantity is equal to itself when its value is positive, and equals to zero when negative. $F_n(x)$ is the filter pair at position x. δ is a small positive real-valued number prevents from division by zero. T is the threshold of

the estimated noise influence. By definition [7], we use the following equations for feature detection at every point of an image.

$$a = \sum (PC(\theta)\cos(\theta))^2 \tag{2}$$

$$b = 2\sum (PC(\theta)\cos(\theta)) \cdot (PC(\theta)\sin(\theta)) \tag{3}$$

$$c = \sum (PC(\theta)\sin(\theta))^2 \tag{4}$$

where $PC(\theta)$ denotes the phase congruency value determined at orientation θ. The angle of the principal axis is given in Eq. 5.

$$\varnothing = \frac{1}{2}atan2\left(\frac{b}{\sqrt{b^2 + (a-c)^2}}, \frac{a-c}{\sqrt{b^2 + (a-c)^2}}\right) \tag{5}$$

The minimum, and maximum moments, m, and M are given in Eq. 6, and Eq. 7, respectively.

$$m = \frac{1}{2}\left(c + a - \sqrt{b^2 + (a+c)^2}\right) \tag{6}$$

$$M = \frac{1}{2}\left(c + a + \sqrt{b^2 + (a+c)^2}\right) \tag{7}$$

The edge, and corner strength of the image are given by the maximum, and minimum moments, respectively [7]. Here, we have used only the edge map of the host image to obtain the feature regions.

2.2 Lifting Wavelet Transform

The Lifting Scheme is the second-generation wavelet transform introduced by Sweldens, which generates integer coefficients for all sub-bands unlike the conventional first-generation scheme, which generates floating point coefficients, that may be altered at any point of time during further processing [8, 9]. LWT offers less computational complexity, and reduced data truncation over first-generation DWT. Conventional lifting scheme consists of three steps namely, Divide, Dual lifting, and Primal lifting (Figs. 1 and 2).

1. Divide Operation: The sample image $I(i, j)$ is divided into two subsets, i.e., even sample set $I_e(i, j)$, and odd sample set $I_o(i, j)$ as described in Eqs. 8, and 9.

$$I_e(i, j) = I(i, 2j) \tag{8}$$

Fig. 1 Forward lifting
scheme
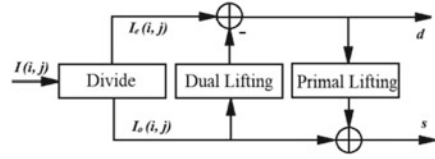
Fig. 2 Reverse lifting
scheme
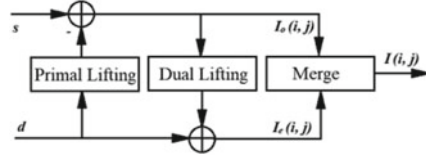
$$I_o(i, j) = I(i, 2j + 1) \tag{9}$$

2. Dual Lifting: The odd sample set $I_o(i, j)$ is predicted from the local neighbour-
   hood even coefficients using a prediction operator P. The error in odd sample
   prediction is used to generate the high-pass coefficients $h(i, j)$ as described in
   Eq. 10.

$$h(i, j) = I_o(i, j) - P[I_e(i, j)] \tag{10}$$

From Eq. 11, we can recover the odd sample set as shown below

$$I_o(i, j) = h(i, j) + P[I_e(i, j)] \tag{11}$$

3. Primal Lifting: The low-pass coefficients $l(i, j)$ is generated by updating the
   even sample with the updating value $U_h(i, j)$ as shown in Eq. 12.

$$l(i, j) = I_e(i, j) + U_h(i, j) \tag{12}$$

2.3 Encoder and Decoder

Let I be the original colour host image of size $m \times n$, having Red, Blue, and Green
channel of 8 bit each. The image is represented as $I = \{I_{red}, I_{green}, I_{blue}\}$.

$$I_{red} = X(i, j)|0 \leq I < m, 0 \leq I < n, X(i, j) \in \{0, 1, 2, \ldots, 255\} \tag{13}$$

$$I_{green} = Y(i, j)|0 \leq I < m, 0 \leq I < n, Y(i, j) \in \{0, 1, 2, \ldots, 255\} \tag{14}$$

$$I_{blue} = Z(i, j)|0 \leq I < m, 0 \leq I < n, Z(i, j) \in \{0, 1, 2, \ldots, 255\} \tag{15}$$

Also, W be the original grayscale watermark of size $p \times q$, and characterized as

$$W = V(i, j)|0 \leq I < p, 0 \leq I < q, V(i, j) \in \{0, 1, 2, \ldots, 255\} \qquad (16)$$

Figures 3 and 4 depict the watermark embedding, and extraction procedure, respectively. After detecting the local feature region using PC, we have come to the conclusion that the most congruent regions are the best ones to impregnate the watermark in. Here, we have proposed a modified version of the *'Cox's Algorithm'* [1], namely *'**Adaptive k- α blending**'*. Equation 17 and Eq. 18 describe Cox's algorithm, and the proposed algorithm respectively,

Fig. 3 Encoder block diagram
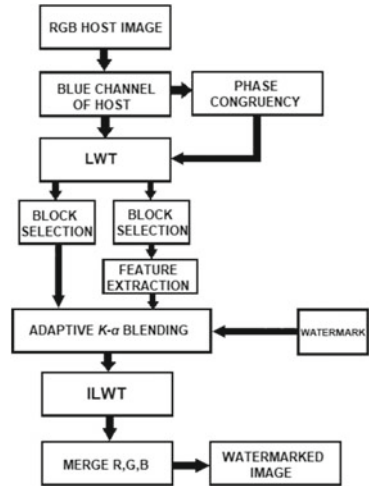
Fig. 4 Decoder block diagram
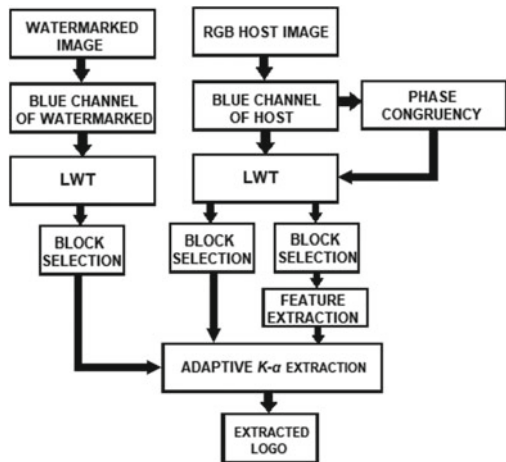
$$I'(i, j) = I(i, j) + \alpha \times W(i, j) \tag{17}$$

$$HH\{I'_{blue}(i, j)\} = k \times HH\{I_{blue}(i, j)\} + \alpha \times W(i, j) \tag{18}$$

In Eq. 18, $HH\{I_{blue}(i, j)\}$ is the 'HH' or the diagonal detail of the host image produced by LWT, $HH\{I'_{blue}(i, j)\}$ is the watermark embedded diagonal detail of the host image, $W(i, j)$ is the watermark, and 'k' and 'α' are the strength of the diagonal detail of the blue channel, and the strength of the watermark, respectively, which are dynamic with variation in values of the PC map of the host image. We have tested our algorithm with numerous values of 'k' and 'α', and it is observed that range given in Eqs. 19 and 20 provides the best result in terms of Peak Signal-to-Noise Ratio (PSNR).

$$k \in \mathbb{R} |\ 0.85 \le k \le 0.95 \tag{19}$$

$$\alpha \in \mathbb{R} |\ 0.010 \le \alpha \le 0.020 \tag{20}$$

Being a Non-Blind watermarking scheme, this algorithm uses the original host image to extract the embedded watermark. The extraction equation is given by

$$W(i, j) = \frac{\left(HH\{I'_{blue}(i, j)\} - [k \times HH\{I_{blue}(i, j)\}]\right)}{\alpha}. \tag{21}$$

3 Result Analysis

To test the imperceptibility, and the robustness of the proposed scheme, we have taken various standard 512×512 colour images, and a grayscale watermark of dimension 70×70. Figure 5 depicts 4 out of the 12 standard cover images we have used. Figures 6 and 7 depict the PC map, and the watermarked images respectively.

To prove imperceptibility of this scheme Table 1 is presented with various popular, and established quality metrics [4]. Metric values in Sl. No: 1–2, 4, 6–8, 11, 13–15

a) Lena b) Fruits c) Splash d) Gold hill
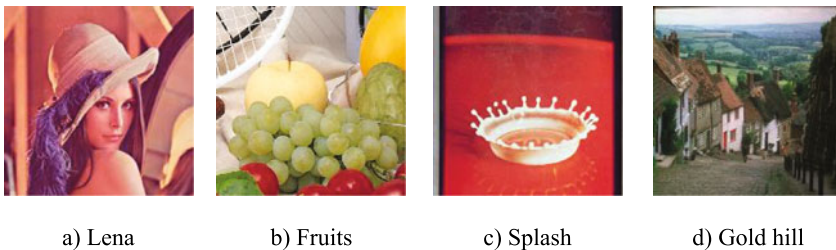
Fig. 5 Original colour cover images

| a) Lena | b) Fruits | c) Splash | d) Gold hill |
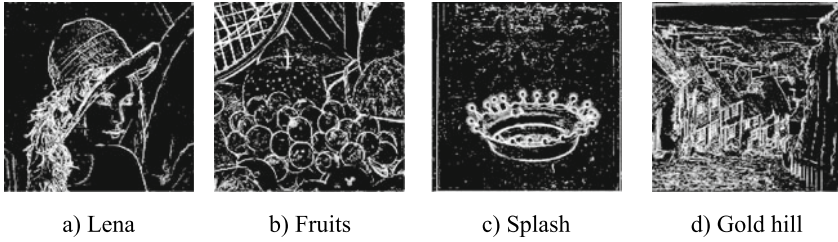
Fig. 6 Phase congruency maps of the blue channel

| a) Lena | b) Fruits | c) Splash | d) Gold hill |
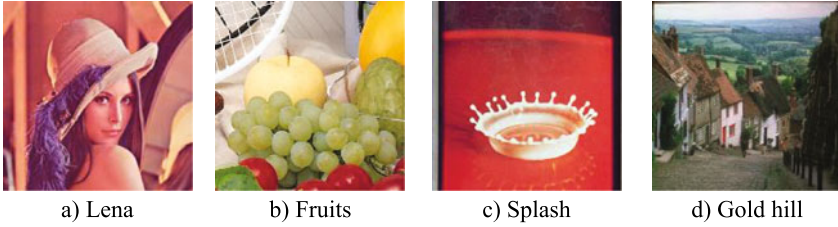
Fig. 7 Watermarked images

imply superior quality as higher metric value denotes better imperceptibility. On the other hand, metric values in Sl. No: 3, 5, 9–10, 12, also illustrate that the hidden logo is invisible as lower metric value denotes less visibility. Hence, the proposed algorithm is proven to be fruitful in every test of imperceptibility. 'Robustness' or 'Ruggedness' refers to the capability of the watermarking algorithm to correctly detect the watermark in the cover image after common geometric, and signal processing impairments [1]. To put the algorithm to test, we have used some well-accepted metrics, i.e.: *Normalized Cross Correlation (NCC), Structural SIMilarity Index (SSIM), Pearson Cross Correlation (PCC), Mutual Information (MI), and Bit Error Rate (BER).* From the various metric values from Table 2, it is evident that stego can sustain the watermark even after strong signal processing, and geometric impairments (Fig. 8).

NCC has been taken as the primary metric to test our approach. In most of the cases, we can see that the NCC value is 1 or almost 1 (Figs. 9 and 10). Other metric values like SSIM, MI, BER are also in support of the superior ruggedness of the algorithm.

In the given scheme, the 24-bit RGB cover image size is 512×512, and the 8-bit grayscale watermark image is of size 70×70. Therefore, embedding payload is

$$(70 \times 70 \times 8)/(512 \times 512 \times 3) = 0.04984(Bit/Pixel)$$

Table 1 The result of various imperceptibility metrics

Sl. no.	Metrics	Value
1	Signal-to-noise ratio	43.8871 dB
2	Peak signal-to-noise ratio	53.9408 dB
3	Mean square error	0.262421
4	Structural SIMilarity Index	1.0000
5	Average difference	0.043922
6	Structural content	1.001307
7	Universal image quality index	0.9999
8	Normalized cross correlation	0.99327
9	Normalized absolute error	0.000795
10	Laplacian mean square error	0.007323
11	Image fidelity	0.9999
12	Entropy difference	0.0015
13	Visual information fidelity	0.9983
14	Image fidelity criterion	7.25793
15	Pixel-based visual information fidelity	0.9900

Table 2 The result of various robustness metrics

Sl. no.	Attacks	NCC	SSIM	PCC	MI	BER
1	No attack	1.0000	1.0000	1.0000	0.8452	0
2	Histogram Eq.	1.0000	1.0000	1.0000	0.8452	0
3	LSB invert	1.0000	1.0000	1.0000	0.8452	0
4	Intensity	1.0000	1.0000	1.0000	0.8452	0
5	Median filtering	1.0000	1.0000	1.0000	0.8452	0
6	Sharpening	0.9686	0.5320	0.9686	0.7854	0.0400
7	Quantization(3 Bit)	0.2628	0.2221	0.2422	0.0678	0.5920
8	Complement	1.0000	1.0000	1.0000	0.8452	0
9	Erode	1.0000	1.0000	1.0000	0.8452	0
10	Dilate	1.0000	1.0000	1.0000	0.8452	0
11	Gaussian noise	0.9382	0.5319	0.9382	0.7133	0.2400
12	Resize (Factor 0.5)	1.0000	1.0000	1.0000	0.8452	0
13	Translation	1.0000	1.0000	1.0000	0.8452	0
14	Cropping (70%)	1.0000	1.0000	1.0000	0.8452	0
15	JPEG 2000 (Q = 20)	0.9225	0.5318	0.9225	0.6905	0.2427
16	Print and Re-Scan	0.4380	0.5563	0.4380	0.1278	0.2562
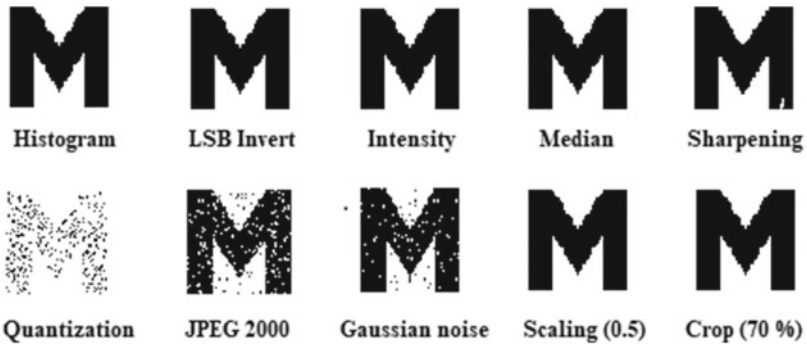
Fig. 8 Extracted watermark against various attacks

Fig. 9 NCC variation with increasing noise
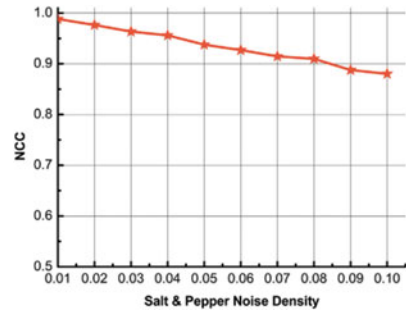
Fig. 10 NCC variation with rotation
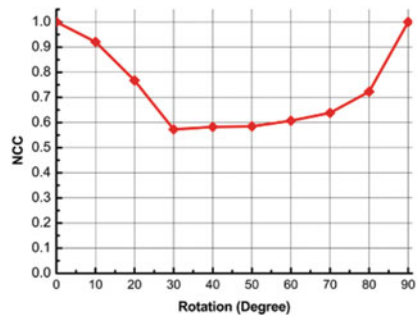
To justify the superiority of the proposed framework, a comparison with some other well-established methods has been drawn in Table 3. Among PC feature-based models, and other models the proposed approach gives the highest PSNR as well as SSIM.

Table 3 Comparison with other models

Sl. no.	Method	PSNR	SSIM
1	Proposed scheme	53.94	1.00
2	Hardware implementation of phase congruency and SVD-based watermarking [10]	45.23	1.00
3	Phase congruency and SVD factorization [11]	41.17	–
4	Visual saliency-based watermarking approach [12]	42.47	–
5	Visual attention model based LSB replacement [4]	42.27	0.949
6	Reversible watermarking based on integer-to-integer wavelet transform [8]	44.34	–

4 Conclusion

A wavelet domain, feature-based, reversible watermarking framework has been presented in this paper. The technique impregnates the watermark by modulating the diagonal detail coefficients of the blue channel of the cover image with the factors *'k'* and *'α'* according to the phase congruency feature map. Experimental results establish the fact that the projected approach has excellent imperceptibility, and is robust against most forms of signal processing impairments. In the near future, we aim to make the algorithm faster for effective hardware implementation.

References

1. Cox IJ, Kilian J, Leighton T, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6(12):1673–1687
2. Mohanty SP (1999) Digital watermarking: a tutorial review. Accessed from http://www.itb.ac.id/
3. Nayak MR, Tudu B, Basu A, Sarkar SK (2015) On the Implementation of a secured digital watermarking frame work. Informat Secur J A Global Perspect 00(1):1–9
4. Basu A, Sarkar SK (2013) On the implementation of robust copyright protection scheme using visual attention model. Informat Secur J A Global Perspect 22(1):10–20
5. Reddy AA, Chatterji BN (2005) A new wavelet based logo watermarking scheme. Pattern Recogn Lett 26(7):1019–1027
6. Meerwald P, Uhl A (2001) A survey of wavelet-domain watermarking algorithms. In Proceedings of SPIE, electronic imaging, security and watermarking of multimedia contents III, CA, USA 4314, pp 505–516
7. Kovesi P (1996) Invariant measures of image features from phase information (Ph.D. Dissertation). University of Western Australia, Perth, Western Australia
8. Lee S, Yoo CD, Kalker T (2007) Reversible image watermarking based on integer-to-integer wavelet transform. IEEE Trans Informat Forensics Secur 2(3):321–330
9. Sweldens W (1998) The lifting scheme: a construction of second generation wavelets. Soc Indust Appl Math 29(2):511–546
10. Nayak MR, Bag J, Sarkar S, Sarkar SK (2017) Hardware implementation of a novel water marking algorithm based on phase congruency and singular value decomposition technique. Int J Electron Commun (AEÜ) 71(2017):1–8

11. Wang W, Wang J, Wang C, Pu D (2010) A novel watermarking algorithm based on phase congruency corner detection and SVD. In: Fifth international conference on frontier of computer science and technology
12. Tian L, Zheng N, Xue J, Li C, Wang X (2011) An integrated visual saliency-based watermarking approach for synchronous image authentication and copyright protection. Signal Process Image Commun Elsevier 26(8–9):427–437

Design of Phased MIMO Radar and Its Performance Analysis Based on Different Beamforming Algorithms

Subhankar Shome, Samarendra Nath Sur, Soumyasree Bera, Debasish Bhaskar, Bansibadan Maji and Rabindranath Bera

Abstract A new approach of multiple-input multiple-output (MIMO) radar is proposed with accumulated antennas, which is called as Phased-MIMO radar. This technique is a bridge between MIMO and phased array, which is basically trying to take advantages of both. This new hybrid technique, MIMO radar can add diversity gain without sacrificing the coherent processing gain at the transmitting side which is the main advantage of phased-array radar. In this technique, the array of phased array system is divided into a numbers of subarrays, each subarray is assigned to coherently transmit or receive an orthogonal waveform. To form a beam towards the desired signal in assured direction weight vector is formulated for each subarray, using this technique coherent processing gain is achieved. Every subarray is combined to form MIMO radar, finally resulting in higher resolution capabilities. Our new proposed Phased-MIMO radar is designed using LCMV and MVDR adaptive beamforming technique, a comparative analysis is presented and simulated results are also shown for corresponding beampattern.

1 Introduction

Radar technology is invented around 90 years back around 1930 during the World War for defense applications [1]. Nowadays radars are used in several fields of society. Considering the growth in the communication field, radar is also upgraded day by day, and incorporating the latest technologies like, Phased-Array, MIMO, Massive

S. Shome (✉) · S. N. Sur · S. Bera · D. Bhaskar · R. Bera
ECE Department, Sikkim Manipal Institute of Technology, Majhitar, Rangpo 737136, Sikkim, India
e-mail: subho.ddj@gmail.com

R. Bera
e-mail: rbera50@gmail.com

B. Maji
ECE Department, National Institute of Technology, Durgapur, Durgapur, West Bengal, India
e-mail: bansibadan.maji@ece.nitdgp.ac.in

MIMO, which is already established in the era of communication. In this paper, the authors approached two well-established communication technologies to be merged in for modern radar. A phased array technology is well defined, and established which is offering coherent processing gain using array antennas. The advantage of array antenna includes beamforming as another advantage of this system. Without any mechanical parts moving, we can scan an area using beamforming, which is very much helpful in radar domain. The array's gain within the beam is very high, which improves radar target detection, but phased array system has some disadvantages also, it suffers from minimal diversity gain, which is an advantageous point for MIMO system which is also a well-established technology today. The proposed system is trying to explore all the advantageous issues of Phased Array and MIMO Technologies, which can be useful for the radar system to improve target detection, good angle estimation accuracy, multi-target detection accuracy, and many more. This modern approach is named as Phased-MIMO radar [2], as phased array and MIMO technologies [3] are merged and used in a new direction. Beamforming technology is a vital part of this new approach.

Array antennas are used in adaptive beamformation technique, even also used in hybrid beamforming technique to achieve maximum reception in the desired direction of user or target, the same frequency signals from other directions are rejected. Weights of each antenna element of the array will vary to achieve adaptive beamformation. The main requirement to design this type of beamforming technique is signal- processing capability to optimize the antenna beam in the desired direction of the target. Two well-known beamforming algorithms namely, LCMV (Linearly Constrained Minimum Variance Beamformer) [4] and MVDR (Minimum Variance Distortionless Response) [5] are implemented for Phased MIMO Radar, and performance is analyzed with reference to the array pattern response.

In this paper, the authors have formulated and simulated Phased MIMO based Radar using MVDR- and LCMV-based beamforming algorithm, and the performance is analyzed with respect to array antenna pattern in the desired direction and also with respect to BER plot.

2 Phased MIMO Radar Model

The number of array antennas is divided into subpart, to formulate sub-array for Phased MIMO Radar. In this paper, K number of subarrays is formed, which is overlapped. Number of elements can be 1–8 for each subarray. $\phi_k(t)$ signal coherently transmitted using all elements of kth subarray.

The signal transmitted by the kth subarray can be modeled as

$$s_k(t) = \sqrt{\frac{M}{K}} \varnothing_k(t) w_k^*, \quad \text{where } k = 1, \dots K \qquad (1)$$

Target reflected signal at θ direction can be modeled as

$$r(t, \theta) \triangleq \sqrt{\frac{M}{K}} \beta(\theta) \sum_{k=1}^{K} w_k^H a_k(\theta) e^{-j\tau_{k(\theta)}} \varnothing_k(t) \tag{2}$$

Here, complex vector of bemaforming weight can be defined as w_k and $a_k(\theta)$ is denoted as steering vector for kth subarray. Reflection coefficient of target is defined by $\beta(\theta)$. The time required for the wave to travel between the first subarray and the kth subarray is denoted by $\tau_{k(\theta)}$.

3 Linearly Constrained Minimum Variance (LCMV)

This algorithm is not suitable for every case of adaptive beamforming applications as the useful signal power may be unknown most of the time. Certain phase and gain of the desired signal allow going through in LCMV algorithm. Signals whose direction are different from the useful signal coming in is preventing as non-desired signal considered as disturbances. We can acquire the weights of LCMV beamformer by doing this peaking and nulling process along with minimizing the output power of the beamformer. So LCMV beamformer has the advantage of preserving the useful signals and suppressing the disturbances coming from the other directions.

Time signal received by the nth sensor is declared by $x_m(t)$, m = 1, 2, … M, where M is the number of elements and J is tap per element [6].

$$x_m(t) = s(t - \tau_m) + n_m(t), \quad \text{where m} = 1, 2 \dots M \tag{3}$$

s(t) is the desired signal from the specific look direction, τ_m denotes propagation delay of difference desired signal at the mth element, the total interference and noise observed at the mth sensor *is denoted by* n_m(t).

Conventional beamforming weight can be calculated using the steering vector [7].

$$a(\theta) = \begin{bmatrix} 1 \\ exp\left(j\frac{2\pi}{\lambda} sin\,\theta_i d\right) \\ exp\left(j\frac{2\pi}{\lambda} (sin\,\theta_i)(m-1)d\right) \end{bmatrix} \tag{4}$$

where d is the space between the elements of the antenna and θ_i is the desired angle, and m number of elements.

The optimization of weights done using Frost algorithm [8] is formulated as

$$\underset{W}{min} P = \underset{w}{min} W^T R_{xx} W \quad \text{Subject to } C_0^T W = F \tag{5}$$

$$R_{xx} = E\left[X(n)X(n)^T\right] \tag{6}$$

where R_{xx} is the array correlation matrix. C_0 is constraint matrix

$$C_0 = [c_1 ... c_j ... c_J] \tag{7}$$

$$c_j = [\underbrace{0...0}_{(j-1)K0_s} \underbrace{1...1}_{K1_s} \underbrace{0...0}_{(J-j)K0_s}]^T, \quad j = 1, ... J. \tag{8}$$

In the proposed LCMV, we have considered K = 10 and J = 1, so C_0 = [1111111111]

F can be described as follows:

$$
\overset{C^T}{\begin{bmatrix} 1\,1\,0\,0\,0\,0\,0\,0 \\ 0\,0\,1\,1\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1\,1 \end{bmatrix}}
\overset{W}{\begin{bmatrix} w_0 \ w_1 \\ w_2 \ w_3 \\ w_4 \ w_5 \\ w_6 \ w_7 \end{bmatrix}}
=
\begin{bmatrix} w_0 + w_1 \\ w_2 + w_3 \\ w_4 + w_5 \\ w_6 + w_7 \end{bmatrix}
=
\overset{F}{\begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix}}
\tag{9}
$$

Finally, using Lagrange multipliers method, W_{opt} is the optimal weight vector of Eq. (5) which is given by

$$W_{opt} = R_{xx}^{-1} C_0 \left(C_0^T R_{xx}^{-1} C_0 \right)^{-1} . F \tag{10}$$

The beamformer optimal output power is given by

$$P_{opt} = F^T (C_0^T R_{xx}^{-1} C_0)^{-1} F \tag{11}$$

4 Minimum Variance Distortionless Response (MVDR)

A consistent beamforming response in the direction of the useful signal which is called the desired direction, LCMV algorithm becomes MVDR algorithm. In the MVDR algorithm, we only need to know the direction of the desired signal, we do not have to know the direction or power of the interfering signal, noise anything. The following expression (12) can be expressed as MVDR algorithm [9, 10].

$$\min P(W) = \min(W^H R_{xx} W) \tag{12}$$

$$W^H a(\theta_0) = 1 \tag{13}$$

The gain of the useful signal equals to one in the direction of the desired signal, ensuring the minimum output power of MVDR beamformer.

The optimization weight vector can be expressed by the following formula:

$$W_{MVDR} = \frac{R_{xx}^{-1} a(\theta_0)}{a^H(\theta_0) R_{xx}^{-1} a(\theta_0)} \tag{14}$$

5   Simulation Work

In our simulations, all transmitter–receiver elements are placed in a way to form Uniform Linear Array. M = 10 number of transmit elements and N = 10 number of receiver elements are considered. K = 5 subarrays which are assumed to be fully overlapped to get the essence of Phased MIMO formation. Additive noise which has identical variances in each array element is modeled using Complex Gaussian zero-mean spatially and temporally white random sequence. We assume two interfering target sources are placed in the direction of $-10°$ and $+30°$. The desired target of interest is assumed to place and the desired signal is coming from the direction of $-20°$ to calculate the transmit beam pattern and diversity beam pattern. An overall beam pattern example is simulated considering another desired signal from $0°$ angle. In this simulation, conventional beamforming is considered to calculate the transmit beampattern, also a comparative outcome is shown based on adaptive beamforming for Phased MIMO Radar using MVDR and LCMV algorithm. Output SINR analysis is also done based on MVDR and LCMV algorithm for adaptive beamforming.

6   Results

Figure 1 shows the transmit beampattern and Fig. 2 shows the diversity beampattern for Phased MIMO radar. Figure 3 is showing the overall beampattern for Phased MIMO Radar using LCMV algorithm and Fig. 4 shows the comparison between LCMV and MVDR algorithm applied to calculate overall beampattern for Phased MIMO Radar. Figure 1 showing phased-array radar has the typical conventional beampattern with main lobe (of width $\pi = M$) centered at $\theta s = -20°$. This figure also shows, this proposed technique achieved almost 10 dB transmit diversity gain, which is absent in MIMO Radar. This is an advantage of combining Phased Array and MIMO system in one; here we are exploring the advantage of phased array system.
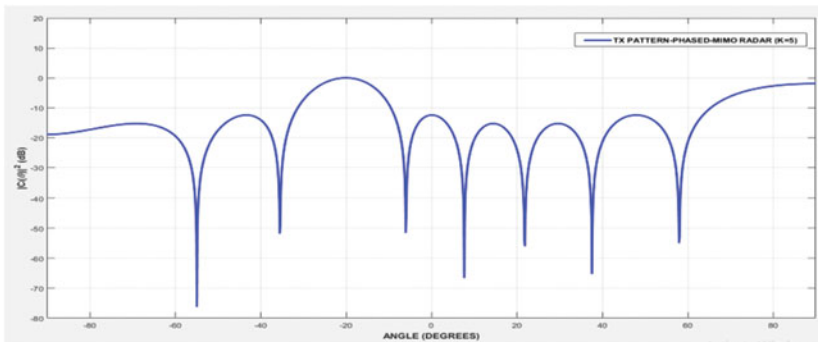
Fig. 1   Transmit beampattern using conventional beamforming for phased MIMO radar
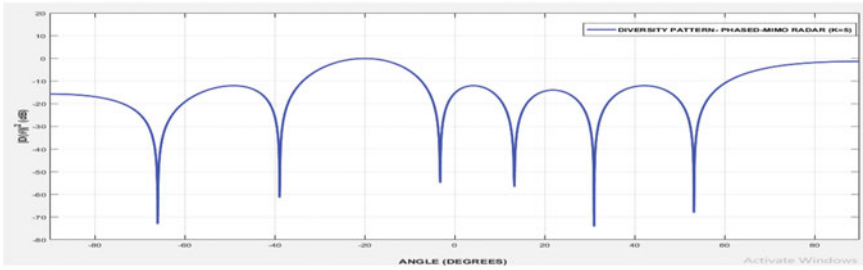
Fig. 2 Diversity beampattern using conventional beamforming for phased MIMO radar
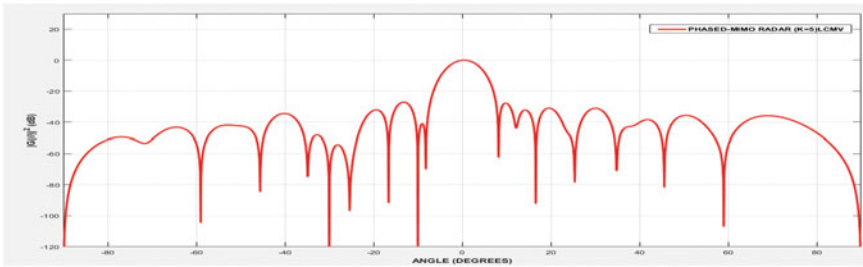
Fig. 3 Overall beampattern for phased MIMO radar using LCMV algorithm
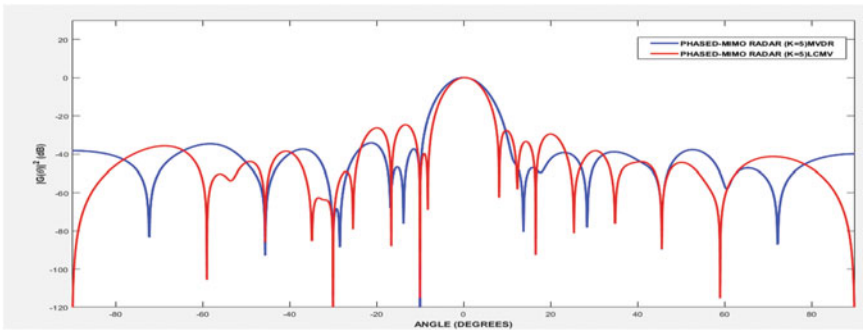
Fig. 4 Comparison of LCMV and MVDR for phased MIMO radar

Figure 2 is showing diversity beampattern in Phased MIMO Radar; here also we achieved around 12 dB diversity gain in Phased MIMO Radar. This parameter is absent in case of a Phased array, the Phased array is suffering from diversity gain, so here is another advantage to combining two technique in a single system.

Figure 3 is showing the overall beampattern using LCMV algorithm, which has a significant improvement in comparison to previous results; here we have achieved almost 20 dB dynamic gain using adaptive beamforming. In this case, beampattern is created with the main lobe at $\theta s = 0°$ which is the direction of the desired signal.
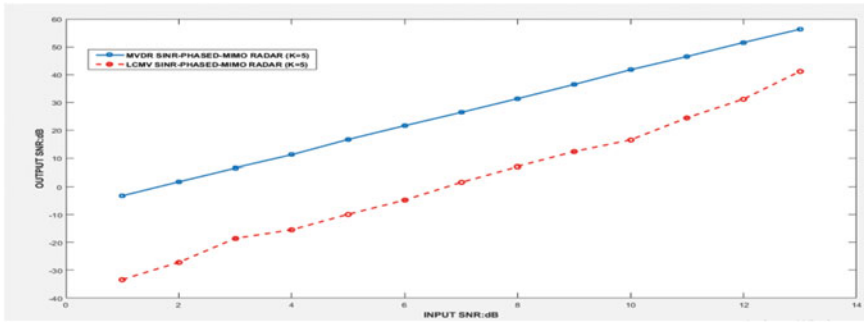
Fig. 5 Input versus output SNR for LCMV and MVDR beamforming for phased MIMO radar

Figure 4 is showing the overall beampattern using two different adaptive beamforming algorithms, one is LCMV and another is MVDR. Both the techniques have some advantages and disadvantages in different scenarios. In our case, the main lobe to side lobe formation is little better in LCMV algorithm.

Figure 5 we have shown the Input–Output SNR analysis for Phased MIMO Radar, from the graph we can see the SNR response is quite better in case of MVDR algorithm, as the desired signal direction is defined for MVDR.

7 Conclusion

The new system simulated for MIMO Radar with an array antenna, this array antenna are forming different subarray, each subarray is forming different beams finally forming a MIMO effect. The beam is forming toward a certain direction in space for which weight vector of each subarray is designed. In our system, Transmit beampattern and diversity gain is shown which results in coherent processing gain. In Phased MIMO Receiver, two different adaptive beamforming is applied namely LCMV and MVDR and a comparison is done in the form of the beam pattern. Input versus Output SNR is also measured, where MVDR is showing better performance in simulated Radar, as the desired signal direction is predefined.

References

1. Skolnik MI (2001) Introduction to radar systems, 3rd edn. Mc-Graw-Hill, New York
2. Hassanien A, Vorobyov SA (2010) Why the phased- MIMO Radar outperforms the phased array and MIMO radars. In: 18 European signal processing conference (EUSIPCO-2010), 23–27 Aug 2010, Aalborg, Denmark
3. Hassanien A, Vorobyov SA (2009) Transmit/receive beamforming for MIMO radar with colocated antennas. In: Proceedings of IEEE international conference on acoustic, speech, and

signal processing, Taipei, Taiwan, pp 2089–2092

4. Rui F, de Lamare RC (2010) An adaptive LCMV beamforming algorithm based on dynamic selection of constraints. Communications Research Group, Department of Electronics, University of York, YO10 5DD, United Kingdom

5. Ferguson BG (1998) Minimum variance distortionless response beamforming of acoustic array data. J Acoustic Soc Am 104:947. https://doi.org/10.1121/1.423311

6. Zou Q, Yu ZL, Lin Z (2004) A robust algorithm for linearly constrained adaptive beamforming. IEEE Signal Process Lett 11(1)

7. Kiong TS, Salem SB, Paw JK, Sankar KP, Darzi S (2014) Minimum variance distortionless response (MVDR) beamformer with enhanced nulling level control via dynamic mutated artificial immune system. Scientif World J 164053I(10)

8. http://www.labbookpages.co.uk/audio/beamforming/frost.html

9. Van Trees HL (2002) Optimum array processing. Wiley, NY

10. Hassanien A, Vorobyov SA, Wong KM (2008) Robust adaptive beamforming using sequential programming: an iterative solution to the mismatch problem. IEEE Signal Process Lett 15:733–736

LabVIEW-Based Software and Hardware Implementation of 16-QAM Modem

Mohini, Anshu Priya, Samarendra Nath Sur and Soumyasree Bera

Abstract In the world of wireless communication, there is a potential expansion in the field of digital communication and many new technologies and applications are growing every day. With the advancement in wireless communication and it becomes an essential part for modern society to share knowledge and information. In this article, 16-QAM modem system is implemented in LabVIEW software and hardware implementation is initialized. LabVIEW is a system-design platform and used in graphical programming and allows a designing system in a block-based manner.

Keywords 16-QAM modem · LabVIEW · Modulator · Demodulator · Function generator · DSO

1 Introduction

This paper explains the 16-QAM Modem using LABVIEW software and its hardware implementation. Basically, a modem is a combination of modulator and demodulator. It helps to convert carrier signals into digital data for transmission and demodulates signals to decode the transmitted data. The main function of the modem is to transmit and decrypt signal that allows digital data to be sent from one place to other places and send the information over the channel by avoiding the loss of information. Modems are widely used in computers and for internet services.

In this paper, 16-QAM is chosen to be the modulation and demodulation scheme using LABVIEW and it is extensively used in the field of telecommunication, radio communications, optical fiber system and also used for delivering the data [1–3].

Mohini · A. Priya · S. N. Sur (✉) · S. Bera
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India
e-mail: samar.sur@gmail.com

2   Quadrature Amplitude Modulation

QAM means the quadrature amplitude modulation is both analog and digital modulation. It is used for modulating data onto the carrier for radio communications. It is the combination of two amplitude signals in a channel and the effective bandwidth is doubled. This is used with the pulse amplitude modulation in the digital and wireless communication. QAM signal comprises two carriers each having the same frequency and differed by 90° phase. These two carriers are divided into In-phase and Q-phase. Mathematically one signal is a sine wave and other is a cosine wave. QAM is more advantageous than other modulation because it carries more number of bits of information per symbol and more efficient than other modulation. The bit rate is increased without increasing the bandwidth [4–6] (Fig. 1).

2.1   16-QAM

The 16-QAM means 16 states quadrature amplitude modulation, it has 4 I Q values and its bandwidth efficiency is four bits/second/hertz. The data is divided into two channels I and Q, it shows two amplitude values. Simultaneously two bits are added to each channel (Fig. 2).
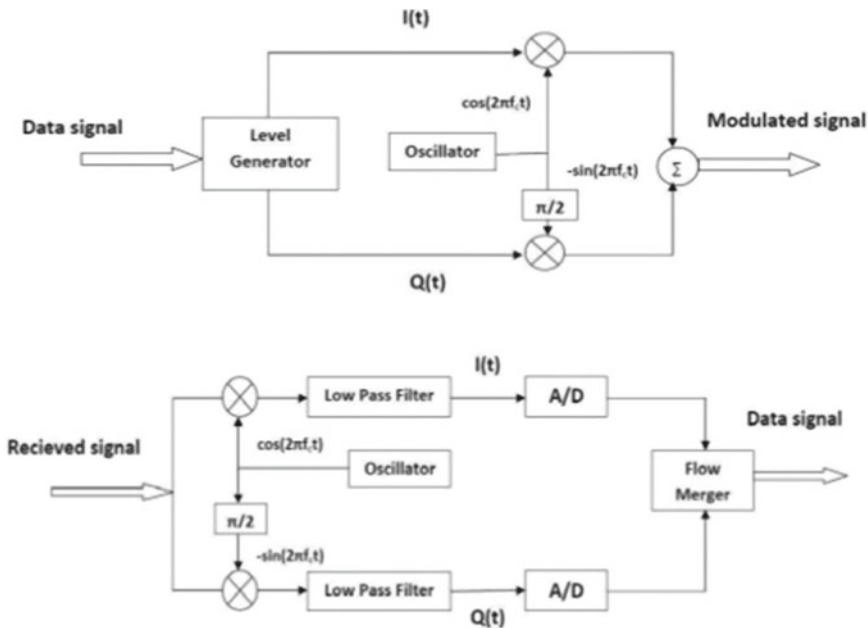
Fig. 1   Block diagram of 16-QAM modulator–demodulator
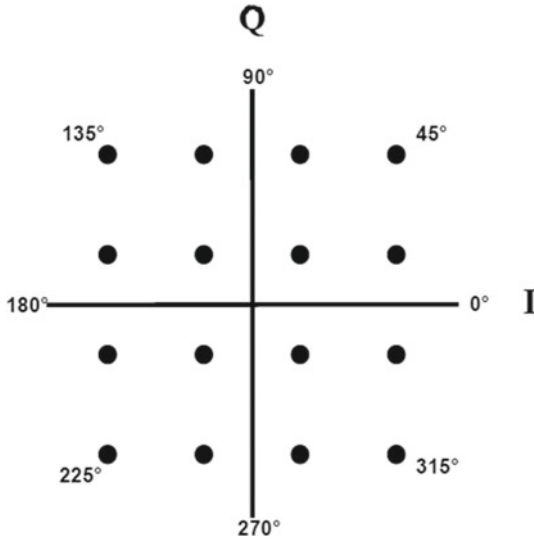
Fig. 2 I and Q plot of 16-QAM

3 LabVIEW Implementation of 16-QAM Modem

This section presents the LabVIEW implementation of 16-QAM modem. LabVIEW is the system designed platform software, which is used for visual programming language from National Instruments. It deals with data and graphical programming and in a block-based manner. As LabVIEW is termed visual instruments (vis). It consists of three important panels block diagram panel, front panel, and connector panel. The block diagram panel is used for building blocks for the given model, all the connections are done in this panel and it gives the graphical source code. The front panel shows the result of the model with the help of graph, input–output sequence, and various indicators. A brief description of each block is given below [7–12] (Fig. 3).

3.1 16-QAM Modulator

In this section, we have implemented the 16-QAM modulator in LabVIEW by using several blocks. In this model, first, we have used array size block, which defines the size of an array then an input bit sequence is connected with this block which helps in adding bits in the block. This array size block is linked with the quotient and remainder block it gives the quotient and remainder of the input and a constant block is connected which divides the bits. Then the reshape array block is used to change the dimension of an array. The main function of this block is to read the data row
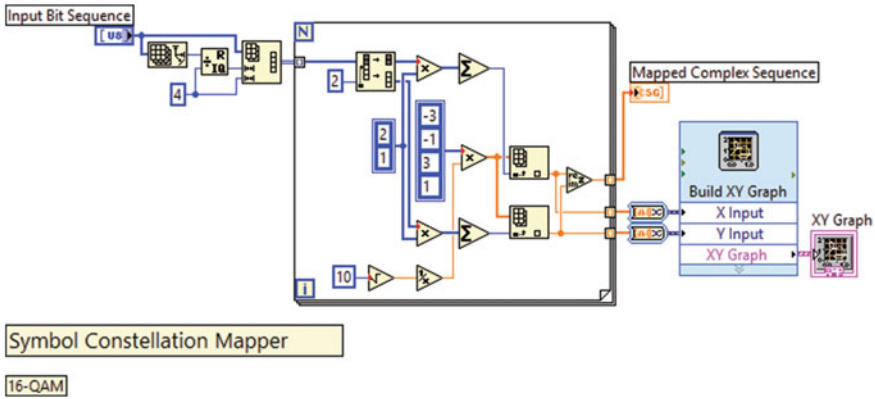
Fig. 3 16-QAM modulator

by row as input bits are given and bits are obtained in a row pattern. The split 1-D array block is linked, which divides array at index and return the two portions with the elements of the index of the second sub-array. The constant block is connected, which divides the bits into two so that it can give values in real and imaginary. Then index array block is used which return the n-dimensional array of the index. The bits get divided in this block and it is connected with the multiplier and summation of the bits are done it resizes the input to show the values of n-dimensional array. Then real–imaginary to complex function block is used which convert into complex value. It is connected with the mapped sequence block which shows the complex values in the front panel. This block is linked with the graph which shows the constellation diagram of 16-QAM modulation in the front panel.

3.2 16-QAM Demodulator

This section shows the demodulation part. In this section, the obtained modulated values are subtracted with all the possible points using case structure in LabVIEW. Then the normalized value is obtained and stored in the form of an array. With the help of an array, we obtained the minimum value by using the Array Max and Min function block. The main function of this block is to give the maximum and minimum values of the array and obtained the index value. The index value we obtained is converted in binary form by using the digital to binary converter as shown in the figure. Hence, the demodulated values are obtained (Fig. 4).
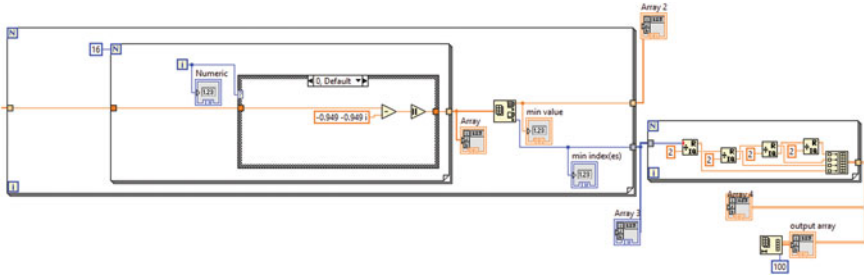
Fig. 4 16-QAM demodulator
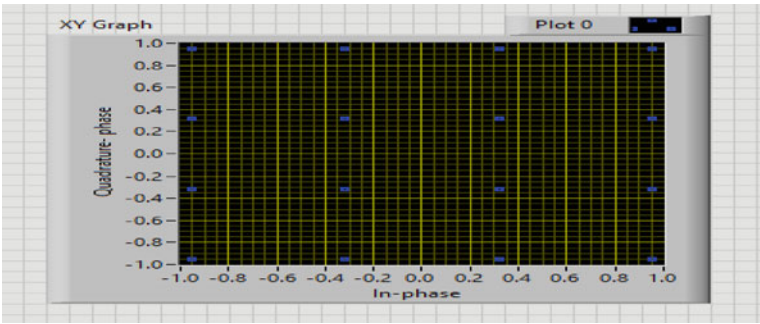
Fig. 5 Result of 16-QAM modulator
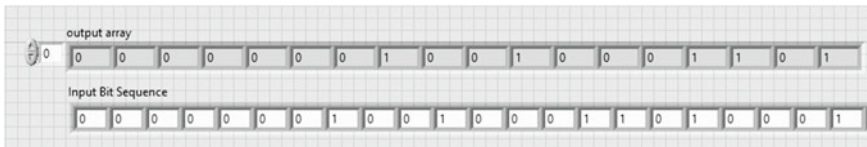
Fig. 6 Result of 16-QAM demodulator

4 Simulation Result

This section shows the simulation results of 16-QAM Modem. In this, XY graph is used to denote the constellation plot in the front panel block (Figs. 5 and 6).

5 Hardware Implementation

This section explains the hardware implementation of 16-QAM. It describes the different setups being used for hardware realization. With the help of DSO (Digital signal oscilloscope) and function generator, we obtained the output. In the hardware
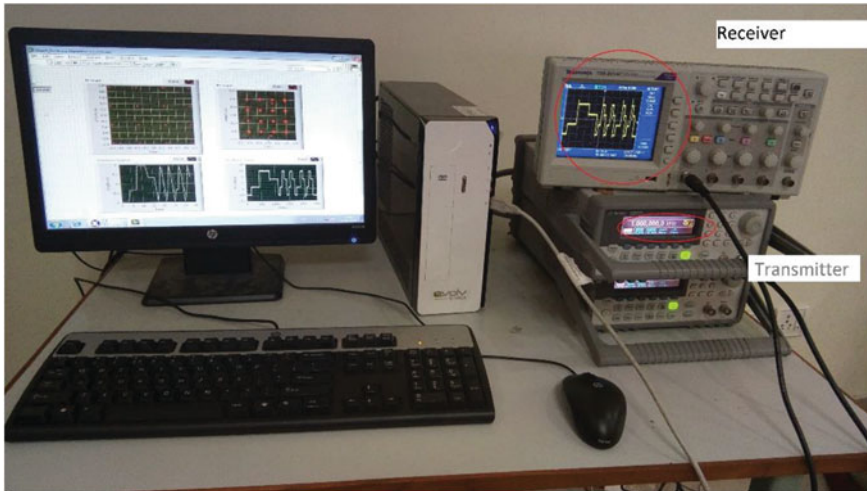
Fig. 7 Hardware implementation of 16-QAM in closed-loop mode
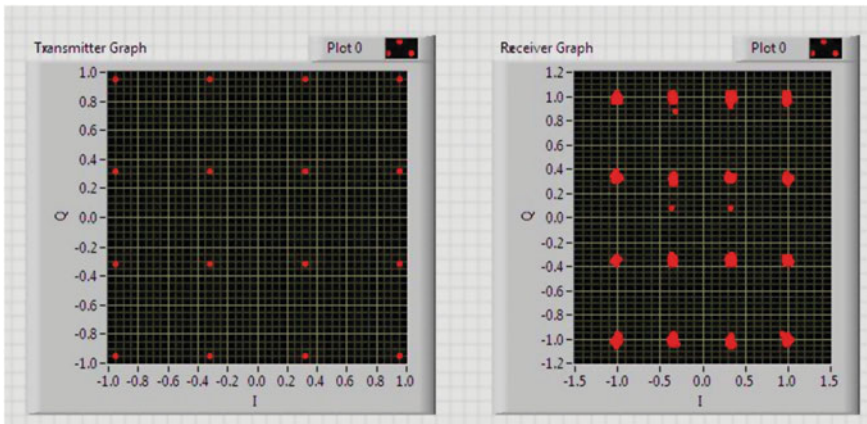
Fig. 8 Output of 16-QAM modulator in closed-loop mode

part, various blocks are used in LabVIEW. In the hardware section, an initializer of function generator is used which is connected with the 16-QAM modulator and basically this initializer establishes communication with the system then a create block is used so that waves can be generated. The function generator part is linked with the DSO section which consists of an initializer, auto setup, and the channel source. It is connected with the receiver part to obtain the output in the DSO (Digital signal oscilloscope) as well as in the front panel (Figs. 7 and 8).

6 Conclusion

In this paper, it is shown that how LabVIEW is used to design the 16-QAM modem. With the help of various blocks like array size block, quotient and remainder block, reshape array, split 1D array, index array, and real to complex value block are used to obtain the modulated values and by using the case structure and converting the values with the help of digital to binary converter, we obtained the demodulated values. Basically, the LabVIEW shows the block-based functions and helped in obtaining the graphical plot of 16-QAM modem. Further, the hardware implementation is done with the help of the function generator and DSO we obtained the output.

References

1. Oyetunji SA, Akinninranye AA (2013) Performance evaluation of digital modulation techniques in AWGN communication channel. Int J Eng Res Technol 2(5): 2100–2106 ISSN 2278-0181
2. Tilwari V, Kushwah AS (2013) Performance analysis of Wi-Max 802.16e physical layer using digital modulation techniques and code rates. Int J Eng Res Advanc Technol 3(4):1449–1454 ISSN 2248-9622
3. Sharma DK, Mishra A, Saxena R (2010) Analog and digital modulation techniques: an overview. TECHNIA Int J Comput Sci Commun Technol 3(1):551–561 ISSN 0974-3375
4. Barnela M (2014) Digital modulation schemes employed in wireless communication. Int J Wired Wireless Commun 2(2):2042–2059
5. Devi D, Sharma A (2013) BER performance of GMSK using MATLAB. Int J Advanc Res Comput Eng Technol 2(4):1389–1392
6. Bryan D (1995) QAM for terrestrial and cable transmission. IEEE Trans Consumer Electron 41(3):383–391
7. National Instruments, LabVIEW User Manual, Part Number 320999E-01, 20 03
8. Gurugubelli J, Indrajit C (2009) Design and implementation of generalised parametrizable modulator for a reconfigurable radio. In: IEEE region 10 conference on TENCON 2009, Sept. 2009, pp 1–6
9. Swami A, Sadler BM (2000) Hierarchical digital modulation classification using cumulants. IEEE Trans Commun 48:416–429
10. Soliman SS, Hsue S-Z (1992) Signal classification using statistical moments. IEEE Trans Commun 40:908–916
11. Yang Y, Liu C-H, Soong T-W (1998) A log-likelihood function-based algorithm for QAM signal classification. Signal Process 70(1):61–71
12. Pfau T, Hoffmann S, Noe R (2009) Hardware efficient coherent digital receiver concept with feed forward carrier recovery for M-QAM constellation. IEEE/OSA J Lightwave Technol 27(8):989–999

A Study on Capacity of MRC over OWDP Fading Channels

Laishram Mona Devi, Nibedita Das, Suparna Goswami and Aheibam Dinamani Singh

Abstract The system capacity degrades because of fading in the wireless channels. The channel capacity expressions for Maximal Ratio Combining (MRC) diversity receiver over One-Wave Diffused Power (OWDP) fading channels are presented in this paper considering an arbitrary number of diversity branches. OWDP fading channel models have multiple diffused components and one specular component. To calculate the channel capacities, Probability Density Function (pdf) is employed for numerous adaptive transmission methods. Also, the comparison is done among the capacities of numerous adaptive transmission methods.

Keywords Adaptive transmission · Channel capacity · MRC · OWDP fading · Pdf

1 Introduction

In wireless communication, to measure spectral efficiency and assure better Quality of Services, the study of capacity is crucial. The existence of fading in the wireless channels degrades the system capacity. This degradation can be removed by employing one of the diversity techniques called Maximal Ratio Combining (MRC). In MRC, the multipath signals are co-phased, weighted, and superimposed at the receiver side in order to enhance the received signal-to-noise ratio (SNR). With coherent detection, MRC gives optimum performance among various diversity methods,

L. M. Devi (✉) · N. Das · S. Goswami · A. D. Singh
Department of ECE, NERIST, Nirjuli, Arunachal Pradesh, India
e-mail: monalaishram16@gmail.com

N. Das
e-mail: nibeditadas577@gmail.com

S. Goswami
e-mail: suparnanerist@gmail.com

A. D. Singh
e-mail: ads@nerist.ac.in

even though it has high complexity due to its requirement to weight every branch [1].

In One-Wave Diffused Power (OWDP) fading model, many diffused propagating wave and a specular component are presented in received signals. This fading channel is encountered in networks (wireless sensor), which are mounted in a cavity equipped with directional antennas. They are mainly used in tunnel, public transportation vehicles, airframe, etc. This fading is the particular condition of Two-Wave with Diffused Power (TWDP) Fading Channels [2]. The upper and lower limit expression for Bit Error Probability (average) of binary phase shift keying system for OWDP fading was developed in [3]. In [4], the work on dual MRC receiver capacity using Probability Density Function (pdf)-based process over fading channels which is nonidentical TWDP was presented. In presence of arbitrary and nonidentical fading parameters performance analysis of Selection Combining (SC) receiver over TWDP fading channel was analyzed [5]. By using the output SNR pdf from [5], the system capacity is calculated and studied over TWDP fading channel for SC receiver [6]. MRC receiver performance using pdf-based method over TWDP fading channels was presented [7]. In [8], Cumulative Distribution Function method was employed to develop the expression of MRC diversity system performance in TWDP fading channels. A new form of fading model over TWDP has developed in [9] using a mixture of Gamma distribution and channel capacity of TWDP channel was observed under different transmission methods. The capacity over TWDP fading channels for MRC receiver with four techniques of adaptive transmission was investigated in [10]. Till now no paper has been accounted in the area of analyzing MRC receiver capacity over OWDP fading channel.

In this paper, the transmitter employed few adaptive transmission methods whereas MRC diversity technique is applied at the receiver. The expressions of capacity for numerous adaptive transmission methods are derived over OWDP fading channels. Finally, the expressions are analyzed and compared among the numerous adaptive transmission methods.

This paper is arranged in the following way. In Sect. 2, the channel and system are discussed. Section 3 gives the combined output SNR pdf of MRC receiver. Various capacity formulas and derivations for adaptive transmission methods are demonstrated in Sect. 4. The results and discussion part are presented in Sect. 5. In Sect. 6, the conclusion of this topic is presented.

2 Channel and System

The fading channels which have multiple diffused components and a specular component are considered as OWDP. The channel is considered as slow and frequency nonselective along with OWDP fading statistics. The received signal expression with complex low-pass function with single symbol duration T_s is given below

$$r'(t) = re^{j\psi} s(t) + n(t), \tag{1}$$

where transmitted data with energy E_b is represented as s(t), zero mean complex Gaussian noise and power spectral density for two sided (2No) is represented as n(t). The phase is usually denoted by random variable ψ and the amplitude of OWDP fading distribution is noted as r. Then, the envelop pdf of OWDP fading channels is

$$P_R(r) = \sum_{i=1}^{M} a_i \frac{r}{\sigma^2} e^{-\left\{\frac{r^2 + 2\sigma^2 K}{2\sigma^2}\right\}} I_0\left(\frac{r}{\sigma^2}\sqrt{2\sigma^2 K}\right), \tag{2}$$

where $I_0(\cdot)$ is the first kind Bessel function with zeroth order, K is the ratio of overall specular power and diffused power, i represents approximation order, M is series order, a_i is the coefficients and σ_i is the variance. Applying the square and random variable transformation, the SNR pdf expression of OWDP fading channels is

$$f_\gamma(\gamma) = \eta \sum_{i=1}^{M} a_i e^{-(\eta\gamma + K)} I_0\left(2\sqrt{K\eta\gamma}\right) \tag{3}$$

where $\eta = \frac{K+1}{\bar{\gamma}}$.

3   The Pdf of Output SNR for MRC Receiver

The pdf of the output SNR for MRC receiver over OWDP (γ_{MRC}) is obtained after solving the Characteristic Function (CF) expression of (2) and the final expression is

$$\begin{aligned}
f_{\gamma_{MRC}}(\gamma) = (\eta)^L \sum_{i_1=1}^{M} \cdots \sum_{i_L}^{M} \left\{\prod_{t=1}^{L} a_{i_t}\right\} \\
\times \sum_{s=0}^{\infty} (\eta L K)^s \frac{e^{-LK} e^{-\eta\gamma}}{\Gamma(L) s! (L)_s} \gamma^{L+s-1}
\end{aligned} \tag{4}$$

where $\Gamma(\cdot)$ is the complete Gamma function, $(\cdot)_s$ is the Pochhammer symbol, L is the number of diversity branches.

4 Adaptive Transmission System Capacity

The capacity of MRC receiver over OWDP fading channels can be numerically derived using the following methods [12, 13].

4.1 *Channel Inversion with Fixed Rate (CIFR)*

The process of CIFR scheme is used to transmit power to ensure fixed received SNR. The channel capacity $C_{CIFR}(\bar{\gamma})$ in (bit/second/hertz) is [12] expressed as

$$C_{CIFR}(\bar{\gamma}) = B \log_2 \left(1 + \frac{1}{R_{CIFR}(\bar{\gamma})} \right), \tag{5}$$

where

$$R_{CIFR}(\bar{\gamma}) \cong \int\limits_0^\infty \left(\frac{f_\gamma(\gamma)}{\gamma} \right) d\gamma. \tag{6}$$

Putting (4) in (6), the $R_{CIFR}(\bar{\gamma})$ integral turn out as

$$R_{CIFR}(\bar{\gamma}) = (\eta)^L \sum_{i=1}^M \ldots \sum_{i_L=1}^M \left\{ \prod_{t=1}^L a_{i_t} \right\} \sum_{s=0}^\infty (\eta L K)^s$$

$$\times \frac{e^{-LK}}{\Gamma(L)s!\,(L)_s} \int\limits_0^\infty \left(\frac{1}{\gamma} \right) e^{-\eta\gamma} \gamma^{L+s-1} d\gamma. \tag{7}$$

The integral is solved using [11, (3.381.4)]. After algebraic manipulation the rate equation is

$$R_{CIFR}(\bar{\gamma}) = \sum_{i=1}^M \ldots \sum_{i_L=1}^M \sum_{s=0}^\infty \frac{\left\{ \prod_{t=1}^L a_{i_t} \right\} (\eta)^{1-s} e^{-LK} (\eta L K)^s}{\Gamma(L)s!\,(L)_s}.$$

$$\times \Gamma(L+s-1) \tag{8}$$

By putting Eqs. (8) into (5) the final expression of CIFR capacity is achieved.

4.2 *Truncated Channel Inversion with Fixed Rate (TIFR)*

In CIFR scheme, a large transmission power is mandatory during deep fades of channel to achieve SNR which is constant. But for the transmitter, it becomes a big challenge to transmit this massive power of signal. Therefore, during deep fade, all the transmissions are noise dominated signal thus disturbance occur in decoding process. Therefore, TIFR scheme is employed to surmount the effect of CIFR scheme. In TIFR transmitter immediately stops transmission when channel goes to deep fade, and channel inversion occurs if the received SNR is higher than γ_0 [12]. Then, the formula of capacity is

$$C_{TIFR}(\bar{\gamma}) = B \log_2\left(1 + \frac{1}{R_{TIFR}(\bar{\gamma})}\right)(1 - P_{out}(\gamma_0)), \qquad (9)$$

where

$$R_{TIFR}(\bar{\gamma}) \cong \int_{\gamma_0}^{\infty} \left(\frac{f_\gamma(\gamma)}{\gamma}\right) d\gamma \qquad (10)$$

and

$$P_{out}(\gamma_0) = \int_0^{\gamma_0} f_\gamma(\gamma) d\gamma \qquad (11)$$

is the outage probability having threshold value γ_0.

For solving the TIFR capacity technique, it is needed to calculate $R_{TIFR}(\bar{\gamma})$ and $P_{out}(\gamma_0)$. The expression of $R_{TIFR}(\bar{\gamma})$ is obtained by putting (4) into (9) as

$$R_{TIFR}(\bar{\gamma}) = \sum_{i_1=1}^{M} \cdots \sum_{i_L=1}^{M} \left\{ \prod_{t=1}^{L} a_{i_t} \right\} (\eta^L)$$

$$\times \sum_{s=0}^{\infty} (\eta LK)^s \frac{e^{-LK}}{\Gamma(L)s!\,(L)_s} \int_{\gamma_0}^{\infty} \gamma^{L+S-1-1} e^{-\eta\gamma} \cdot d\gamma. \qquad (12)$$

Using [11, (3.381.3)] the integral in (12) is solved. Then, the final expression after simplification is

$$R_{TIFR}(\bar{\gamma}) = \sum_{i_1=1}^{M} \cdots \sum_{i_L=1}^{M} \times \sum_{s=0}^{\infty} \frac{\left\{ \prod_{t=1}^{L} a_{i_t} \right\} \eta (LK)^s e^{-LK} \Gamma(L+s-1, \eta\gamma_0)}{\Gamma(L)s!\,(L)_s} \qquad (13)$$

where $\Gamma(.,.)$ is upper incomplete Gamma function. The expression for $P_{out}(\gamma_0)$ is deduced from (4) as

$$P_{out}(\gamma_0) = \sum_{i_1=1}^{M} \ldots \sum_{i_L=1}^{M} \sum_{s=0}^{\infty} \left\{ \prod_{t=1}^{L} a_{i_t} \right\} e^{-LK} (LK)^s \frac{g(L+s, \eta\gamma_{th})}{\Gamma(L)s!\,(L)_s}, \qquad (14)$$

where $g(.,.)$ is lower incomplete Gamma function. Putting both Eqs. (13) and (14) into (8) and the final expression for this scheme is obtained.

4.3 Constant Transmitting Power

If the transmission rate changes corresponding to channel condition even though the transmission power is constant, then the transmitter adopts Optimal Rate Adaptation (ORA) [12]. The channel capacity formula (bit/second/hertz) for ORA scheme is

$$C_{ORA}(\bar{\gamma}) = B\log_2 e \int_0^{\infty} \ln(1+\gamma) f_{\gamma}(\gamma) d\gamma. \qquad (15)$$

The ORA capacity expression is derived by putting (4) into (15). After rearranging the expression, the expression is found as

$$C_{ORA}(\bar{\gamma}) = B\log_2 e(\eta)^L \sum_{i_1=1}^{M} \ldots \sum_{i_L=1}^{M} \left\{ \prod_{t=1}^{L} a_{i_t} \right\}$$
$$\times \sum_{s=0}^{\infty} (\eta LK)^s \frac{e^{-LK}}{\Gamma(L)s!\,(L)_s} \int_0^{\infty} \ln(1+\gamma)\gamma^{L+S-1} e^{-\eta\gamma} d\gamma. \qquad (16)$$

Then, the final expression of ORA capacity is deduced by simplifying (16) using [14, (78)] as

$$C_{ORA}(\bar{\gamma}) = B\log_2 e \sum_{i_1=1}^{M} \ldots \sum_{i_L=1}^{M} \sum_{s=0}^{\infty} \sum_{k=1}^{L+s} \left\{ \prod_{t=1}^{L} a_{i_t} \right\} \eta^{s+L-k} e^{-LK} e^{\eta} (LK)^s$$
$$\times \frac{(L+s-1)!\,\Gamma[-L-s+k,\eta]}{\Gamma(L)s!\,(L)_s}. \qquad (17)$$

4.4 Optimal Power and Rate Adaption at the Transmitter

If the transmitter varies the transmitted power and rate based on the channel condition, then the transmitter is said to adopt the Optimal Power and Rate Adaptation (OPRA) scheme [12]. Mathematically, OPRA channel capacity (bits/second/hertz) is

$$C_{OPRA}(\bar{\gamma}) = B \int_{\gamma_0}^{\infty} \log_2\left(\frac{\gamma}{\gamma_0}\right) f_{\gamma}(\gamma) d\gamma, \tag{18}$$

where B Hz represents channel bandwidth, pdf of the output SNR is represented as $f_{\gamma}(\gamma)$ and γ_0 is the most effective cut-off of SNR; below this level, no transmission is occurred. The capacity expression for OPRA is deduced by putting (4) into (18) as

$$C_{OPRA}(\bar{\gamma}) = B \log_2 e(\eta)^L \sum_{i=1}^{M} \cdots \sum_{i_L=1}^{M} \sum_{s=0}^{\infty} \frac{\left\{\prod_{t=1}^{L} a_{i_t}\right\}(\eta L K)^s e^{-LK}}{\Gamma(L) s! \, (L)_s}$$
$$\times \int_{\gamma_0}^{\infty} \ln\left(\frac{\gamma}{\gamma_0}\right) \cdot \gamma^{L+s-1} \cdot e^{-\eta\gamma} d\gamma. \tag{19}$$

After substituting $x = \frac{\gamma}{\gamma_0}$, the expression obtained is given as

$$C_{OPRA}(\bar{\gamma}) = B \log_2 e(\eta)^L \sum_{i=1}^{M} \cdots \sum_{i_L=1}^{M} \sum_{s=0}^{\infty} \frac{\left\{\prod_{t=1}^{L} a_{i_t}\right\}(\eta L K)^s e^{-LK}}{\Gamma(L) s! \, (L)_s}$$
$$\times \gamma_0^{L+s} \int_{\gamma_0}^{\infty} \ln(x) \cdot x^{L+s-1} \cdot e^{-\eta\gamma_0 x} dx. \tag{20}$$

The integral is solved by using [14, (64)]. After simplification of (20) the final OPRA expression is as

$$C_{OPRA}(\bar{\gamma}) = B \log_2 e \sum_{i=1}^{M} \cdots \sum_{i_L=1}^{M} \sum_{s=0}^{\infty} \frac{\left\{\prod_{t=1}^{L} a_{i_t}\right\}(L K)^s e^{-LK}}{\Gamma(L) s! \, (L)_s}$$
$$\times (L+s-1)! \sum_{k=0}^{L+s-1} \frac{\Gamma(k, \eta\gamma_0)}{k!}. \tag{21}$$

The most effective cut-off of SNR needs to satisfy the condition mentioned below

$$\int\limits_{\gamma_0}^{\infty} \left(\frac{1}{\gamma_0} - \frac{1}{\gamma} \right) f_\gamma(\gamma) d\gamma = 1. \tag{22}$$

The equation of output SNR pdf is involved to calculate the capacities, i.e., from (5)–(20). The expression for the pdf of the output SNR of MRC receiver is given in (4).

5   Numerical Results and Discussion

Capacity of various techniques for adaptive transmission which had been mathematically derived in the above section is studied. Figure 1 exhibits capacity versus average SNR ($\bar{\gamma}$) graph using the CIFR method. The figure shows that channel capacity improves with an increase in the number of branches L at a fixed value of K. This is because with the increasing number of diversity branches the probability of entire fading channels going to a period of deep fade is reduced. From the curves given in Fig. 2, it can be observed that for a fixed value of L, while the value of K increases the system capacity is improving. The main reason for this nature is because the higher value of K indicates greater power of direct waves resulting to a better channels. In Fig. 3, the comparison of numerous adaptive transmission schemes for MRC receiver is shown. It is observed from the plot that a certain capacity is obtained with lowest SNR with ORA. When the SNR is low, the capacity obtained from TIFR capacity performance for larger SNR. In this situation, the transmission of TIFR is suspended. The threshold SNR γ_0 is fixed at 2 dB.

Fig. 1 MRC receiver capacity with CIFR method
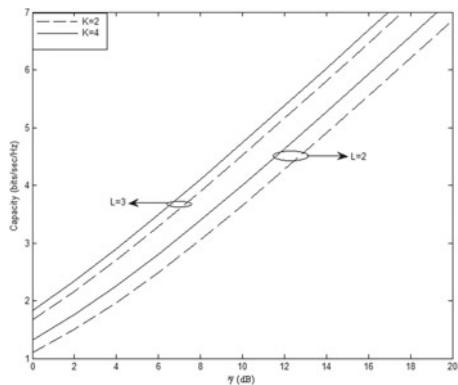
Fig. 2 MRC receiver capacity with TIFR method
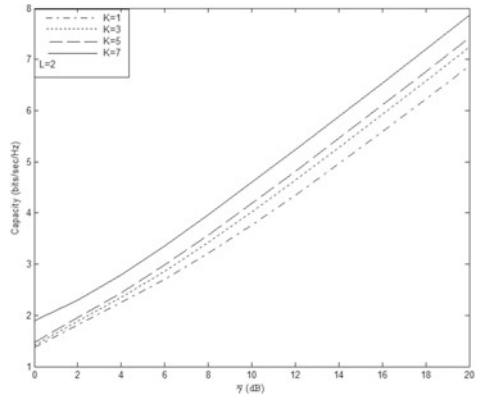
Fig. 3 MRC receiver capacity using adaptive transmission methods
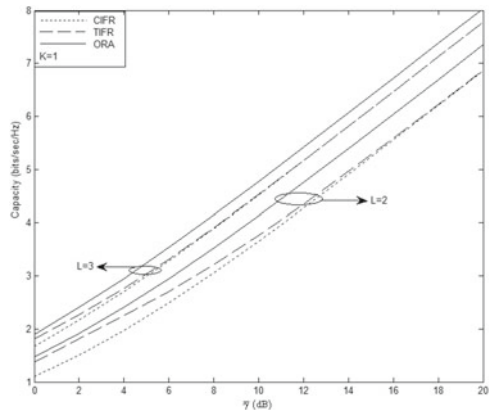
6 Conclusions

For numerous adaptive transmission methods, channel capacity expressions of MRC diversity receiver over OWDP fading channels are investigated. These expressions are obtained using the output SNR pdf of diversity system. The expression consists of gamma function as well as Pochhammer functions. The numerically evaluated results are plotted and the results are compared and analyzed.

References

1. Simon MK, Alouini M-S (2005) Digital communications over fading channels, 2nd edn. Wiley, New York
2. Durgin GD, Rappaport TS, de Wolf DA (2002) New analytical models and probability density functions for fading in wireless communication. IEEE Trans Commun 50(6):1005–1015

3. Goswami S, Singh SH, Singh AD (2010) ABEP upper and lower bound of BPSK system over OWDP fading channels Advanc Wireless Mobile Commun 10:307–313. ISSN 0973-6972.10 © Research India Publications

4. Subadar R, Singh AD (2015) Capacity of dual MRC receiver with adaptive transmitters over non identical TWDP fading channels. International conference on advances in computer engineering and applications (ICACEA), IMS Engineering College, Ghaziabad, India

5. Subadar R, Singh AD (2013) Performance analysis of M-SC receiver over TWDP fading channels. IEEE WCL 2(3):267–270

6. Subadar R, Singh AD (2014) Capacity analysis of SC receiver over TWDP fading channel. Int J Electron Commun (AEU) 68(2):166–17I

7. Subadar R, Singh AD (2014) Performance of MRC system over TWDP fading channels. Int J Electron Commun (AEU) 68(6):569–572

8. Yao L, Yang N (2011) Symbol error probability of QAM with MRC diversity in two wave with diffuse power fading channels. IEEE Commun Lett 15(1):10–12

9. Ermolova NY (2016) Capacity analysis of two-wave with diffuse power fading channels using a mixture of gamma distributions. IEEE Commun Lett 20(11)

10. Dinamani A, Sub R (2015) Capacity analysis of MRC receiver with adaptive transmitters over TWDP fading channels. ISACC (2015)

11. Gradshteyn S, Ryzhik 1M (2000) Table of integrals, series, and products, 6th edn. Academic, San Diego, CA

12. Goldsmith AJ, Variaya PP (1997) Capacity of fading channels with channel side information. IEEE Trans Theor 43(6):1986–1982

13. Cheng J, Berger T (2003) Capacity of Nakagami-q (Hoyt) fading channels with channel side information. Proc ICCT 1915–1918

14. Alouini M-S, Goldsmith AJ (1999) Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques. IEEE Trans Vehicle Telecommun 48:1165–1181

A Novel Wavelet-Based Image Defogging Using Dark Channel Prior and Guided Filter

Manas Sarkar, Ujjwal Mondal and Debashis Nandi

Abstract Recently, fog removal from images and videos has got tremendous importance in image and video processing for object detection, tracking and surveillance system, advance image editing, and many more which are poorly affected by the fog. A lot of works have been proposed so far for removing fog from image and video. It includes different methods like contrast enhancement, local color line model based and many more fog removal technique. In this paper, we propose a novel defogging method where guided filter based dark channel prior (DCP) is applied on low–low (LL) band of discrete wavelet transformation (DWT) coefficient of the intensity rectified image which is obtained after homomorphic filtering. To improve the sharpness of the output, unsharp masking (USM) is applied on high–low (HL) and low–high (LH) bands of the DWT coefficient. Finally, sharp and high contrast output is generated by applying contrast limited adaptive histogram equalization technique (CLAHE) to the inverse transformed image. The proposed method improves the overall quality of the defogged output image with respect to contemporary methods. The measured quality metrics of the proposed method are compared with some recent works. The quantitative and qualitative results confirm the claims.

Keywords Dark channel prior · Depth estimation · Homomorphic filtering · Wavelet transform · Contrast enhancement · Unsharp masking

M. Sarkar (✉)
Haldia Institute of Technology, Haldia, India
e-mail: manasm.sarkar@gmail.com

U. Mondal
Calcutta University, Kolkata, India
e-mail: ujjwalmondal18@gmail.com

D. Nandi
National Institute of Technology, Durgapur, Durgapur 713209, West Bengal, India
e-mail: debashisn2@gmail.com

1   Introduction

In computer vision applications, many systems like outdoor object detection, tracking and surveillance, navigation and many more are poorly affected by fog because the foggy atmosphere makes the visibility poor in terms of contrast, intensity, sharpness, and color quality. It causes the information loss of the objects from the scene and the overall image quality becomes degraded. Therefore, we should apply a proper fog removal technique to discover the depth information [1] from a foggy image or video sequence.

   Image fog is a function of scattering coefficient and unknown depth of the scene object. Defogging and dehazing is a process of eliminating fog and haze from images to rebuild the frames of the original scene in terms of intensity, color, contrast, and resolution. Conventionally, two types of defogging techniques based on the number of frames are available from different applications: (i) Single frame based, [2, 1, 3–5, 6] and (ii) Multi-frame based [7, 8]. In single frame based defogging technique, only one frame of a scene is used whereas in multi-frame based defogging technique, multiple frames of the same scene are used for removing fog. Single frame based defogging techniques have become more popular because multi-frame based techniques need prior information of the scene from the reference images. Though the single frame based techniques are recent and more popular, these also have some limitations like fails to give all scene information in dense fog and sky region, does not physically improve the depth but just enhances the visibility [3]. In some of the cases, color quality, contrast, sharpness, and intensity are not properly maintained together. Hence, after identifying the shortcomings of previous fog removal techniques, we are motivated to propose a new technique which gives superior results.

2   Review of Previous Works

Tarel and Hautiere [2] proposed a fog removal algorithm which can be used in in-vehicle camera. The technique works based on considering the low saturation colors of tiny objects and few parameters dominate the technique like atmospheric veil inference, image smoothing, and tone mapping. The restored defogged image is $R(x, y) = \frac{I(x,y) - V(x,y)}{1 - \frac{V(x,y)}{I_s}}$, where $V(x, y)$ is the atmospheric veil and I_s is the sky intensity. This method fails to restore the actual color of the image and cannot handle dense fog.

   He et al. [1] introduced a novel image defogging method, depending on the features of outdoor defogged image. Based on the thickness and transmission map, the fog is estimated and the equation is as follows: $\tilde{t}(x) = \omega \min_{y \epsilon \Omega(x)}(\min_c \frac{I^c(y)}{A^c})$, where $\tilde{t}(x)$ is transmission and $\min_{y \epsilon \Omega(x)}(\min_c \frac{I^c(y)}{A^c})$ is the dark channel of the foggy image. Sometimes, this method fails to estimate dense fog and does not work well for infinite depth and sky region.

Tripathy and Mukhopadhyay [4] proposed a technique which uses anisotropic diffusion for refining air light map from DCP. This technique is applied for recovering the scene contrast and to smooth the air light map. Here, the de-foggy image $I_0(x, y)$ can be restored as: $I_0(x, y, c) = \frac{I(x,y,c)-A(x,y)}{1-(A(s,y)/I_\infty(c))}$. This method works very well in hazy scene but image clarity is not up to the mark in some cases.

Fattal [5] developed a color image defogging technique which works based on a local color line model which shows a one-dimensional distribution of intensity values of small image patches. These approaches are used to obtain the transmission map. But in the sky region, the method is not capable to produce good results.

Sijiu et al. [6] propose a wavelet and dark channel based fog removal method which works on different frequency ranges but it has its own limitations of estimating ambient light and sometimes dense fog depth.

3   Proposed Algorithm

As the limitations of previously proposed algorithms are observed, we are trying to overcome it to get optimum output. Mainly, we are focused to solve the problems generated by dense fog or to extract more information behind the dense fog. Along with that, other problems are also taken care of. Our proposed algorithm can be separated into two units such as preprocessing and postprocessing unit whereas the main processing unit works based on the principle of novel DCP method. Preprocessing unit consists of homomorphic filtering and DWT whereas postprocessing unit consists of guided filtering, USM, and local contrast enhancement technique. A flow diagram of the proposed defogging technique is depicted in Fig. 1.

Image fog model [1, 4] is represented as:

$$O(x) = J(x) * t(x) + A(1 - t(x)) \tag{1}$$

where $O(x)$ is an observed foggy image intensity, $J(x)$ is the defogged image intensity or scene radiance, $t(x)$ is medium transmission, and A is global atmospheric light. Fog is generated [1, 4] in the atmosphere mainly due to two factors: (a) Direct attenuation and (b) Air light. In foggy image, scene radiance is reduced exponentially with the scene depth and it is related to transmission. To obtain defogged image or to recover $J(x)$, the value of A and $t(x)$ must be calculated. In our proposed method, (1) Homomorphic filter [9] is used in the foggy image to improve the intensity by correcting the nonuniform illumination. Low-frequency component has to be decreased and high frequency has to be increased for maintaining the image illumination. (2) After applying DWT [10] on the intensity rectified foggy image, DCP with guided filter is applied on low-frequency component (LL band) to get more information with preserved edge as the low-frequency part is affected by fog energy. Dark channel is defined as described by He et al. [1]:

$$J^{Dark}(x) = \min_{y \in \Omega(x)}(\min_{c \in \{R,G,B\}} J^c(y)) \tag{2}$$
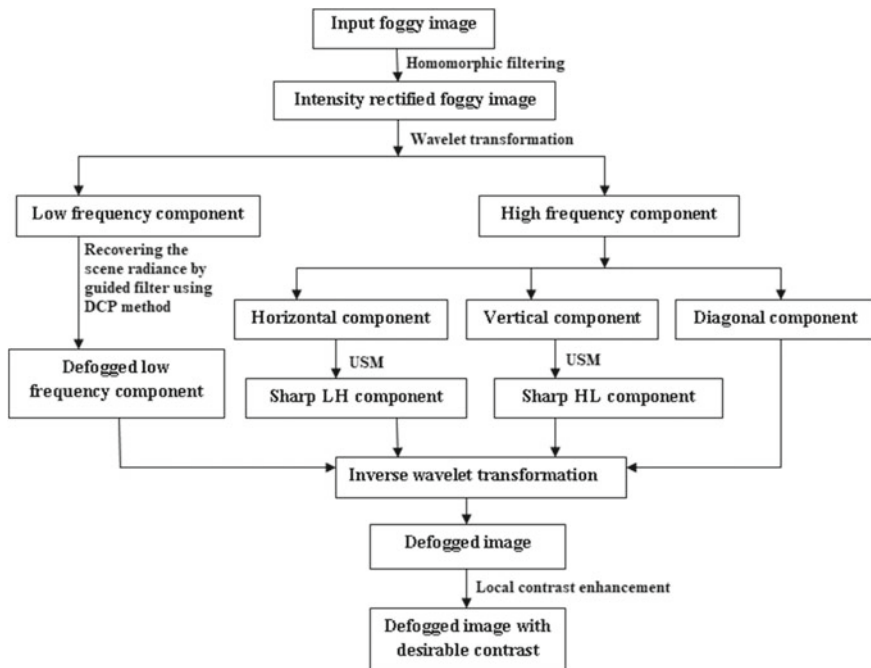
Fig. 1 Block diagram of the proposed fog removal algorithm

where J^c indicates color channel of J and $\Omega(x)$ represents local patch centered at x. DCP is used to estimate the fog depth. Finally, transmission $\tilde{t}(x)$ is estimated as:

$$\tilde{t}(x) = \omega \min_{y \in \Omega(x)} (\min_c \frac{I^c(y)}{A^c}) \tag{3}$$

where ω lies between 0 and 1. In our work, it is considered as 0.95. Because of the halo effects and block artifacts, the transmission map obtained from Eq. (3) is refined by guided filter [11]. Its performance is better near edges compared to Laplacian used in [1]. (3) In this algorithm, instead of getting three high-frequency components (LH, HL, HH), USM technique [12] is applied on LH and HL sub-bands for sharpness enhancement because artifacts are generated for the reason of overshoot in HH component. USM scaling constant value is set to be 1.4. Defogged image is obtained after joining all the modified DWT components by IDWT technique. (4) Optimum visibility is obtained after enhancing the contrast by applying CLAHE [13] on the defogged image. It enhances local contrast and edges in each region or tiles of an image. Here, the tile size is considered as [8 8]. Contrast enhancement limit is introduced to tune or to limit the level of contrast and it is to be set between 0.02 and 0.08.

4 Simulation Results and Performance Analysis

Simulation results of few current techniques along with the proposed method are shown for various test images. We see that the proposed technique performs better with respect to the others qualitatively and quantitatively. The performance of the different techniques is quantified with the help of some quality metrics like (a) contrast gain (CG) [4], (b) color information entropy (CIE) [14] and (c) anisotropic quality index (AQI) [15].

CG is obtained by taking the difference of mean contrast between the defogged output and the input foggy image. The large value of CG implies high-quality performance.

$$CG = C_{I,def} - C_{I,fog} \tag{4}$$

where $C_{I,def}$ and $C_{I,fog}$ are the mean contrast of the defogged output and the foggy input image respectively.

The color image information quantity is measured by CIE. Based on the information quantity, image clarity is defined. So, higher value of CIE signifies better performance.

$$CIE = -\sum_{k=0}^{L-1} P_k log_2(P_k) \tag{5}$$

where L defines the number of intensity levels and P_k is used to indicate the probability associated with intensity level k or histogram count.

AQI metric is a no reference which is sensitive to noise and blur and is one of the important quality metrics for quality analysis. Evaluation of AQI does not require any prior information of the reference image. Higher the value of AQI implies better performance of a method.

Figure 2a–h shows the output images at different steps the proposed algorithm.

The outputs of different techniques for a foggy 'Image 1' and 'Image 2' are shown in Figs. 3 and 4. We feel that the output of the proposed technique (Figs. 3e and 4e) are visually better with respect to the others.

'Image 1' is considered for light fog whereas 'Image 2' has intense fog. Our proposed result reveals more information with a clear view of the scene in both cases compared to the other techniques.

Quantitative performance analysis of different techniques is carried out by evaluating the quality metrics. Tables 1 and 2 show the numerical values of the quality metrics of different algorithms. The value implies that the proposed technique performs better in terms of quality metrics with respect to other recent techniques.
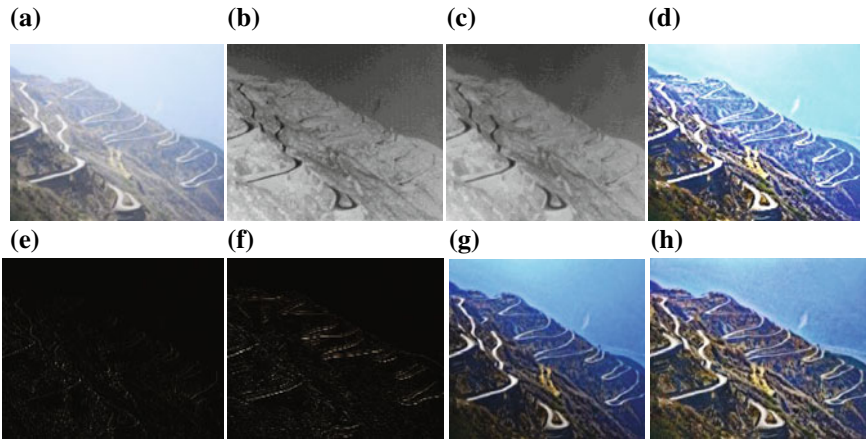
Fig. 2 Image 'juluk': **a** Foggy image, **b** Transmission map, **c** Refined map, **d** Restored LL band, **e** Sharp LH band, **f** Sharp HL band, **g** defogged result, **h** Contrast-enhanced defogged result
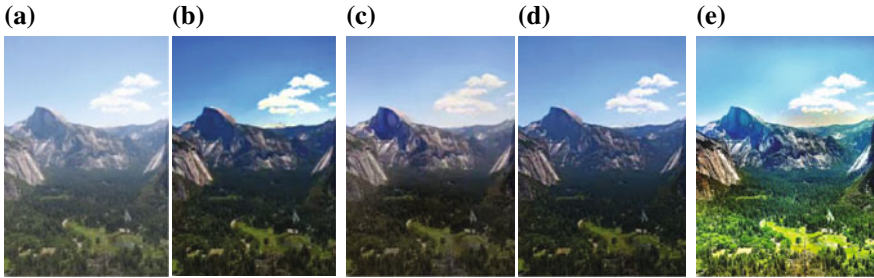
Fig. 3 'Image 1': **a** Foggy image, **b** He et al. [1], **c** Fattal [5], **d** Tripathy and Mukhopadhyay [4], **e** Proposed result
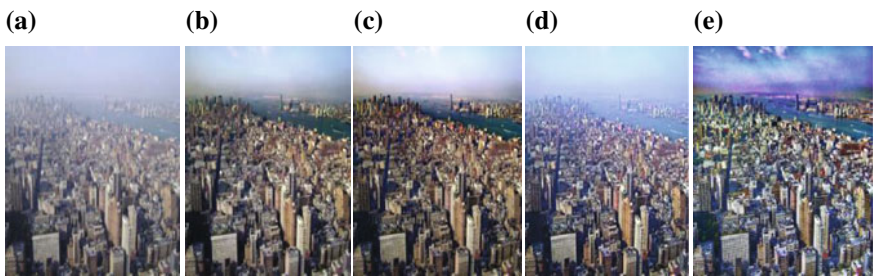
Fig. 4 'Image 2': **a** Foggy image, **b** He et al. [1], **c** Fattal [5], **d** Tripathy and Mukhopadhyay [4], **e** Proposed result

Table 1 Comparison of CG and AQI of previous methods with the proposed method

Method	He et al. [1]		Tripathy and Mukhopadhyay [4]		Fattal [5]		Proposed	
Metrics	CG	AQI	CG	AQI	CG	AQI	CG	AQI
'Image1'	0.08	0.0030	0.11	0.0047	0.04	0.0059	0.18	0.0084
'Image2'	0.02	0.0118	0.16	0.0125	0.02	0.016	0.16	0.0171

Table 2 Comparison of CIE of previous methods with the proposed method

Method	He et al. [1]	Tripathy and Mukhopadhyay [4]	Fattal [5]	Proposed
Metric	CIE	CIE	CIE	CIE
'Image1'	7.431	7.572	7.584	7.715
'Image2'	5.133	7.680	6.335	7.748

5 Conclusions

This article proposes a novel image defogging algorithm which can reduce the fog from the image by preserving the contrast and color information. In the proposed method, dark channel prior method along with guided filter is applied to LL band in wavelet domain of the foggy image for removing the fog and unsharp masking is applied to LH and HL bands for improving the sharpness of the image. The defogged image is retrieved by performing inverse wavelet transform on the processed wavelet coefficients. Finally, the improved defogged image is generated through local contrast enhancement on the defogged image. The performance of the proposed algorithm is compared with some recent image defogging techniques through qualitative and quantitative evaluation of the outputs. The outputs show the proposed technique provides a better result with respect to the other recent techniques.

References

1. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. IEEE Trans Pattern Anal Mach Intell 33(12):2341
2. Tarel JP, Hautiere N (2009) Fast visibility restoration from a single color or gray level image. In: Proceedings of 12th IEEE international conference on computer vision, pp 2201–2208
3. Tan R (2008) Visibility in bad weather from a single image. In: Proceedings of IEEE conference computer vision and pattern recognition, June (2008)
4. Tripathi AK, Mukhopadhyay S (2012) Single image fog removal using anisotropic diffusion. IET Image Proc 6(7):966–975
5. Fattal R (2014) Dehazing using color-lines. ACM Trans Graph 34(1):13
6. Sijiu W, Haiyan Z, Weidong C (2015) Research for image haze-removal algorithm using the dark-channel prior based on wavelet transform. Open Cybernet System J 9:1378–1384

7. Schechner YY, Narasimhan Srinivasa G, Nayar Shree K (2001) Instant dehazing of images using polarization. In: Proceedings of IEEE computer society conference on computer vision and pattern recognition, vol 1, pp 1–325

8. Kopf J, Neubert B, Chen B, Cohen M, Cohen-Or D, Deussen O, Uyttendaele M, Lischinski D (2008) Deep photo: model-based photograph enhancement and viewing. ACM Trans Graph 27(5):116:1–116:10

9. Schreiber WF (1978) Image processing for quality improvements. In: Proceedings of the IEEE 66(12):1640–1651

10. Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans PAMI 7(11):674–693

11. He K, Sun J, Tang X (2010) Guided image filtering. In: Computer vision—ECCV2010, vol 6311, pp 1–14

12. Ying L, Tek M, BengKeat N, Liew G (2008) A wavelet based image sharpening algorithm. In: International conference on computer science and software engineering, pp 1053–1056

13. Zuiderveld, K (1994) Contrast limited adaptive histograph equalization. Graphic Gems IV. San Diego: Academic Press Professional, pp 474–485

14. Wang Y, Fan C (2014) Single image defogging by multiscale depth fusion. IEEE Trans Image Process 23(11):4826–4837

15. Gabarda S, Cristóbal G (2007) Blind image quality assessment through anisotropy. J Opt Soc Am A 24:B42–B51

New Approach to Image Segmentation Based on Neighborhood-Influenced Fuzzy C-Means Clustering

Samiran Kr. Banik, Tiyasa Chakraborty and Debashis Nandi

Abstract In recent years, accurate segmentation of images is a very challenging task for image processing applications. Image segmentation can also be stated as clustering problem in which image pixels are clustered according to the homogeneity of their feature values. Crisp K-means clustering algorithm can achieve the solution of this problem. But it is not suitable for coinciding partition and it is unable to handle noisy data. Fuzzy form of C-means clustering can manage overlapping partition problem, but traditional FCM is also sensitive to noise pixels. In this paper, neighborhood-influenced Fuzzy C-means (NFCM) algorithm is proposed where spatial neighborhood information of pixels is incorporated with the traditional fuzzy c-means algorithm. NFCM is giving more accurate segmentation result compared to hard c-means and fuzzy c-means based segmentation techniques.

Keywords Image segmentation · K-means · FCM · Neighborhood-influenced fuzzy C-means

1 Introduction

Image segmentation is a procedure that splits an image into meaningful non-overlapping sections depending upon some uniformity or similarity measure [1–3]. This is very useful in diverse applications like geographical information system, remote sensing, medical image processing, computer vision, and other image analy-

S. Kr. Banik (✉) · T. Chakraborty (✉) · D. Nandi
National Institute of Technology, Durgapur, Durgapur 713209, West Bengal, India
e-mail: samiran.phd@gmail.com

T. Chakraborty
e-mail: tiyasachakraborty@gmail.com

D. Nandi
e-mail: debashisn2@gmail.com

sis applications. Image segmentation also can be defined as a clustering problem [4], where segmentation is achieved in terms of cluster of pixels that 'belong together' [5]. Different clustering techniques are there that are mainly driven by bivalent or multivalued logic which comprises crisp or hard clustering and fuzzy or soft clustering [16–18].

In hard K-means clustering, data points are assigned into K number of clusters where K is pre-assigned integer value. Each cluster has a centroid point or cluster representative such as the centroid of jth cluster is denoted as C_j. The ith element to be assigned into a cluster is expressed by a feature vector Z_i. The centroid values are updated iteratively to minimize the objective function. The assignment of data points is done depending upon the lowest distance between Z_i to centroids. Thus, the objective function of hard K-means is furnished based on the total variance of data points to its corresponding centroids [6].

$$\text{OBJ}_{Kmeans} = \sum_{i=1}^{n} \sum_{k=1}^{M} dis(C_k, Z_i) \tag{1}$$

where n is the total number of data points, M is the total number of clusters, and $dis(C_k, Z_i)$ is indicating the traditional Euclidian distance between data point Z_i to its corresponding cluster centroid C_k. But K-means is not appropriate for noisy data value or coinciding partitioning problem. To manage those problems, fuzzy C-means clustering is beneficial [7–9].

In real applications, where no crisp or hard boundaries are present between clusters, fuzzy clustering is often fit in those circumstances. In the FCM algorithm, the data items are assigned to more than one cluster associated with some values between 0 and 1 known as membership values or 'degree of belongingness' [10]. The objective function of traditional fuzzy C-means is written as follows [11]:

$$\text{OBJ}_{FCM} = \sum_{i=1}^{n} \sum_{j=1}^{M} \mu_{ij}^{\eta} dis(c_j, z_i) \tag{2}$$

Here, μ_{ij} indicates the membership value of ith data point (i.e., Z_i) for jth cluster (i.e., C_j) and the term η is fuzziness controller or fuzzifier. Each membership value μ_{ij} lies between the closed interval 0–1. For a particular data point, the summation of membership values associated with all clusters is equal to 1. The summation of membership values of all data points for a particular cluster lies between the open interval 0–n. Image segmentation using fuzzy clustering procedure delivers a mean of categorizing pixel data values with a great scope of accuracy [12].

To achieve efficient image segmentation, several research approaches have been introduced by modifying the conventional FCM algorithm. Many researchers have spoken about the effectiveness by modifying the objective function and/or membership function of the FCM. Among them, one popular approach is the integration of spatial constraints [13–15].

In this paper, we use a new version of Fuzzy C-means clustering for image segmentation by incorporating spatial information. Instead of using traditional objective function of fuzzy c-means, neighborhood-influenced objective function is considered. Accordingly, centroid and membership updating equations are also modified.

2 Proposed Method

In this study, a new approach named as neighborhood-influenced Fuzzy C-means (NFCM) is introduced for image segmentation by considering spatial information (Fig. 1).

Here, the objective function of conventional FCM is modified by incorporating the neighborhood influence within it.

$$\text{OBJ}_{NFCM} = \sum_{i=1}^{n} \sum_{j=1}^{M} \mu_{ij}^{\eta} dis(c_j, z_i) + \chi \psi \tag{3}$$

The first part of the above equation is similar to ordinary FCM and the second part includes two new factors in which χ is denoted as neighborhood influence controller and ψ represents the neighborhood influence. The term ψ is formulated as follows:

$$\psi = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij}^{\eta} \left[\frac{1}{|Nb_i|} \sum_{p=1}^{|Nb_i|} dis(c_j, x_p) \right] \tag{4}$$

where $|Nb_i|$ is the number of neighbor associated to ith data point, i.e., Z_i and x_p denotes pth neighbor of Z_i. The membership function and centroids are also modified as

$$\mu_{ij} = \frac{1}{\sum_{r=1}^{M} \left[\left[\frac{dis(c_j, z_i) + \chi \cdot \left[\frac{1}{|Nb_i|} \sum_{p=1}^{|Nb_i|} dis(c_j, x_p) \right]}{dis(c_r, z_i) + \chi \cdot \left[\frac{1}{|Nb_i|} \sum_{p=1}^{|Nb_i|} dis(c_r, x_p) \right]} \right] \right]^{\frac{2}{\eta-1}}} \quad \begin{array}{l} \text{For } l \leq k \leq M; \\ 1 \leq i \leq n \end{array} \tag{5}$$
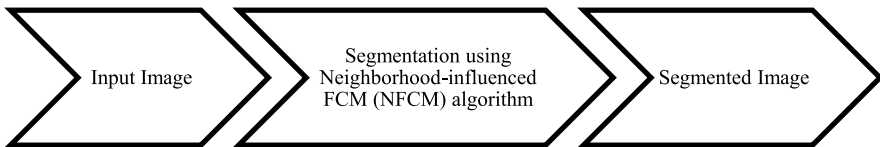
Fig. 1 Block diagram of the proposed methodology

$$c_{j.new} = \frac{\sum_{i=1}^{n} \mu_{ij}^{\eta}(z_i + \chi \cdot \left[\frac{1}{|Nb_i|} \sum_{p=1}^{|Nb_i|} dis(c_{j.old}, x_p) \right])}{(1 + \chi) \sum_{i=1}^{n} \mu_{ij}^{\eta}} \quad \text{For } 1 \leq j \leq M \quad (6)$$

Algorithm NFCM:

Step 1: Initialize fuzzy partitioning matrix μ in random manner, χ as 1, η as 2.
Step 2: Calculate Centroid using Eq. (6)
Step 3: Update Membership using Eq. (5)
Step 4: Calculate Objective function OBJ_{NFCM}
Step 5: If $\left\| (OBJ_{NFCM})^{it+1} - (OBJ_{NFCM})^{it} \right\| < \varepsilon$ then Stop
Otherwise, return to step 2.

where it denotes present iteration number and ε is small threshold value taken as 0.001 for the convergence of the algorithm.

3 Result

In this section, we describe the experimental results where three different images are taken for testing of those image segmentation algorithms. K-means and conventional FCM-based segmentation algorithms are also applied to those test images along with NFCM segmentation algorithm. The Results are presented as follows in which NFCM gives a much better result than other also in visual perspective (Table 1).

Table 1 Comparing NFCM with FCM and K-means based segmentation Algorithm

4 Performance Analysis

Numerous performance metrics have been defined in recent years for measuring the performance of an image segmentation algorithm. In this study, sensitivity, efficiency, and accuracy are computed for NFCM, FCM based, and K-means based segmentation methods. Subsequently, algorithms are compared to each other based on those computed values (Figs. 2 and 3 and Table 2).

Fig. 2 Analysis of segmentation accuracy (%)
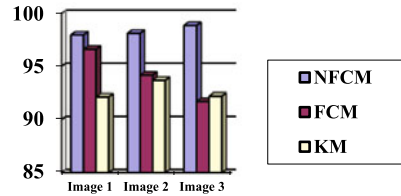
Fig. 3 Analysis of segmentation efficiency (%)
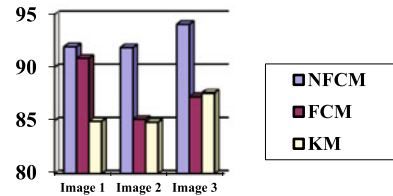
Table 2 Performance Analysis of NFCM, FCM based and K-means based image segmentation algorithm

	Original image		NFCM algorithm	FCM algorithm	KM algorithm
Image 1		Sensitivity	0.21	0.32	0.41
		Accuracy	0.9796	0.9661	0.9213
		Efficiency	0.9194	0.9085	0.8493
Image 2		Sensitivity	0.31	0.42	0.44
		Accuracy	0.9811	0.9417	0.9368
		Efficiency	0.9186	0.8513	0.8485
Image 3		Sensitivity	0.28	0.4	0.36
		Accuracy	0.9885	0.9171	0.9220
		Efficiency	0.9406	0.8724	0.8758

5 Conclusion

In this paper, a new method for segmentation of object region is proposed based on neighborhood-influenced FCM. Influence of neighbor pixels makes the clustering robust to the noise as the spatial information is accounted for calculation of membership and centroid values along with objective function. This proposed algorithm is compared with conventional FCM based and K-means based image segmentation techniques. The tabulated result shows that proposed NFCM algorithm outperforms FCM and K-means based segmentation methods.

References

1. Fu KS, Mui JK (1981) A survey on image segmentation. Pattern Recogn 13(1):3–16
2. Dong G, Xie M (2005) Color clustering and learning for image segmentation based on neural networks. IEEE Trans Neural Netw 16(4):925–936
3. Pal NR, Pal SK (1993) A review on image segmentation techniques. Pattern Recogn 26(9):1277–1294
4. Pan J, McInnes F, Jack M (1996) Fast clustering algorithms for vector quantization. Pattern Recogn 29:511–518
5. Pappas TN (1992) An adaptive clustering algorithm for image segmentation. IEEE Trans Signal Process 40:901–914
6. Hartigan J, Wang M (1979) A K-means clustering algorithm. Appl Statist 28:100–108
7. Ghosh S, Dubey SK (2013) Comparative analysis of K-means and fuzzy C means algorithms. Int J Advanc Comput Sci Appl (IJACSA) 4(4)
8. Zadeh LA (1965) Fuzzy sets. Inform Control 8:338–353
9. Bezdek JC (1981) Pattern recognition with objective fuzzy algorithms. Plenum Press, New York
10. Zhang XB, Jiang L (2009) An image segmentation algorithm based on fuzzy c–means clustering. In: International conference on digital image processing
11. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. IEEE Trans Pattern Anal Mach Intell 13(8)
12. Abdulghafour M (2003) Image segmentation using Fuzzy logic and genetic algorithms. J WSCG 11(1)
13. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T (2002) A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imag 21:193–199
14. Pham DL, Prince JL (1999) Adaptive Fuzzy segmentation of magnetic resonance images. IEEE Trans Med Imag 18(9):737–752
15. Mohamed NA, Ahmed MN, Farag A (1999) Modified Fuzzy C means in medical image segmentation. In: Proceedings of international conference acoustics, speech and signal process, pp 3429–3432
16. Mohamed Jafar OA, Sivakumar R (2013) A comparative study of hard and fuzzy data clustering algorithms with cluster validity indices. In: Proceedings of international conference on emerging research in computing, information, communication and applications. Elsevier Publications, pp 775–782
17. Gong M, Tian D, Su L, Jiao L (2015) An efficient bi-convex fuzzy variational image segmentation method. Inf Sci 293(2015):351–369
18. Rubio E, Castillo O, Valdez F, Melin P, Gonzalez CI, Martinez G (2017) An extension of the fuzzy possibilistic clustering algorithm using type-2 fuzzy logic techniques. Advanc Fuzzy Syst 2017

Morphological Change Detection in Terror Camps of Area 1 and 2 by Pre- and Post-strike Through MOAB: A

Amit Kumar Shakya, Ayushman Ramola, Akhilesh Kandwal, Prag Mittal and Rishi Prakash

Abstract Change Detection (CD) techniques have emerged out as an effective tool to monitor various positives and negatives aspects of any event, whether these events are related to the development or destruction of human civilization. These techniques provide us binary (Yes–No) and detailed information about various types of statistical changes that get developed in the land cover after any event. In this context, MOAB (Mother of All Bomb) strike on Islamic State militants in Achin, District of Nangarhar province, Afghanistan can be referred as a step taken to establish peace and stability in that disturbed region. Here, the morphological changing conditions are analysed by the texture based post classification CD technique by comparing the second-order statistics of pre-strike and post-strike images of Area 1 and Area 2. A novel technique to compute circular texture varying from 0° to 360° is fused with Decorrelation Stretcher (DS) technique. The possible hideouts of the terrorist before the strike and demolished structures after strike are highlighted by colour processing in Red, Green, and Blue (RGB) band forms. Quantification of the texture features along with visual interpretation through DS establishes a pattern in the changing texture features of the land cover in the affected areas.

Keywords Change detection · MOAB · Circular texture · Decorrelation stretcher · Colour processing

1 Introduction

CD from the Earth environment is an important area of investigation right from the beginning of human civilization. Nowadays, these investigation techniques have crossed their conventional limits and are also playing a very important role in the field of defence and security. Today, World's powerful nations are working on upgrading

A. K. Shakya (✉) · A. Ramola · A. Kandwal · P. Mittal · R. Prakash
Department of Electronics and Communication, Graphic Era (Deemed to Be University),
Dehradun 248002, Uttarakhand, India
e-mail: xlamitshakya.gate2014@gmail.com; xlamitshakya.gate2014@ieee.org

their military satellite communication systems skills to have an eye to counter global terrorism. Terrorism has now become a global threat to World economic, social and cultural structure. Iraq, Afghanistan, Nigeria, Pakistan, and Syria are the World top five countries affected by terrorist activities [1]. The attack on the World Trade Centre [2], on November 13, 2015, Paris attack [3] of 2008, and Mumbai attack [4] are some of the major terrorist incidents happened across the World in the last decades. Afghanistan ranks second in the list with Global terrorism index (GTI) with a score of 9,444, and shows an increase of (+0.229) in the various forms of terrorist activities from 2015 to 2016. Our country score in terms of GTI was 7,484 showing a decline of (−0.059) [1]. The terrorist activities across the Globe have cost a tremendous loss of innocent human life and imposed unnecessary financial burdens on the World economy. The funds which were to be used for feeding the hungry and poor are now being invested in the security and weaponry.

1.1 Background of Mother of All Bombs (MOAB)

The MOAB is technically known as Massive Ordnance Air Blast. It was developed by the United States Air Defence expert, A. L. Weimorts for the United States Military. Till today, it is the World biggest non-nuclear weapon, designed to deliver from MC-130 Hercules aircraft [5] (Fig. 1).

Fig. 1 Pictorial representation of attack by MOAB

2 Change Detection Through Second-Order Statistical Parameters

CD from the satellite images has found a lot of applicability in the field of remote sensing. Techniques like deep learning through the artificial neural network (ANN) [6], object-based CD [7], image fusion CD [8], and grey level co-occurrence matrix (GLCM) [9] are widely used for detecting changes in the land cover. Here, a novel version of GLCM is used by computing texture features (TF) along 0°–360°. GLCM was invented by Haralick [10]. He proposed 14 different statistical features related to texture classification. These features are gathered in the category of contrast, correlation, energy, homogeneity, sum-average, sum-variance, homogeneity, sum entropy, image contrast, correlation measures, and maximum correlation coefficients [10]. These features are further grouped into four different categories by scientist Gotlieb [11]. These categories include measures of correlation, entropy, statistical parameters, and visual characteristics of texture.

Let an image contain 'm' numbers of rows and 'n' number of the column is expressed as $I(m, n)$. Then, the total number of pixels under consideration is $(m \times n)$. P_p is the pixel position in $(m \times n)$ image matrix, where (1×1) is the first pixel and $(m \times n)$ is the last image pixel. Now, the TF responsible for the visual variation of the images is computed through GLCM, usually, the GLCM of an image is computed along four different directions but we have extended the orientation area and computed GLCM features along eight directions. The eight directions are represented in form of eight angles, i.e., (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) (Fig. 2).

These variations are evaluated on the basis of changes in the pixel values. Four textural features responsible for the change in visual interpretation are contrast, correlation, energy, and homogeneity, they are expressed mathematically as under.

Fig. 2 Diagrammatic representation of different directions
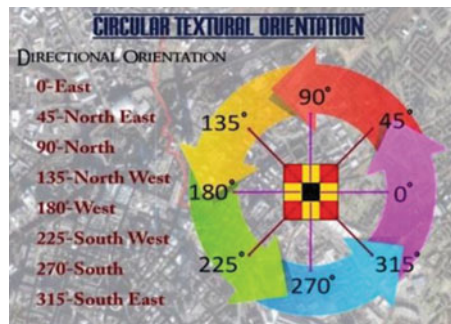
1. ***Contrast***

$$f_m = \sum_{P_p=1\times1}^{P_p=m\times n} \times d_g^2 \times \left\{ \sum_{P_p=1\times1}^{P_p=m\times n} \sum_{P_p=1\times1}^{P_p=m\times n} I(m,n) \right\} \tag{1}$$

where $|m - n| = d_g^2$, this indicates the difference in the grey levels of image, pixels.

2. ***Correlation***

$$f_m = \sum_{P_p=1}^{m} \sum_{P_p=1}^{n} \frac{(m,n)I(m,n) - \mu_r \mu_c}{\sigma_r \sigma_c} \tag{2}$$

where μ_r, σ_r and μ_c, σ_c represents mean and standard deviation of row and column respectively. P_p is the position of the pixel in the image. The range of the normalize correlation is in between $[-1, 1]$.

3. ***Energy***

$$f_m = \sum_{P_p=1}^{m} \times \sum_{P_p=1}^{n} |I(m,n)|^2 \tag{3}$$

This feature provides us mean squared value of the image pixel. This feature is also known as Angular Second Moment (ASM), and the normalize range lies in the interval [0, 1].

4. ***Homogeneity***

$$f_m = \sum_{P_p=1}^{m} \sum_{P_p=1}^{n} \frac{1}{1 + (m - n)^2} \times I(m,n) \tag{4}$$

This feature is also known as Inverse Difference Moment (IDM). The normalize value of this feature lies in the range [0, 1].

3 Experimental Results

In this experiment, we have included Area 1 and Area 2 of terrorist hideout which was severely damaged by the MOAB explosion (Figs. 3 and 4).

Now, the TF is calculated for the pre- and post-images of Area 1 and Area 2 for (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°) directions following a circular pattern (Tables 1, 2, 3 and 4).
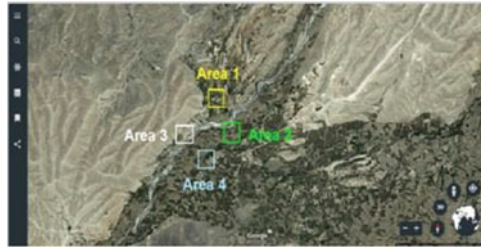
Fig. 3 Area 1 (Yellow) and Area 2 (Green) under the explosion impact
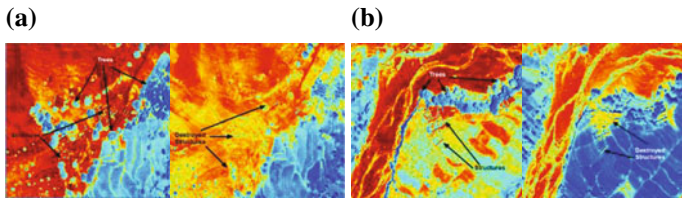
(a)   **(b)**

Fig. 4 **a**, **b** Pre-Strike image acquired on 10, November 2016 and Post-strike image acquired on 14, April 2017

Table 1 Variation in the texture features of the pre-strike image of Area 1

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.2739	0.4200	0.5362	0.6660	0.7831	0.8896	0.9850	1.0719
Corr#		0.9654	0.9469	0.9321	0.9156	0.9008	0.8873	0.8752	0.8642
E#		0.1465	0.1220	0.1094	0.1016	0.0960	0.0918	0.0886	0.0859
H#		0.9154	0.8617	0.8288	0.8061	0.7885	0.7745	0.7630	0.7533
Con#	d = 2	0.4137	0.6691	0.7987	0.8954	1.0312	1.1525	1.2625	1.3629
Corr#		0.9478	0.9153	0.8987	0.8862	0.8690	0.8535	0.8396	0.8268
E#		0.1330	0.1093	0.0979	0.0911	0.0864	0.0829	0.0801	0.0778
H#		0.8877	0.8305	0.7973	0.7747	0.7577	0.7443	0.7333	0.7238
Con#	d = 3	0.2520	0.4764	0.6010	0.6727	0.7921	0.9018	1.004	1.0916
Corr#		0.9682	0.9398	0.9241	0.9149	0.8998	0.8859	0.8735	0.8620
E#		0.1456	0.1206	0.1073	0.0992	0.0936	0.0893	0.0861	0.0835
H#		0.9157	0.8618	0.8276	0.8039	0.7857	0.7708	0.7590	0.7491
Con#	d = 4	0.4391	0.7263	0.8708	0.9734	1.1137	1.2407	1.3571	1.4658
Corr#		0.9446	0.9080	0.8896	0.8764	0.8585	0.8424	0.8277	0.8139
E#		0.1303	0.1069	0.0966	0.0902	0.0856	0.0818	0.0789	0.0765
H#		0.8809	0.8223	0.7909	0.7697	0.7533	0.7393	0.7280	0.7182

Table 2 Variation in the texture features of the post-strike image of Area 1

TF	D	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.1770	0.3148	0.3677	0.3974	0.4473	0.4896	0.5262	0.5593
Corr#		0.9609	0.9304	0.9186	0.9118	0.9007	0.8913	0.8831	0.8757
E#		0.2112	0.1853	0.1716	0.1630	0.1571	0.1527	0.1493	0.1465
H#		0.9335	0.8917	0.8673	0.8504	0.8380	0.8282	0.8206	0.8140
Con#	d = 2	0.3148	0.5552	0.6090	0.5819	0.6330	0.6746	0.7101	0.7421
Corr#		0.9302	0.8763	0.8641	0.8693	0.8577	0.8482	0.8401	0.8328
E#		0.1921	0.1685	0.1576	0.1510	0.1463	0.1427	0.1400	0.1377
H#		0.9023	0.8574	0.8352	0.8208	0.8095	0.8006	0.7940	0.7884
Con#	d = 3	0.2228	0.4109	0.4807	0.4826	0.5307	0.5700	0.6046	0.6365
Corr#		0.9508	0.9090	0.8935	0.8927	0.8819	0.8732	0.8654	0.8583
E#		0.2048	0.1784	0.1653	0.1573	0.1521	0.1482	0.1451	0.1425
H#		0.9248	0.8803	0.8562	0.8402	0.8284	0.8194	0.8119	0.8056
Con#	d = 4	0.3099	0.5506	0.6085	0.5847	0.6372	0.6811	0.7210	0.7587
Corr#		0.9313	0.8773	0.8642	0.8687	0.8569	0.8469	0.8379	0.8294
E#		0.1938	0.1705	0.1587	0.1516	0.1469	0.1432	0.1402	0.1379
H#		0.9047	0.8605	0.8374	0.8222	0.8111	0.8019	0.7943	0.7881

Table 3 Variation in the texture features of the pre-strike image of Area 2

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.2738	0.4054	0.5221	0.6441	0.0768	0.8612	0.9580	1.0483
Corr#		0.9530	0.9302	0.9100	0.8889	0.8695	0.8514	0.8347	0.8191
E#		0.1160	0.0932	0.0819	0.0752	0.0705	0.0668	0.0639	0.0615
H#		0.9110	0.8556	0.8214	0.7982	0.7801	0.7652	0.7528	0.7423
Con#	d = 2	0.4122	0.6451	0.7606	0.8374	0.9609	1.0738	1.1793	1.2782
Corr#		0.9291	0.8885	0.8683	0.8547	0.8332	0.8136	0.7954	0.7782
E#		0.1040	0.0834	0.0734	0.0673	0.0629	0.0595	0.0567	0.0545
H#		0.8826	0.8259	0.7922	0.7688	0.7500	0.7344	0.7209	0.7094
Con#	d = 3	0.2644	0.4980	0.6231	0.6925	0.8113	0.9193	1.0183	1.1094
Corr#		0.9546	0.9143	0.8928	0.8806	0.8601	0.8415	0.8245	0.8088
E#		0.1147	0.0911	0.0790	0.0719	0.0668	0.0629	0.0599	0.0575
H#		0.9887	0.8505	0.8137	0.7886	0.7686	0.7522	0.7386	0.7271
Con#	d = 4	0.4477	0.7284	0.8799	0.9839	1.1314	1.2670	1.3941	1.5130
Corr#		0.9230	0.8741	0.8477	0.8292	0.8036	0.7801	0.7580	0.7374
E#		0.1001	0.0792	0.0698	0.0639	0.0598	0.0565	0.0539	0.0517
H#		0.8719	0.8107	0.7767	0.7530	0.7343	0.7188	0.7053	0.6937

Table 4 Variation in the texture features of the post-strike image of Area 2

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.2007	0.3740	0.4873	0.5733	0.6641	0.8529	0.9327	1.0044
Corr#		0.9707	0.9453	0.9287	0.9161	0.9028	0.8974	0.8878	0.8791
E#		0.1276	0.1073	0.0972	0.0906	0.0858	0.0928	0.0902	0.0880
H#		0.9182	0.8682	0.8382	0.8166	0.7997	0.7794	0.7684	0.7592
Con#	d = 2	0.3537	0.6492	0.7829	0.8176	0.9099	1.1175	1.1950	1.2631
Corr#		0.9481	0.9040	0.8843	0.8791	0.8654	0.8643	0.8549	0.8466
E#		0.1131	0.0953	0.0865	0.0812	0.0775	0.0871	0.0851	0.0834
H#		0.8834	0.8308	0.8014	0.7818	0.7666	0.7517	0.7423	0.7342
Con#	d = 3	0.2658	0.5118	0.6500	0.7082	0.9171	1.0101	1.0920	1.1645
Corr#		0.9611	0.9247	0.9044	0.8957	0.8891	0.8778	0.8678	0.8590
E#		0.1228	0.1020	0.0921	0.0859	0.0925	0.0897	0.0874	0.0856
H#		0.9065	0.8516	0.8205	0.7989	0.7758	0.7633	0.7528	0.7442
Con#	d = 4	0.3642	0.6859	0.8479	0.9038	1.1670	1.2752	1.3720	1.4623
Corr#		0.9465	0.8986	0.8747	0.8664	0.8584	0.8453	0.8335	0.8225
E#		0.1131	0.0959	0.0876	0.0821	0.0885	0.0856	0.0833	0.0813
H#		0.8812	0.8276	0.7976	0.7762	0.7511	0.7384	0.7278	0.7183

Notations Con# = Contrast, Corr# = Correlation, E# = Energy, H# = Homogeneity, d = distance, D = Degree

3.1 Comparative Plot for Pre and Post-strike Texture Feature Variations for Area 1

See Fig. 5.

3.2 Comparative Plot for Pre and Post-strike Texture Feature Variations for Area 2

See Fig. 6.

3.3 Image Enhancement Using Decorrelation Stretcher (DS)

Decorrelation stretcher is an image enhancement process through which the colour qualities of the images are improved for better visualization [12]. It is considered as an effective technique for satellite image enhancement (Figs. 7 and 8).

Fig. 5 Area 1 **a** Pre strike
Con# **b** Post strike Con#
c Pre strike Corr# **d** Post
strike Corr# **e** Pre strike E#
f Post strike E# **g** Pre strike
H# **h** Post strike H#
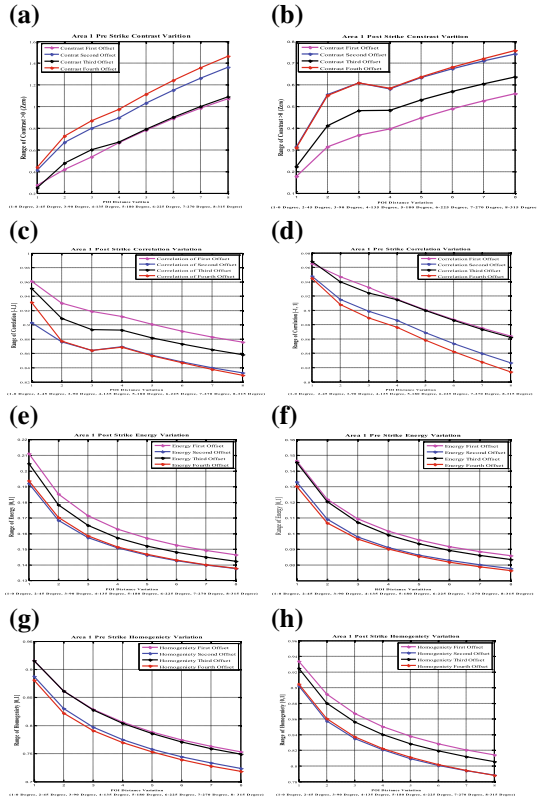
3.4 Colour Maps Through Decorrelation Stretcher

In the pre-strike image of Area 1 terrorist camps are represented by red squares, these
are covered with trees and shrubs due to which they are not easy to get monitored
by the satellite system. After the explosion, the morphology of the area suggests
complete destruction shown in Fig. 9b.

In the pre-strike image of Area 2 terrorist camps can be seen close to the river
bank but soon after the explosion, this area also got completely destroyed (Fig. 10).

Finally, it is concluded that initially, the TF contrast has increased whereas texture
features correlation, energy, and homogeneity have decreased. This pattern is shown
in Fig. 11 which is obtained by the GLCM features average values.

Fig. 6 Area 2 **a** Pre strike
Con# **b** Post strike Con#
c Pre strike Corr# **d** Post
strike Corr# **e** Pre strike E#
f Post strike E# **g** Pre strike
H# **h** Post strike H#
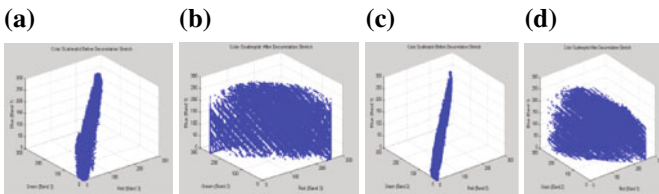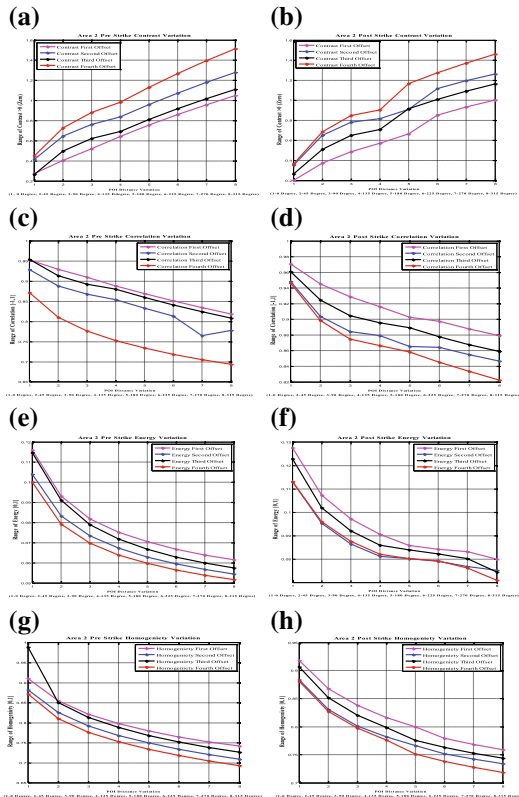
Fig. 7 Area 1 **a** RGB scatter plot of pre-strike image. **b** RGB scatter plot of pre-strike image after applying DS. **c** RGB scatter plot of post-strike image. **d** RGB scatter plot of post-strike image after applying DS

4 Conclusion

The proposed scheme of the CD provides us information in term of quantitative and visual parameters. This method of texture classification through GLCM provides detailed information about second-order statistical parameter changes. The Area 1 and Area 2 were severely damaged during the explosion due to which severe change in the texture of the LC has occurred. DS proves to be an effective technique for the
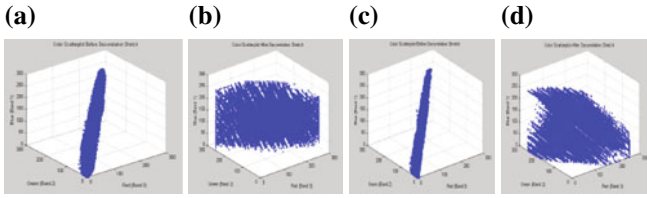
Fig. 8 Area 2 **a** RGB scatter plot of pre-strike image. **b** RGB scatter plot of pre-strike image after applying DS. **c** RGB scatter plot of post-strike image. **d** RGB scatter plot of post-strike image after applying DS

Fig. 9 **a** Pre-strike DS image for Area 1. **b** Post-strike DS for image Area 1
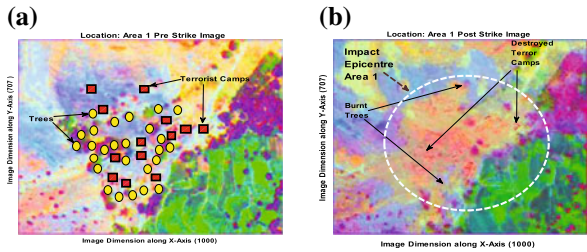
Fig. 10 **a** Pre-strike DS image for Area 2. **b** Post-strike DS for image Area 2
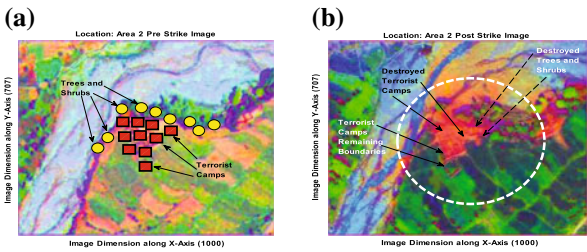
Fig. 11 Pattern obtained in the texture features
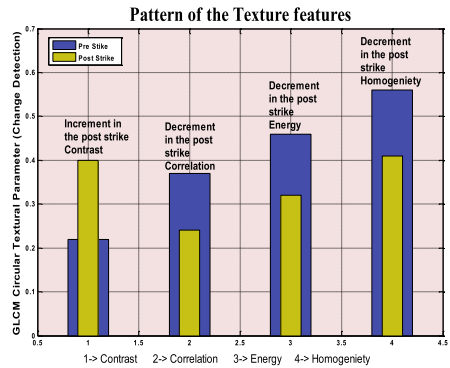
image enhancement. The main advantage of this technique is easy to implement and reliable. The shortcoming of this technique can be the interpretation of the changing pattern with low-resolution satellite images and which can be an area of investigation for further research.

References

1. USA M (2016) Global terrorism index 2016. Survey, Institute for Economics and Peace, Maryland
2. Wikipedia: GBU-43/B MOAB. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/GBU-43/B_MOAB. Accessed 30 Mar 2018
3. Philippe J-M, Philippe J-M, Philippe J-M, Tourtier J-P, Riou B (2016) French Ministry of Health's response to Paris attacks of 13 November 2015, French Ministry of Health, April 2016
4. Wikipedia: 2008 Mumbai attacks. Wikipedia, the free Encyclopaedia. https://en.wikipedia.org/wiki/2008_Mumbai_attacks. Accessed 26 Jan 2018
5. Wikipedia T: GBU-43/B MOAB. In: Wikipedia. https://en.wikipedia.org/wiki/GBU-43/B_MOAB. Accessed 6 May 2017
6. Li W, Wu G, Du Q (2017) Transferred deep learning for anomaly detection in hyper-spectral imagery. Remote Sens Lett 5
7. Yousif O, Ban Y (2018) A novel approach for object-based change image generation using multi-temporal high-resolution SAR images. Int J Remote Sens 38(7):1765–1787
8. Carreiras J, Jones J, Lucas R, Shimabukuro Y (2017) Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data. Remote Sens Environ 194:16–32
9. Wellmanna T, Haase D, Knapp S, Salbach C (2018) Urban land use intensity assessment: the potential of spatial-temporal spectral traits with remote sensing. Ecol Ind 28:190–203
10. Howlick R (1979) Statistical and structural approaches to texture. Proc IEEE 67(5):786–804
11. Gotlieb CC, Kreyszi EH (1990) Texture descriptors based on co-occurrence matrices. Comput Vision Graph Imag Process 51(1):70–86
12. Alley R (1999) Algorithm theoretical basis document for decorrelation stretch. Pasadena, CA 9110

Morphological Change Detection in Terror Camps of Area 3 and 4 by Pre- and Post-strike Through MOAB: B

Amit Kumar Shakya, Ayushman Ramola, Akhilesh Kandwal, Prag Mittal and Rishi Prakash

Abstract Change Detection (CD) techniques have attracted researcher to monitor the changes in the land use and land cover. These techniques provide us binary as well as detailed information about various types of changes. In the part A, we have monitored the textual changes that have occurred in the land cover of Area 1 and Area 2 due to MOAB (Mother of All Bombs) bombing. Here, we have continued the approach to identify the pattern of the changing texture of Area 3 and Area 4. Here, we have evaluated the GLCM (Gray level co-occurrence matrix) features of the land cover, followed by image enhancement using Decorrelation stretcher (DS). Later, through DS images terrorist locations are identified and highlighted. Finally, the patterns of the changing texture provide information about changes that have occurred in the texture features (TF) of the land cover.

Keywords Change detection · Mother of all bomb · Gray level Co-occurrence matrix · Decorrelation stretcher · Texture features

1 Introduction

Earth observation is an important area of investigation right from the beginning of human civilization [1]. This intentionally gives rise to the development of several CD techniques. Nowadays, CD techniques are classified into two streams pre-classification CD and post-classification CD [2]. The pre-classification techniques provide us only binary information about changes, but post-classification techniques provide us detailed information about several types of changes like change in texture, tone, pattern, etc. Pre-classification techniques are pixel-based CD technique whereas post-classification CD techniques are both pixel-based and object-based CD. Pre-classification techniques include image differencing [3], image ratio [4], change

A. K. Shakya (✉) · A. Ramola · A. Kandwal · P. Mittal · R. Prakash
Department of Electronics and Communication, Graphic Era (Deemed to be University),
Dehradun 248002, Uttrakhand, India
e-mail: xlamitshakya.gate2014@gmail.com

vector analysis [5], vegetation index differencing [6], and principal component analysis [7]. Post-classification CD techniques include machine learning [8], multi-date classification [9], GIS-based CD technique [10], GLCM-based CD technique [11], and deep learning [12]. Here, we are using GLCM based post-classification technique to identify the pattern of the changing texture of Area 3 and Area 4, which got devastated by the MOAB explosion. We have analyzed the change in the texture pattern through changes in the texture features, i.e., contrast, correlation, energy, and homogeneity, which are texture visual features. Later, through DS images, we have identified exact locations of the terrorist camps which are highlighted through RGB, color scheme. Finally, the average of GLCM features is computed to detect the pattern of change developed in the land cover through bombarding in the region.

2 Methodology for Working of GLCM (Gray Level Co-occurrence Matrix)

GLCM was invented by scientist Haralick, and he proposed 14 features for texture classification. Later, another scientist C.C Gotlieb classified these features into 4 different groups, based on correlation, entropy, statistical parameters, and visual characteristics of texture.

Now, Fig. 2 shows how to create GLCM from the input image shown in Fig. 1. Let an image 'i' of dimension (6×6) is to be converted into a GLCM image, then the procedure followed will be 'one pixel immediately to the right'. The newly created image after applying GLCM is shown in Fig. 2.

The GLCM for the image is usually calculated across four directions, we have calculated GLCM across eight directions. These directions are North, South, East, West, Northeast, Northwest, Southeast, and Southwest. The Offsets combinations for these directions are shown in Table 1.

Fig. 1 Input image 'i'

Fig. 2 GLCM obtained
from input image 'i'

	1	2	3	4	5	6	7	8
1	0	1	0	0	0	0	1	1
2	3	0	0	1	0	0	0	0
3	0	1	0	1	0	0	0	0
4	0	0	0	0	1	0	2	0
5	1	0	0	0	0	0	0	2
6	1	0	0	0	0	0	1	1
7	1	1	0	1	1	1	0	0
8	0	0	1	1	0	1	0	1

Table 1 Offsets combination along with the angle and distance orientations

Sl. no.	Offset	Angle 0°	Angle 45°	Angle 90°	Angle 135°	Angle 180°	Angle 225°	Angle 270°	Angle 315°
1	[0, D]	[0, 1]	[0, 2]	[0, 3]	[0, 4]	[0, 5]	[0, 6]	[0, 7]	[0, 8]
2	[−D, D]	[−1, 1]	[−2, 2]	[−3, 3]	[−4, 4]	[−5,5]	[−6, 6]	[−7, 7]	[−8, 8]
3	[−D, 0]	[−1, 0]	[−2, 0]	[−3, 0]	[−4, 0]	[−5, 0]	[−6, 0]	[−7,0]	[−8, 0]
4	[−D, −D]	[−1, 1]	[−2, −2]	[−3, −3]	[−4, −4]	[−5, −5]	[−6, −6]	[−7,7]	[−8, −8]

Fig. 3 Area 3 and Area 4
under the explosion impact
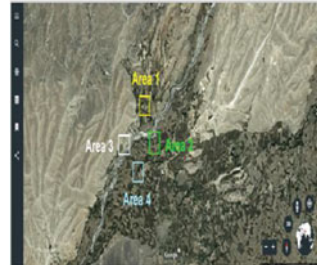
3 Experimental Results

Now we have investigated the variation in the texture of Area 3 and Area 4, after MOAB bombing (Figs. 3 and 4).

In Area 3, the bomb explosion has severely damaged the inner dimension of the terrorist camps, buildings, and structures, but the outer dimensions of the terrorist camps are not affected to a great extent. Finally, in Area 4, it also seems that the terrorist compounds have retained their outer boundaries, but the inner structures of the camps are destroyed. Now, the variations of the TF for the pre- and post-strike images of area 3 and area 4 are evaluated (Tables 2, 3, 4 and 5).

Table 2 Variation in the texture features for the pre-strike image of Area 3

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.3641	0.6476	0.9355	1.1833	1.3949	1.5806	1.7483	1.8994
Corr#		0.9516	0.9138	0.8756	0.8427	0.8146	0.7901	0.7679	0.7480
E#		0.0989	0.0788	0.0696	0.0641	0.0603	0.0573	0.0548	0.0529
H#		0.8730	0.8063	0.7683	0.7432	0.7247	0.7096	0.6972	0.6869
Con#	d = 2	0.6413	1.0419	1.2504	1.5251	1.7624	1.9697	2.1519	2.3108
Corr#		0.9147	0.8611	0.8324	0.7957	0.7640	0.7364	0.7121	0.6909
E#		0.0889	0.0710	0.0632	0.0583	0.0548	0.0522	0.0502	0.0485
H#		0.8414	0.7759	0.7404	0.7159	0.6977	0.6827	0.6711	0.6609
Con#	d = 3	0.4480	0.7757	0.9123	1.1352	1.3338	1.5129	1.6755	1.8213
Corr#		0.9404	0.8968	0.8779	0.8481	0.8215	0.7975	0.7757	0.7561
E#		0.1011	0.0800	0.0709	0.0651	0.0612	0.0582	0.0558	0.0539
H#		0.8791	0.8123	0.7760	0.7506	0.7317	0.7165	0.7041	0.6936
Con#	d = 4	0.6897	1.1379	1.3700	1.6620	1.9156	2.3231	2.3231	2.4887
Corr#		0.9082	0.8483	0.8164	0.7740	0.7435	0.6892	0.6671	0.6671
E#		0.0860	0.0687	0.0612	0.0563	0.0527	0.0477	0.0460	0.0460
H#		0.8325	0.7665	0.7322	0.7076	0.6886	0.6602	0.6498	0.6498

Table 3 Variation in the texture features for the post-strike image of Area 3

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.3636	0.6752	1.3285	1.1236	1.3127	1.4780	1.6243	1.7513
Corr#		0.9439	0.8959	0.8268	0.7925	0.7977	0.7724	0.7402	0.7305
E#		0.0909	0.0702	0.0558	0.0559	0.0523	0.0497	0.0476	0.0459
H#		0.8616	0.7942	0.7224	0.7302	0.7112	0.6961	0.6833	0.6730
Con#	d = 2	0.7110	1.0824	1.1906	1.5410	1.7122	1.8554	1.9762	2.0785
Corr#		0.8899	0.8316	0.8153	0.7595	0.7329	0.7108	0.6921	0.0763
E#		0.0785	0.0627	0.0616	0.0517	0.0489	0.0468	0.0451	0.0438
H#		0.8184	0.7551	0.7512	0.7003	0.6840	0.6710	0.6603	0.6510
Con#	d = 3	0.5269	0.9317	1.2830	1.4026	1.4426	1.5825	1.7069	1.8155
Corr#		0.9187	0.8559	0.8004	0.7810	0.7756	0.7539	0.7345	0.7176
E#		0.0906	0.0699	0.0540	0.0568	0.0536	0.0511	0.0493	0.0478
H#		0.8558	0.7863	0.7151	0.7283	0.7116	0.6982	0.6879	0.0782
Con#	d = 4	0.7144	1.2210	0.9056	1.6588	1.8644	2.0308	2.1656	2.2758
Corr#		0.8894	0.8105	0.8604	0.7411	0.7092	0.6834	0.6625	0.6456
E#		0.0776	0.0613	0.0611	0.0498	0.0470	0.0449	0.0435	0.0422
H#		0.8161	0.7502	0.7359	0.6913	0.6742	0.6606	0.6499	0.6411

Table 4 Variation in the texture features for the pre-strike image of Area 4

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.3332	0.5047	0.6485	0.7994	0.9341	1.0545	1.1604	1.2551
Corr#		0.9338	0.8992	0.8703	0.8401	0.8132	0.7892	0.7680	0.7491
E#		0.1261	0.0995	0.0878	0.0814	0.0772	0.0741	0.0718	0.0700
H#		0.8958	0.8338	0.7982	0.7754	0.7586	0.7454	0.7351	0.7266
Con#	d = 2	0.5000	0.7963	0.9625	1.0674	1.0168	1.3483	1.4643	1.5635
Corr#		0.9004	0.8404	0.8066	0.7843	0.7542	0.7278	0.7044	0.6846
E#		0.1118	0.0897	0.0797	0.0742	0.0706	0.0679	0.0659	0.0643
H#		0.8630	0.8020	0.7671	0.7451	0.7291	0.7163	0.7059	0.6974
Con#	d = 3	0.3323	0.6446	0.8449	0.9539	1.1109	1.2497	1.3725	1.4821
Corr#		0.9339	0.8716	0.8316	0.8087	0.7773	0.7495	0.7250	0.7031
E#		0.1205	0.0938	0.0821	0.0762	0.0723	0.0695	0.0673	0.0656
H#		0.8856	0.8181	0.7792	0.7555	0.7382	0.7247	0.7142	0.7051
Con#	d = 4	0.5679	0.9470	1.1505	1.2753	1.4474	1.6037	1.7429	1.8676
Corr#		0.8864	0.8102	0.7688	0.7422	0.7076	0.6762	0.6482	0.6232
E#		0.1038	0.0821	0.0743	0.0700	0.0671	0.0647	0.0630	0.0616
H#		0.8435	0.7758	0.7441	0.7246	0.7100	0.6977	0.6880	0.6796

Table 5 Variation in the texture features for the post-strike image of Area 4

TF	d	0 D	45 D	90 D	135 D	180 D	225 D	270 D	315 D
Con#	d = 1	0.2924	0.5347	0.6875	0.8070	0.9326	0.0389	0.1317	0.2139
Corr#		0.9486	0.9058	0.8789	0.8577	0.8356	0.8168	0.8005	0.7860
E#		0.1058	0.0826	0.0732	0.0677	0.0642	0.0618	0.0599	0.0585
H#		0.8855	0.8224	0.7896	0.7668	0.7505	0.7383	0.7282	0.7199
Con#	d = 2	0.4992	0.9081	1.0974	1.1495	1.2693	1.3715	1.4585	1.5354
Corr#		0.9117	0.8384	0.8048	0.7948	0.7734	0.7552	0.7396	0.7259
E#		0.0886	0.0712	0.0647	0.0612	0.0589	0.0573	0.0560	0.0548
H#		0.8392	0.7768	0.7482	0.7309	0.7182	0.7087	0.7008	0.6938
Con#	d = 3	0.3890	0.7706	0.9952	1.0976	1.2253	1.3285	1.4202	1.5048
Corr#		0.9314	0.8635	0.8238	0.8052	0.7825	0.7642	0.7479	0.7329
E#		0.0964	0.0737	0.0662	0.0622	0.0598	0.0582	0.0568	0.0555
H#		0.8617	0.7891	0.7568	0.7374	0.7242	0.7144	0.7058	0.6979
Con#	d = 4	0.5461	1.0142	1.2458	1.3260	1.4690	1.5917	1.6974	1.7928
Corr#		0.9035	0.8195	0.7784	0.7633	0.7378	0.7159	0.6970	0.6800
E#		0.0854	0.0685	0.0623	0.0590	0.0567	0.0550	0.0536	0.0523
H#		0.8292	0.7652	0.7358	0.7183	0.7046	0.6936	0.6846	0.6760

Notations *Con#* = Contrast, *Corr#* = Correlation, *E#* = Energy, *H#* = Homogeneity, *D* = Degree, *d* = distance)
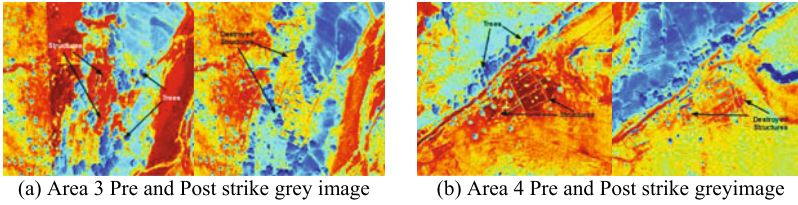
(a) Area 3 Pre and Post strike grey image   (b) Area 4 Pre and Post strike greyimage

Fig. 4 **a**, **b** Pre-strike image acquired on 10, November 2016 and post-strike image acquired on 14, April 2017

Fig. 5 Area 3 **a** Pre-strike contrast **b** Post-strike contrast **c** Pre-strike correlation **d** Post-strike correlation **e** Pre-strike energy **f** Post-strike energy **g** Pre-strike homogeneity **h** Post-strike homogeneity
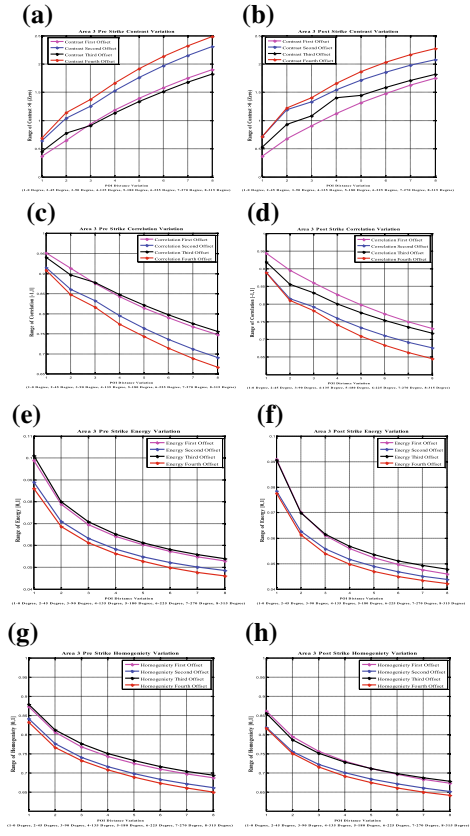
3.1 Comparative Plots for Pre- and Post-strike Texture Feature Variations for Area 3

Now, we have plotted the variations that have occurred in the pre- and post-strike images of Area 3 (Fig. 5).

Fig. 6 Area 4 **a** Pre-strike
contrast **b** Post-strike
contrast **c** Pre-strike
correlation **d** Post-strike
correlation **e** Pre-strike
energy **f** Post-strike energy
g Pre-strike homogeneity
h Post-strike homogeneity
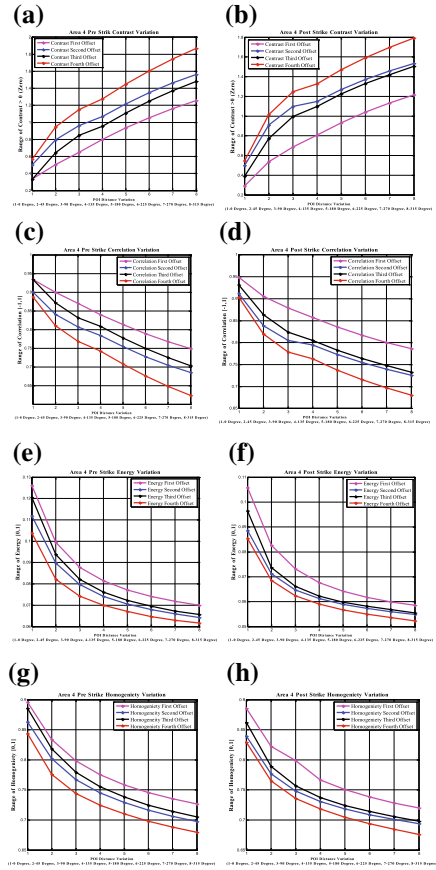
3.2 Comparative Plots for Pre- and Post-strike Texture Feature Variations for Area 4

Now, we have plotted the variations that have occurred in the land cover features of the pre- and post-strike images of Area 4 (Fig. 6).

3.3 Image Enhancement Using Decorrelation Stretcher (DS)

Here, we have applied the DS on the pre- and post-strike images of Area 3 and Area 4 to obtain a better visualization of the hidden terrorist camps. The scatter plot suggests the variation of color before and after applying DS (Figs. 7 and 8).
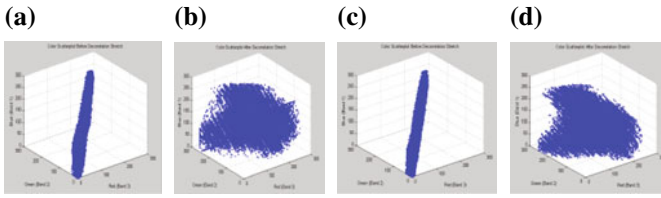
Fig. 7 Area 3 **a** RGB scatter plot of pre-strike image **b** RGB scatter plot of pre-strike image after applying Decorrelation stretcher **c** RGB scatter plot of post-strike image **d** RGB scatter plot of post-strike image after applying Decorrelation stretcher
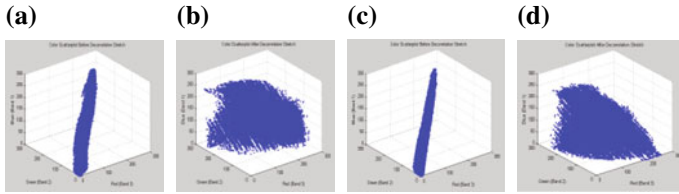
Fig. 8 Area 4 **a** RGB scatter plot of pre-strike image **b** RGB scatter plot of pre-strike image after applying decorrelation stretcher **c** RGB scatter plot of post-strike image **d** RGB scatter plot of post-strike image after applying decorrelation stretcher
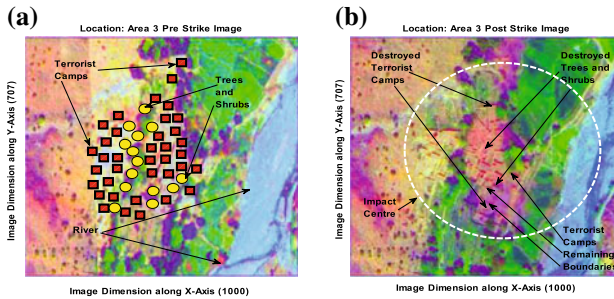
Fig. 9 **a** Pre-strike DS image of Area 3 **b** Post-strike DS image of Area 3

3.4 Change Maps Through Decorrelation Stretcher

In the Area 3, numbers of the terrorist camps are much larger than the Area 1 and 2, the trees and shrubs are lying in between these areas. This area is the densest area in term of urbanization as Fig. 9 a that suggests large numbers of urban structures are lying close to each other making this area as the densest terrorist camp compared to Area 1 and Area 2.

In the final Area 4, there are the more numbers of terrorist camps, compared to Area 2 but less than Area 3. The terrorist's establishment is mostly near the riverside. Here, around 17–18 suspicious structures were present. The numbers of the trees around the terrorist camps are also less in number as compared to the previous areas.

Fig. 10 **a** Pre-strike DS image of Area 4 **b** Post-strike DS image of Area 4
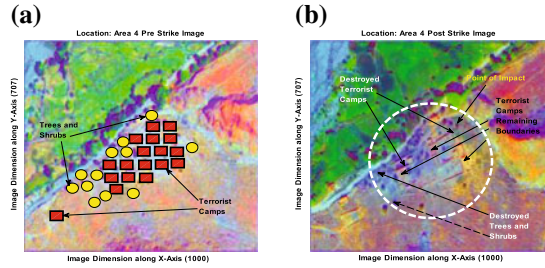
Table 6 Change pattern in texture features before and after strike

S. no	Texture parameters	Pre- strike	Post-strike	Change category
1.	Contrast	Low	High	Positive
2.	Correlation	High	Low	Negative
3.	Energy	High	Low	Negative
4.	Homogeneity	High	Low	Negative

This area is situated along the riverside, which can also be a way to exit during any sought of attack (Fig. 10).

In the post-strike images, it is clearly visible that the roofs of the structure are destroyed, but the walls remain in their positions. This area has got the least damage compared to other areas.

4 Conclusion

Here we have concluded that the structural damage to the suspicious structures has occurred, but not up to a greater extent. This is also concluded that among all the areas, the area which is having the least impact, influence is Area 4 and the area which is most affected by the explosion is Area 1. It is also concluded that the contrast was initially less and after the strike, it becomes high, whereas the feature correlation, energy, and homogeneity were initially high and later on they show declination. When we are comparing pre-strike and post-strike areas as a single event that changes have occurred in the features as a positive change and negative change. Contrast has developed a positive change, while correlation, energy, and homogeneity have developed negative change (Table 6).

Finally, the pattern that got developed in the texture features of the land cover after applying GLCM can be concluded by the overlay plot shown in Fig. 11.

This approach of texture classification through GLCM parameters provides us a useful quantification of the TF. Many other cases related to the terrorist activities can also be monitored and investigated through this approach.

Fig. 11 Texture features changing pattern
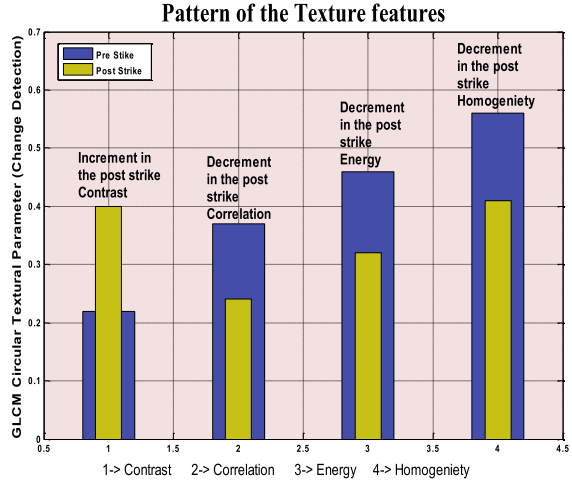
Pattern of the Texture features

References

1. Belward A, Skoien J (2015) Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. ISPRS J Photogramm Remote Sens 103:115–128
2. Hussain M, Chen D, Cheng A, Wei H, Stanley D (2013) Change detection from remotely sensed images: from pixel-based to object-based approaches. ISPRS J Photogramm Remote Sens 80:91–106
3. Coppin P, Bauer M (1996) Digital Change detection in forest ecosystems with remote sensing imagery. Remote Sens Rev 13:207–234
4. Howarth PJ, Wickware GM (2007) Procedures for change detection using Landsat digital data. Int J Remote Sens 25
5. Ye S, Dongmei C, Jie Y (2016) A targeted change-detection procedure by combining change vector analysis and post-classification approach. ISPRS J Photogramm Remote Sens 114:115–124
6. Ma B, Pu R, Wu L, Zhang S Vegetation index differencing for estimating foliar dust in an ultra-low-grade magnetite mining area using landsat imagery. Accessed 5 May 2017
7. Li C, Joinner J, Krotkov N, Bhartia P (2013) A fast and sensitive new satellite SO_2 retrieval algorithm based on principal component analysis: application to the ozone monitoring instrument. Geo-Phys Res Lett 40:6314–6318
8. Cracknell M, Reading A (2014) Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput Geosci 63:22–33
9. Hart S, Veblen T (2015) Detection of spruce beetle-induced tree mortality using high and medium resolution remotely sensed imagery. Remote Sens Environ 168:134–145

10. Brown G, Kytta M (2014) Key issues and research priorities for public participation GIS (PPGIS): a synthesis based on empirical research. Appl Geogr 46:122–136
11. Zhang Y, Zhang H, Lin H (2014) Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sens Environ 141:155–167
12. Zhang L, Lefei Zhang B (2016) Deep learning for remote sensing data. Geosci Remote Sens Magazine 22–40

A Study of Digital Communication over Generalized K-Fading Channels

Nibedita Das, Laishram Mona Devi, Suparna Goswami and Aheibam Dinamani Singh

Abstract The Generalized K (K_G) fading channels is currently taken as generic and versatile model due to its accurate illustration of both large-scale fading (represented by means of Gamma distribution) and small-scale fading (represented by means of Nakagami-m distribution). For this model, a detailed performance of digital modulations over K_G fading channels is investigated. The exact closed form expressions of outage probability and average symbol error rate over K_G fading have been evaluated here. These expressions are calculated using probability distribution function based method. The consequences of the different modulation methods and the fading parameter k and m on the system performance have been presented. The numerical results found are verified by Monte Carlo simulation.

Keywords ASER · Digital modulations · K_G fading channels · Outage probability · PDF

1 Introduction

The propagation of radio signals via wireless medium is a complex phenomenon influenced by effects like multipath fading, shadowing, path loss, etc. Depending upon the radio signal used, there are various types of models illustrating the analytical nature of multipath fading [1]. The focus of wireless communication receiver

N. Das (✉) · L. M. Devi · S. Goswami · A. D. Singh
Department of ECE, NERIST, Nirjuli, Arunachal Pradesh,
India
e-mail: nibeditadas577@gmail.com

L. M. Devi
e-mail: monalaishram16@gmail.com

S. Goswami
e-mail: suparnanerist@gmail.com

A. D. Singh
e-mail: ads@nerist.ac.in

277

design is on removing the effect of small-scale fading. To mitigate the effect, the main requirement is the accurate mathematical modeling of the communication channels. Generalized K is the composite Nakagami-Gamma distribution which closely approximates Nakagami lognormal distribution [2]. K-fading is the special case of Generalized K, and also it incorporates many of the fading and shadowing effects found in wireless channels [3], which later can be mostly utilized in RADAR applications [4, 5]. In [6, 7], the probability density function (PDF), cumulative density function (CDF), moments generating function (MGF), moments, and average Shannon's channel capacity are derived for K_G fading channels without diversity combining, along with the outage probability and bit error probability for a great variety of modulation formats [3]. In [8, 9], the capacity of K_G fading channel is studied for various adaptive transmission techniques. The performance of multihop links and the performance of maximal ratio combining (MRC) receiver over K_G fading channels have been investigated in [10, 11]. Several works done on the performance and capacity analysis over fading channels are found in literature. From literature survey, it is noticed that for Generalized K-fading, the performance analysis is not available for various coherent and noncoherent modulation methods such as DBPSK, 4-FSK, Rect.QAM, etc. This extends the idea and motivation to work in the field of performance of digital modulation methods over K_G fading channels.

The latter part of the paper is written in the following order. In Sect. 2, the channel and system have been elaborated. In Sect. 3, the performance calculation, i.e., the outage probability and average symbol error rate (ASER) performance over K_G fading channel are carried out. The numerical results are discussed in Sect. 4. Lastly, the conclusion part is done in Sect. 5.

2   Channel and System Model

The channel characteristics are considered as slow and frequency nonselective with K_G fading statistics. The expression of the received signal of the complex low pass equivalent per symbol duration, T_s is shown in [12] as

$$r(t) = re^{j\varphi}s(t) + n(t) \tag{1}$$

where s(t) corresponds to the transmitted symbol having energy E_s and n(t) is the complex Gaussian noise with zero mean and two-sided power spectral density $2N_0$. Random variable φ is the phase and r represents the K_G distributed fading amplitude whose SNR PDF is rewritten as [8]

$$f_\gamma(\gamma) = \frac{\eta^{\beta+1}}{2^\beta \Gamma(m)\Gamma(k)} \gamma^{\frac{\beta-1}{2}} K_\alpha(\eta\sqrt{\gamma}), \tag{2}$$

where fading parameter of K_G fading channels are k and m with $\alpha = k - m$, $\beta = k + m - 1$, K(.) is the second kind modified Bessel function having order α and $\Gamma(.)$ is the Gamma function. In (2), $\eta = \sqrt{\frac{4km}{\overline{\gamma}}}$ in which $\overline{\gamma}$ represents average SNR [8].

3 Performance Calculation of Generalized K-Fading

In this section, the ASER and outage probability over generalized K-fading have been derived.

3.1 ASER

Using the expression of SNR PDF given in (2), the closed form expression of ASER is calculated as follows [12]:

$$P_e(\gamma) = \int_0^\infty P_e(\varepsilon/\gamma) f_\gamma(\gamma) d\gamma \tag{3}$$

where $P_e(\varepsilon/\gamma)$ is the conditional symbol error rate for different modulation techniques. The ASER calculated for coherent and noncoherent system are described below.

3.1.1 Coherent Modulations

For coherent modulation schemes, the conditional symbol error rate can be written as

$$P_e(\varepsilon/\gamma) = a Q\left(\sqrt{b\gamma}\right). \tag{4}$$

The Q function is expressed in [12] as

$$Q\left(\sqrt{b\gamma}\right) = \frac{a}{2\pi} \Gamma\left(\frac{1}{2}, \frac{b\gamma}{2}\right) \tag{5}$$

where a and b are taken from [12] on the basis of modulation method used. Substituting $P_e(\varepsilon/\gamma)$ and $f_\gamma(\gamma)$ in Eq. (3), the expression is rewritten as

$$P_e(\gamma) = \frac{a\eta^{\beta-t+0.5}}{\sqrt{\pi} 2^{\beta+t+0.5}\Gamma(m)\Gamma(k)} \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma(\alpha+t+0.5)}{\Gamma(t+1)\Gamma(\alpha-t+0.5)}$$

$$\times \int_0^\infty x^{(\beta-t+0.5)-1}\ell^{-\eta x}\Gamma\left(\frac{1}{2},\frac{bx^2}{2}\right)dx \tag{6}$$

The integral of Eq. (6) can be solved using [14] and final solution after simplification is obtained as

$$P_e(\gamma) = \left(\frac{b}{2}\right)^{\frac{-\beta+t-0.5}{2}}\left(\frac{\Gamma\left[\frac{\beta-t+2.5}{2}\right]{}_PF_Q\left[\left\{\frac{\beta-t+0.5}{2},\frac{\beta-t+1.5}{2}\right\},\left\{\frac{1}{2},\frac{\beta-t+2.5}{2}\right\},\frac{\eta^2}{2b}\right]}{\beta-t+0.5}\right.$$

$$\left.-\frac{\eta\Gamma\left[\frac{\beta-t+2.5}{2}\right]{}_PF_Q\left[\left\{\frac{\beta-t+1.5}{2},\frac{\beta-t+2.5}{2}\right\},\left\{\frac{3}{2},\frac{\beta-t+3.5}{2}\right\},\frac{\eta^2}{2b}\right]}{(\beta-t+1.5)\sqrt{\frac{b}{2}}}\right) \tag{7}$$

where $_PF_Q$ is a Hypergeometric function.

3.1.2 Noncoherent Modulation

The conditional symbol error rate for noncoherent modulation schemes is written as

$$P_e(\varepsilon/\gamma) = a\,\exp(-b\gamma) \tag{8}$$

where a and b are taken as given in [12]. Substituting $P_e(\varepsilon/\gamma)$ and $f_\gamma(\gamma)$ into (3), it is rewritten as

$$P_e(\gamma) = \frac{a\sqrt{\pi}}{\Gamma(m)\Gamma(k)}\sum_{t=0}^{\alpha-\frac{1}{2}}\sum_{p=0}^{\infty}\frac{(-1)^p\eta^{\beta+p-t+0.5}}{p!\,2^{\beta+t+0.5}}\frac{\Gamma(\alpha+t+0.5)}{\Gamma(t+1)\Gamma(\alpha-t+0.5)}$$

$$\times\int_0^\infty \gamma^{\left(\frac{2\beta+2p-2t+1}{4}\right)-1}\ell^{-bx}d\gamma \tag{9}$$

The integral is solved using [13, (3.381.4)], the final expression is found as below

$$P_e(\gamma) = \frac{a\sqrt{\pi}}{\Gamma(m)\Gamma(k)}\sum_{t=0}^{\alpha-\frac{1}{2}}\sum_{p=0}^{\infty}\frac{(-1)^p\eta^{\beta+p-t+0.5}}{p!\,2^{\beta+t+0.5}}\frac{\Gamma(\alpha+t+0.5)}{\Gamma(t+1)\Gamma(\alpha-t+0.5)}$$

$$\times b^{\frac{-2\beta-2p+2t-1}{4}}\Gamma\left(\frac{2\beta+2p-2t+1}{4}\right) \tag{10}$$

3.2 Outage Probability

A significant indicator of performance for wireless systems over various fading channels is outage probability. The P_{out} is the probability that the instantaneous received SNR falls below the required threshold, γ_{th} [1], which is given below

$$P_{out} = \int_{0}^{\gamma_{th}} f_\gamma(\gamma) d\gamma \tag{11}$$

The outage probability is calculated by using Eq. (11), i.e.,

$$P_{out} = \frac{\sqrt{\pi}}{\Gamma(m)\Gamma(k)} \sum_{t=0}^{\alpha - \frac{1}{2}} \sum_{p=0}^{\infty} \frac{(-1)^p \eta^{\beta+p-t+0.5}}{p! \, 2^{\beta+t+0.5}} \frac{\Gamma(\alpha+t+0.5)}{\Gamma(t+1)\Gamma(\alpha-t+0.5)}$$
$$\times \int_{0}^{\gamma} \gamma^{\left(\frac{2\beta+2p-2t-3}{4}\right)} d\gamma \tag{12}$$

The integration is solved using [13, (3.351.1)] and it is obtained as

$$P_{out} = \frac{\sqrt{\pi}}{\Gamma(m)\Gamma(k)} \sum_{t=0}^{\alpha - \frac{1}{2}} \sum_{p=0}^{\infty} \frac{(-1)^p \eta^{\beta+p-t+0.5}}{p! \, 2^{\beta+t-1.5}(2\beta+2p-2t+1)} \frac{\Gamma(\alpha+t+0.5)}{\Gamma(t+1)\Gamma(\alpha-t+0.5)}$$
$$\times \gamma^{\left(\frac{2\beta+2p-2t+1}{4}\right)} \tag{13}$$

4 Numerical Results and Discussions

The expressions found in Sect. 3 are numerically calculated and the results are plotted for various values of k for the illustration purpose. Figures 1, 2, and 3 shows the plot for ASER of various modulation methods and outage probability. P_{out} vs average SNR ($\overline{\gamma}$) is plotted in Fig. 1. It is seen from the graph that as the value of k increases, P_{out} decreases (i.e., outage probability performance improves). Moreover, the gap among the curves reduces with the increase k value. It is because the increase in the value of k signifies the amount severity of shadowing effect is less hence it represents better channel. Likewise, Figs. 2 and 3 show the plot for average symbol error rate versus $\overline{\gamma}$ for different coherent and noncoherent systems, respectively. The graph is plotted for various k values. From Fig. 2, it is noticed that the performance of BPSK is better than that of BFSK. The reason behind this is that the separation between the message points is more in BPSK, which results in lesser amount of error probability in presence of fading.

Fig. 1 Outage probability of K_G fading channels
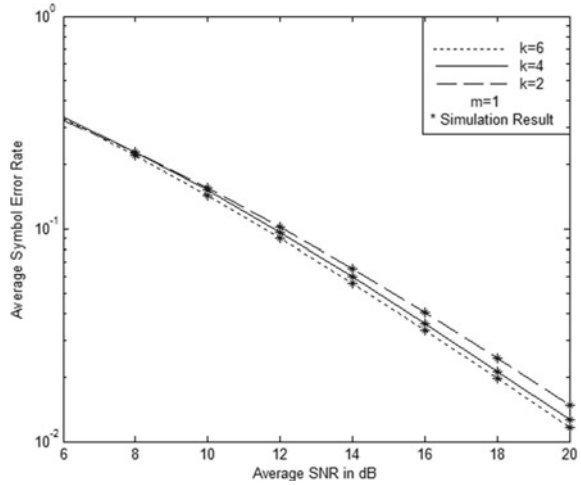
Fig. 2 ASER of coherent method over K_G fading channels
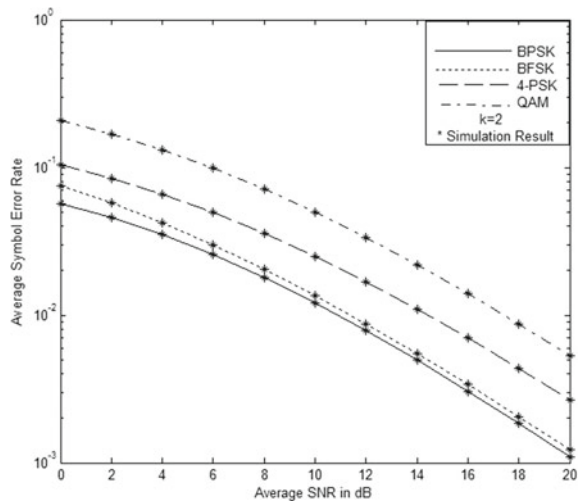
The performance of ASER decays with the increase in order of modulations due to the lesser distances among the message bits. For the higher order modulations at the same time for a short duration of deep fading more numbers of bits will be in error. Moreover, ASER performance of rect. QAM is inferior to that of 4-PSK. This is because in QAM, message is stored in amplitude of the carrier where the signals can be affected by fading and noise.

In Fig. 3, the ASER versus $\bar{\gamma}$ in dB for various noncoherent modulation methods have been plotted. It is clear from the graph that the performance improves with higher values of k. Here, the performance of DBPSK is more efficient than BFSK since in DBPSK the message is stored in phase variation and in BFSK the message is stored in frequency variation. As the noise can easily affect the frequency of the

Fig. 3 ASER of noncoherent method over K_G fading channels
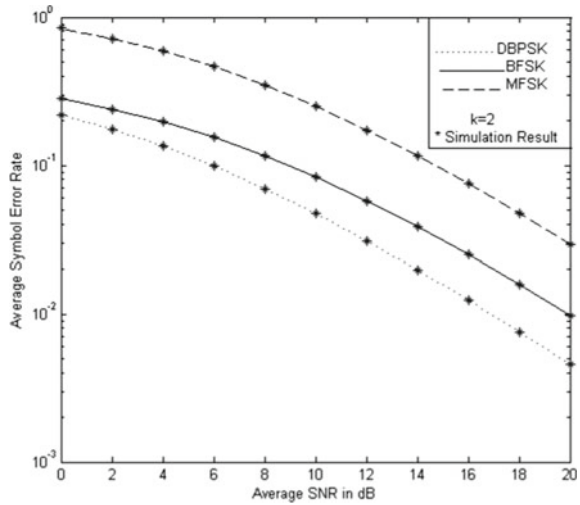
Fig. 4 Comparison between coherent and noncoherent methods over K_G fading channels
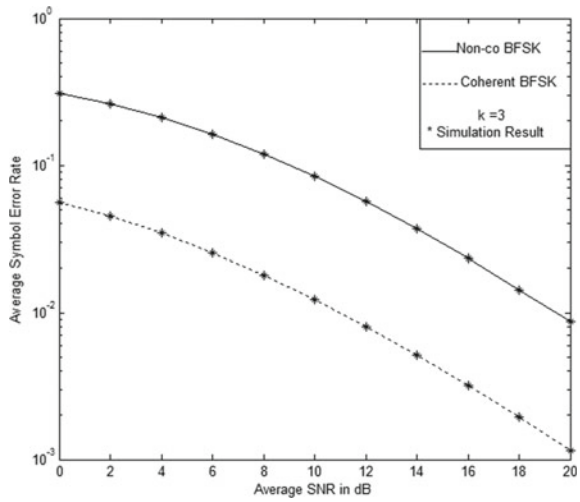
signal, hence DBPSK performs superior to BFSK. But, the performance of 4-PSK is very poor as compared to other modulation schemes.

The performance comparison of coherent BFSK and noncoherent BFSK is given in Fig. 4. It shows that the ASER performance of coherent BFSK is better than noncoherent BFSK. It is on the basis of perfect phase and frequency synchronization of the coherent detection. For higher order modulation process also observation is similar as above.

5 Conclusions

The closed form expressions for outage probability, ASER for coherent and noncoherent modulation system over generalized K-fading have been derived. The expressions are found by using PDF of output SNR of the system. The numerically calculated results are obtained for different values of k using Mathematica and validated with Monte Carlo simulation.

References

1. Marvin K (2005) Simon and Mohammed –Slim Alouini: digital communication over fading channels, a unified approach to performance analysis, 2nd edn. Wiley, New York
2. Aleksic Danijela, Stefanovic Mihajlo, Popovic Zoran, Radenkovic Dragan, Ristic Jovan D (2009) On the K and K_G fading channels. Serbian J Electric Eng 6(1):187–201
3. Subadar R, Das P (2017) Performance of L-SC receiver over K-Fading Channels, 8th ICCCNT
4. Iskander DR, Zoubir AM (1999) Estimation of the parameters of the K-distribution using higher order and functional moments. IEEE Trans Aerosp Electron Syst 35(4):1453–1457
5. Chitroub S, Houacine A, Sansal B (2002) Statistical characterisation and modelling of SAR images. Elsevier Signal Process 82(1):69–92
6. Bithas PS, Sagias NC, Mathiopoulos PT, Kkaragiannidis GK, Rontogiannis AA (2005) Digital communications over generalized-K fading channels In: Proceedings of international workshop on satellite and space communications, pp 684–687
7. Bithas PS, Sagias NC, Mathiopoulos PT, Kkaragiannidis GK, Rontogiannis AA (2006) On the performance analysis of digital communications over generalized-K fading channels. IEEE Commun Lett 10(5):353–355
8. Laourine A, Alouni MS, Affes S, Stephenne A (2008) On the capacity of generalized-K fading channels. IEEE Trans Wireless Commun 7(7)
9. Dwivedi VK, Singh G (2012) A novel MGF based analysis of channel capacity of generalized-K fading with maximal-ratio combining diversity. Prog Electromagnet Res 26:153–165
10. Cao J, Yang LL, Zhong Z (2010) Performance of multihop wireless links over generalized-k fading channels. Proceedings of IEEE vehicle technology conference, Fall
11. Dwivedi VK, Singh G (2014) Moment generating function based performance analysis of maximal-ratio combining diversity receivers in the generalized-K fading channels. Wireless Personal Commun 77(3)
12. Singh AD, Goswami S (2015) Performance of digital modulations over TWDP fading channels. In: International conference of electronic design, computer networks and automated verification (EDCAV), pp 135–138
13. Gradshteyn IS, Ryzhik IM (2000) Table of integrals, series and products, 6th edn. Academic Press, San Diego, CA
14. Wolfram Mathworld. http://mathworld.wolfram.com/incompleteGammaFunction.html

An Improved Genetic Algorithm Based Solution to Vehicle Routing Problem over OpenMP with Load Consideration

Rahul Saxena, Monika Jain, Aashis Kumar, Vidyanshu Jain, Tushar Sadana and Siddharth Jaidka

Abstract Vehicle Routing Problem being a combinatorial class of problems has implications in various areas and applications where traditional methods to find a search space either fail or slow down especially in case of real time systems. Using the most popular heuristic searching approach, Genetic Algorithm, there have been solutions in the state of the art to the problems like Vehicle Routing, Traveling Sales Person, etc. In this paper, a parallelized version of genetic algorithm has been proposed for vehicle routing problem over OpenMP programming model. The problem has been solved taking into consideration the constraint that congested cities of the network have been pushed behind after obtaining a fit set of chromosomes in every iteration so that as the time pass by the load sheds on the cities and we optimize in terms of time as well. The concept can very well be mapped to the network routing problem as well. Experimental results show that the efficiency of the proposed model is high in comparison to the serial version where the machine certainly meets the restrictions in terms of execution and processing power to get the intensive computations done as the number of cities or nodes in the network increases. Finally, the paper sums up with the future endeavors of the problem with more complex constraints getting involved to fetch the best possible route and how parallel processing must deliver such solutions.

Keywords Vehicle routing · Genetic algorithm · Parallel processing · OpenMP

1 Introduction

Traveling Salesperson Problem is a classic NP-hard combinatorial optimization problem to find shortest route for an N-vertices graph where each vertex gets visited exactly once. With one of the most interesting problems in the field of graph theory to be solved, TSP has find its implication in many real-world problems like crew

R. Saxena (✉) · M. Jain (✉) · A. Kumar · V. Jain · T. Sadana · S. Jaidka
Manipal University Jaipur, Jaipur, India
e-mail: rahul.saxena@jaipur.manipal.edu

scheduling, vehicle routing, connection decision problem for a PCB, and many more. As a result, the problem has attracted many researchers and scholars to investigate upon it and find its applicability to solve a real-world problem in a most comprehensive way and manner. The problem has a number of problem solving approaches. The document focusses on one of the solving strategy to VRP, by formulating the problem in terms of another graph-based problem called Traveling Salesperson Problem (TSP) during processing. The VRP problem first appeared in a paper by George Danzig and John Ramser in 1959. Vehicle Routing Problem determines the optimal set of routes for a fleet of vehicles to traverse in order to deliver customer demands where each customer demand is fulfilled exactly once and the vehicles starts and ends at the same station/depot. The main components of VRP are depots, customers, vehicles, and routes. In this paper, we have considered the solution for a single vehicle however the same approach can be extended for a fleet of vehicles. To calculate the optimal solution in vehicle routing problem is a NP-hard problem and is a generalization of traveling salesman problem. Figure 1 shows the example of graph of vehicle routing problem. It depicts the optimized routes represented with a complete graph G (V, E).

Decision will be based on value of $x_{i,j}$, i.e., if vehicle x travels from node i to node j, $x_{i,j} = 1$ else $x_{i,j} = 0$. In simple words, the problem can be viewed as covering all vertices of a graph starting from a source vertex exactly once with minimum cost of traversing the nodes. There have been solutions in the state of the art with respect to the problem in the past but genetic algorithm based solutions to find a solution search space have gained popularity due to combinatorial nature of the problem which increases the number of routes to be tested by the factor of the factorial value of the number of nodes. In such cases, exact solution cannot be achieved as the machine reaches its computational limitations. GA-based solutions making a trade-off with the solution accuracy to a legible extent achieve the desired task. In fact, for real-time applications, the approach performs quite well. In this paper, we have worked out the problem using the same concept. A parallel version of GA algorithm has been utilized to further enhance the efficiency of the proposed solution in terms

Fig. 1 Graphical representation of a vehicle routing scenario
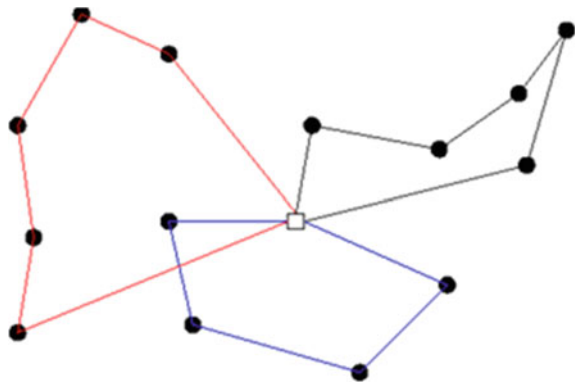
of execution time. But it is interesting to note that always finding the optimum path or route based on the cost-distance metric is not sufficient. If any route has high congestion or high traffic, no matter the path is optimum, the busyness of the route may hamper the performance. So with an added constraint, the proposed solution identifies the congested nodes of the network, and in the final solution pushes the congested city two to three nodes back in order to optimize for time as pushing gives the time to the node to shed off the load or the traffic and thus the request can be processed quick.

The upcoming sections of the paper guide the flow to approach the problem. Section 1 presents a general introduction to the problem. Section 2 discusses the genetic algorithm and related approaches to vehicle routing problem using genetic algorithm. Section 3 discusses the genetic algorithm model for VRP. Section 4 presents the parallel programming approach to genetic algorithm to find the most optimum route. Further, the congested nodes are found by making a parallel search over the network and then pushing the congested nodes back in the final routes selected. Section 5 presents the experimental results where the serial approach is compared up with the parallel approach on the basis of execution time and producing solutions. Finally, paper concludes with some future possibilities to extend the solution for efficient routing and data forwarding mechanism in networks.

2 Related Work

Before discussing the state of the art approaches for Vehicle Routing Problem using Genetic Algorithm, a brief overview of what genetic algorithm is and how it actually works as a heuristic is discussed.

2.1 Genetic Algorithm: An Overview

Genetic algorithm is a natural selection mechanism introduced by Darwin. It is a process of finding the solution that is close to optimal solution. The process consists of initially selecting 10–15% chromosomes, then choosing appropriate chromosomes by using fitness function, then performing crossover and mutations between the remaining chromosomes and repeating this process for successive iterations. A survey on genetic algorithm has been presented by Srinivas and Patnaik [1]. GA is an iterative strategy. A common scope of eras for a basic GA can extend from 50 to more than 500. The whole arrangement of eras is known as a **run**. Utilizing GA as a seeking heuristic, the process will be started by an arrangement of **populace or population**.

These populaces will be referred to as **chromosome** development of a grouping of bit or numeric string where every individual from the arrangement will be referred to as qualities. Figure 2 shows a step by step procedure to a GA process. The termination
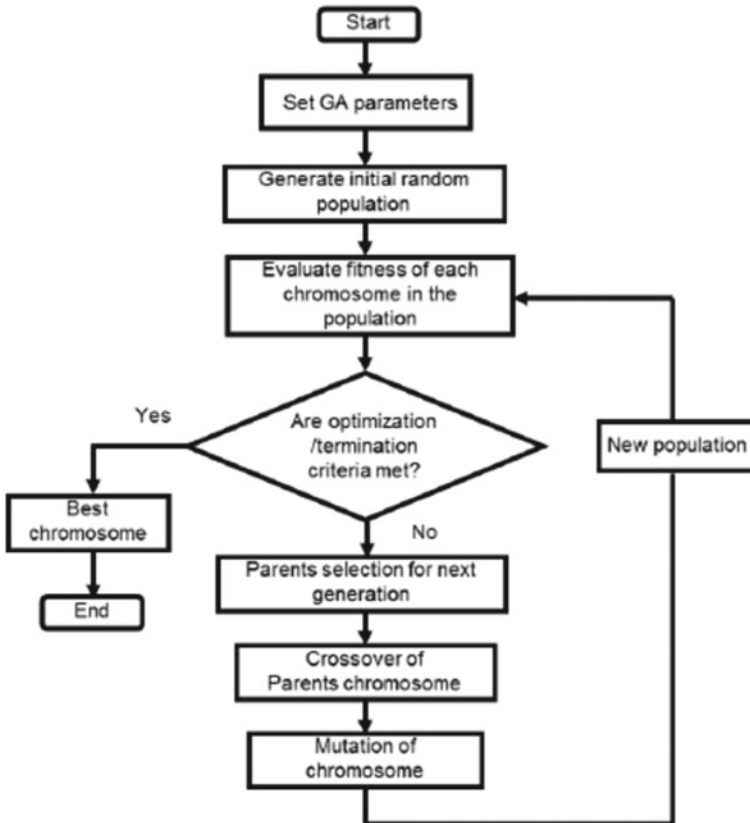
Fig. 2 Genetic algorithmic procedure

criteria are set as predefined number of iterations or when the generation stabilizes for successive iterations. The population generation, crossover, and mutation operators and techniques vary as per the requirement. Fitness function along with the operators and techniques will be discussed later in Sect. 3 of the paper.

2.2 State-of-the-Art Approaches to VRP

A lot of investigative work has been done to find the solution to Vehicle Routing Problem. Both VRP and TSP (Traveling Salesperson Problem) involves a Hamiltonian Cycle of least weightage inside a fully connected undirected graph G = (V, E). The possible number of paths in the naive algorithm is n! and as the n increases, it becomes increasingly difficult to find the optimal solution in terms of time and space complexity. The researchers keeping the view in mind approached the problem in an

optimized manner in order to achieve the solution which cannot be achieved using brute force and classical approaches like integer programming, branch, and bound technique. Lin et al. [2] discuss "Research on Optimization of Vehicle Routing Problem for Ride sharing Taxi" suggesting that shared ride can save resources and prevent traffic congestion. Nazif and Lee [3] "Optimized crossover genetic algorithm for capacitated vehicle routing problem". In this paper, "Capacitated" VRP was solved by applying optimized crossover in genetic algorithm. Masum et al. [4] "Solving the Vehicle Routing Problem using Genetic Algorithm". In this paper, the authors solved VRP using Genetic Algorithms by tuning the system to find a balance between a fast solution and limiting the search space. Chand and Mohanty [5], "A Multi-Objective Vehicle Routing Problem using Dominant Rank Method". In this paper, a multi-objective Vehicle Routing Problem using Dominant Rank Method was proposed to find optimum solution. Cooray [6] discusses the Machine Learning-Based Parameter Tuned Genetic Algorithm for Energy Minimizing Vehicle Routing Problem. Nazif et al. [7], "Optimized Crossover Genetic Algorithm for Vehicle Routing Problem with Time Windows". This study presented a genetic algorithm with optimized crossover operator having time windows to find an optimized route. Enrique Alba et. al. [8], "Solving the Vehicle Routing Problem by Using Cellular Genetic Algorithms" discuss the VRP problem with genetic algorithm. Ákos Kovács "Solving the Vehicle Routing Problem with Genetic Algorithm and Simulated Annealing" discusses the VRP algorithm along with Genetic algorithm and simulated annealing. As we can see that genetic algorithm is used for a while to get more accurate results in vehicle routing problem.

The state of the art suggests that VRP solution using genetic algorithm approach has been the best of the approaches to the problem as the other heuristic methods also produce a computational complexity of $O(n^3)$ or $O(n^2 \log n)$ which is quite high for large n values. But still targeting the problem with GA heuristic approach has boundations. If the initial population is taken low, the accuracy with the solution has to be compromised and picking up larger number of routes may still hang up the system. Keeping this in view, there have been proposed parallelized solutions for TSP and VRP [9, 10, 11]. The solutions discussed in this paper accelerate the solution achievement to great extent. Carrying this idea further, the paper here proposes a parallel version of the solution to the problem over OpenMP programming model with an added constraint of dense nodes in the network which still hampers the efficiency if traversing through the shortest path. How this can be achieved is discussed in the following sections of the paper.

3   Genetic Programming Approach to VRP

As discussed in Sect. 2, the basic model of VRP can be modeled as per the following steps discussed below. Taking a reference figure as Fig. 3 where a connected graph of 5 cities has been shown. The steps of the algorithm have been explained in consideration to this example.

Fig. 3 A 5-connected graph
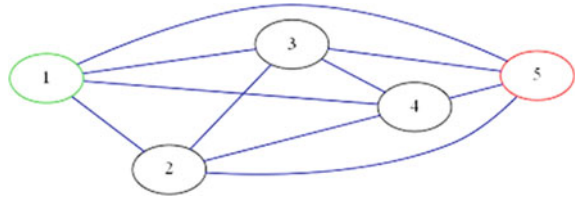
Initialization: Initial population is a set of chromosomes where each chromosome represents a route. In the first step, the initial population is randomly generated. The routes can be simply picked up by using Permutation encoding. The routes can be taken as

$$1-2-3-4-5$$
$$1-2-4-3-5$$
$$1-3-2-4-5$$
$$1-4-2-3-5 \text{and so on.}$$

Selection based on fitness function: Chromosomes whose path length is less than the value of fitness function are retained in the population and the rest are discarded. Fitness Function refers to the function used to filter the children in the new born population which will be selected to breed, the fitness function is applied to every generation. Mathematically, it can be defined as

$$\text{Minimize } \Sigma i \ \Sigma j \ dij \ xij$$
$$\text{Subject to: } \Sigma \ xij = 1, i = 1 \ldots N \text{ and } j = 1 \ldots N$$

For N vertex, the above conditions defined the fitness function. The criteria for minimum can be chosen by assigning ranks to the route or by putting some threshold. The routes below the threshold value are selected as fit population. The threshold for this is calculated as (max + min)/2

max indicates the maximum path length available in the current population and **min** indicates the minimum path length available in the current population.

Crossover: Next step is to produce the offsprings by mating the fit population obtained. A number of crossover operators are available to perform the task in [15]. From the point of view of mapping the problem to the parallel architecture, Ordinal Crossover technique has been taken.

Mutation: During the mutation process, dense nodes for the network will be evaluated. This evaluation is done before the genetic procedure for evaluating fit population only once. Now, in a route, the nodes with heavy traffic on it will be pushed back so as to trade off for the time in the chromosomes obtained from crossover stage in the routes where the nodes with high-density values appear in series or very early in the route.

These set of operations are repeated for a predefined number of epochs or till the time solution converges.

4 Parallelized Version of the Proposed Method Over OpenMP

4.1 *OpenMP Programming Model*

To improve computational time and to solve VRP, we used functions of OpenMP API for parallelization as our algorithm has scope of parallelization. OPENMP is a popular and easiest platform to write a parallel code to run in multiple core architecture. OPENMP consists of basically three main components namely Directives, Environmental variables and Clauses. Main concept behind the parallelism is multithreading. Multiple threads are created in a program and work is distributed among the threads using directives, clauses, and environmental variables. OPENMP program is a combination of serial and parallel code. The programmer has to explicitly specify which code to run in parallel using compiler directives [12]. OPENMP supports C and Fortran language. The application here utilizes C language for parallel implementation of algorithm and compared the performance of among various number of cores and threads. Figure 4 shows the basic architecture of OpenMP.

OpenMP has been an indigenous API platform for many applications [13, 14, 15] where the results have been achieved to support applications in real time. The parallel processing power of multicore based architecture machines has certainly enabled the solutions to some of the most important problems [17] which require fast processing over a huge data set with complex mathematical calculations.

Fig. 4 Parallel tasks in OPENMP
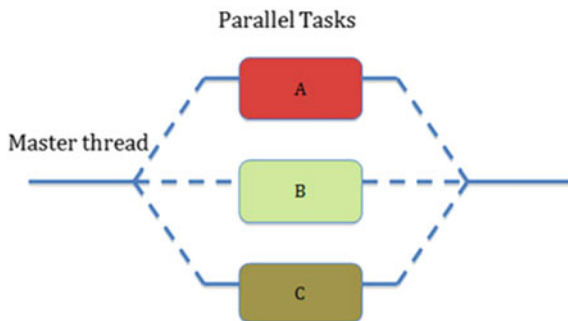
4.2   Algorithmic Modeling of the Problem Over OpenMP

The problem here basically has few independent portions over which parallelization can be achieved. The first process is the evaluation of fit population based on the fitness function. So if we suppose that for 10 cities out of 10! paths, 10,000 routes are selected using *Permutation encoding* scheme, the fitness procedure can be spawn over the threads to run in parallel as:

Procedure 1: Minimum Route

```
Begin
#pragma omp parallel for omp_num_threads (n)
        /* running the code over n threads*/
                for initial 10000 routes for 10 cities
                {
                        Min=Initial route
                        If (Min < Current route)
                        Min=Current Route
                }
End
```

A function used to filter the children in the newly created population which will be favored to breed, the fitness function is applied to every generation given as: (max + min)/2 where *Max* indicates the maximum path length available in the current population and *Min* indicates the minimum path length available in the current population. Procedure 1 evaluates the route with minimum cost. A similar procedure can be used to find the route with maximum cost. Using these two values, the populace can be selected by following the fitness function $= (Max + Min)/2$

Procedure 2: Population Selection

```
Begin
        #pragma omp parallel for omp_num_threads (10)
                /* running the code over ten threads*/
                        for initial 10000 routes for 10 cities
                        {
                                If (Current route Cost < (Min+Max)/2)
                                        Store Current route in Possible route vector.
                        }
End
```

Once the population set, i.e., fit chromosomes being identified, ordinal crossover operation is used to generate offsprings [15]. The OpenMP functions over here uses Intel Thread Building Blocks library along with C++. The pair of routes is selected by the threads and crossover operation is done parallely by all the threads.

Procedure 3: Crossover

Begin
        #pragma omp parallel for omp_num_threads (n)
                / running the code over n threads*/*
                        for selected chromosomes
                        {
                                for each pair of chromosome do crossover
                                store in vector child.
                        }
End

For mutation process, dense nodes of the network need to be identified. This step is done offline before the start of genetic procedure. The threshold criteria over here are taken as the number of connections over a node must be less than half of the number of nodes.

Procedure 4: Dense node identification

Begin
        #pragma omp parallel for omp_num_threads (n)
                / running the code over n threads*/*
                        for each node
                        {
                                If node has degree >= half of the number of nodes
                                    Mark as dense node.
                        }
End

So, if one of the best chromosome obtained after the crossover is 1-2-3-4-5 for a 5 city problem and node '2' evaluates out to be dense then the route traversal can be altered as 1-3-2-4-5 or 1-3-4-2-5 where the next connected nearest node to 1 can be obtained by brute force technique.

These set of steps are done repeatedly in each iteration till a termination criterion is met or after a considerable no. of iterations. Thus, parallelization optimizes in each epoch and a considerable amount of execution time is cut down in comparison to the serial execution.

5 Experimental Results

Execution of vehicle routing problem using genetic algorithm over single core and multicore architecture has been depicted in this section. Time taken by single-core architecture is higher as comparison to multicore architecture. We have reduced the computation time by executing the problem parallely over multiple cores and threads. Open MP 4.1 runtime environment is used with quad core specification. We have varied number of threads according to number of cities. We have taken different number of cities from 5 to 50 for VRP using GA over single-core and multicore (four core). Results obtained are given in Table 1.

Obtained result shows that time taken by VRP using multicore architecture is nearly to half in comparison to single-core architecture. Approximately, 50% speedup

Table 1 Time taken by VRP using genetic algorithm over single-core and multicore architecture

S. no.	Number of cities	Execution time over single core (s)	Execution time over quad core (s)
1	10	17.82	7.245
2	50	34.567	16.914
3	60	30.147	18.965
4	500	10000.56	700.89
5	1000	30000.12	15000.45

Fig. 5 Ordinal crossover for the routes
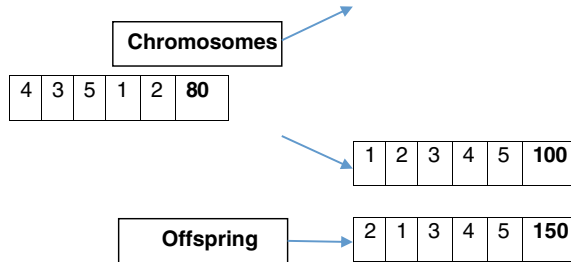
Fig. 6 Execution time of VRP using GA over single-core and multicore processors
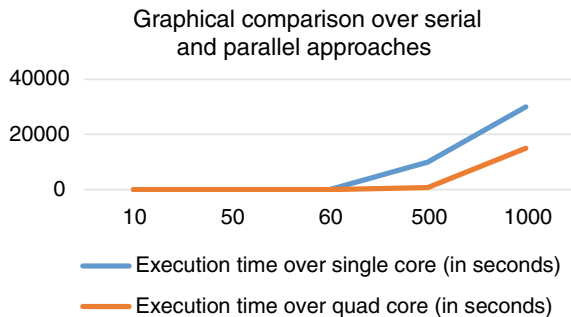
in time can be obtained by using quad cores. However, exact speed up time will depend upon the number of iterations and optimal solution. Figure 5 shows the graph of execution time obtained over single-core and multicore architecture with different number of cities. **X-axis represents the various number of cities and Y-axis shows execution time in seconds** (Fig. 6).

The above results have been obtained up to thousand cities. With brute force approach to the solution, the machine hangs up and is enable to produce the results after 10 cities. Executing genetic algorithm approach over single core do produce results but needs to relax the constraints regarding number of crossovers and thresholds and, thus, compromising with the results accuracy. Executing the process by mapping the program model to parallel architecture enables to work up with strict constraints thus not compromising with the efficiency of the solution.

Table 2 Time taken for evaluating dense nodes of the network over single-core and multicore architecture

S. no.	Number of cities	Execution time over single core (s)	Execution time over quad core (s)
1	50	0.98	0.70
2	500	2.04	1.65
3	1000	15.404	10.55

Also calculating the most dense nodes of the network, as discussed in previous section, parallely saves a significant amount of time as depicted in following table. The results over here have been reported by running the process for 10 times and evaluating the average of the execution time values (Table 2).

6 Conclusion and Future Scope

The paper discusses the application of traveling salesman problem, i.e., vehicle routing problem. Special scenarios of vehicle routing problem are also discussed in the paper. Approach of vehicle routing problem using genetic algorithm heuristic has been the focus of this research article. To identify the most optimal route for a vehicle among the combination, traditional searching methods stuck up with increase in the problem size. Thus, heuristic algorithms like GA are required to account for the computational limitations of machine. However, as the complexity of the problem and problem size further increases, even the heuristic algorithmic computations become intensive and relaxing the process may not yield good solutions and we have to compromise the accuracy. With the advancement in the processor architecture, use of parallel processing can greatly enhance the efficiency of the algorithms which slows down or halt due to intensive computations involved in the process for a large data. The paper demonstrates a comparative analysis of vehicle routing problem using single core and using multicore and reports a major speedup in execution timing. However, the computation time is highly dependent on the initial population and population generated. But performance in executing VRP using Open MP (multicore architecture) gives better results (nearly halftime) in most of the cases. Further, the paper focusses on the fact that optimization is not only the case of optimizing on the basis of cost metric or distance metric. The load on a node too is crucial for optimization. So, the final solutions have been mutated based on the dense nodes of the network by pushing them behind in the traversal path so that the load lowers on the node due to delay introduced and the waiting time can be cut off to process the request.

This model can be worked out for various routing problems of network environment [16]. The packet forwarding over the most efficient path can be done by evaluating the shortest path. The densest nodes can be simulated with link capacity across the routers and few alternative best routes can be cached to the router to pass

the data in case of delay in usual route. The concept can very well be mapped for DOS attacks in networks to identify the malicious nodes as well. On the algorithmic point of view of efficiency improvement, tests can be conducted by varying the crossover methods and threshold functions and criteria. Further executing the idea over high-end processing machines, cluster and GPU cards over a large number of cities will yield a high speedup.

References

1. Srinivas M, Patnaik LM (1994) Genetic algorithms: a survey. Computer 27.6:17–26
2. Lin Y, Li W, Qiu F, Xu H (2012) Research on optimization of vehicle routing problem for ride-sharing taxi. Procedia-Soc Behav Sci 43:494–502
3. Nazif H, Lee LS (2012) Optimized crossover genetic algorithm for capacitated vehicle routing problem. Appl Math Model 36(5):2110–2117
4. Masum AKM, Shah Jalal M, Faruque F, Sarker IH (2011) Solving the vehicle routing problem using genetic algorithm. Int J Advanc Comput Sci Appl 2(7):126–131
5. Chand P, Mohanty JR (2013) A multi-objective vehicle routing problem using dominant rank method. Int J Comput Appl 29–34
6. Cooray PLNU, Rupasinghe TD (2017) Machine learning-based parameter tuned genetic algorithm for energy minimizing vehicle routing problem. J Indust Eng
7. Nazif H, Lee LS (2010) Optimized crossover genetic algorithm for vehicle routing problem with time windows. Am J Appl Sci 7(1):95
8. Alba E, Dorronsoro B (2004) Solving the vehicle routing problem by using cellular genetic algorithms. In: EvoCOP vol 3004, pp 11–20
9. Saxena R, Jain M, Bhadri S, Khemka S (2017) Parallelizing GA based heuristic approach for TSP over CUDA and OPENMP. In: Advances in computing, communications and informatics (ICACCI), pp 1934–1940. IEEE
10. Roberge V, Tarbouchi M, Labonté G (2018) Fast genetic algorithm path planner for fixed-wing military UAV using GPU. IEEE Trans Aerospace Electron Syst
11. Rey A, Prieto M, Gómez JI, Tenllado C, Hidalgo JI (2018) A CPU-GPU parallel ant colony optimization solver for the vehicle routing problem. In: International conference on the applications of evolutionary computation. Springer, Cham, pp 653–667
12. Mustafa B, Ahmed W Parallel algorithm performance analysis using OpenMP for multicore machines. Int J Advanc Comput Technol (IJACT), ISSN 2319–7900
13. Saxena R, Jain M, Singh D, Kushwah A (2017) An enhanced parallel version of RSA public key crypto based algorithm using OpenMP. In: 10th international conference on security of information and networks (SIN). ACM, pp 37–42
14. Jain M, Saxena R Parallelization of video summarization over multi-core processors. Int J Pure Appl Math 118(9):571–584. ISSN 1311-8080
15. Jean-Yves et al (1996) Genetic algorithms for the traveling salesman problem. Ann Operat Res 63:339–370
16. Jain M, Saxena R (2017) Overview of VANET: requirements and its routing protocols. In: 2017 International conference on communication and signal processing (ICCSP). IEEE, pp 1957–1961
17. Saxena R, Jain M, Yaqub SM (2018) Sudoku Game Solving Approach Through Parallel Processing. In: Proceedings of the Second International Conference on Computational Intelligence and Informatics. Springer, Singapore, pp 447–455

Capacity Analysis in Generalized *K*-Fading Channels

**Laishram Mona Devi, Nibedita Das, Suparna Goswami
and Aheibam Dinamani Singh**

Abstract In wireless communication, fading results the signal quality to reduce. The generalized *K*-fading model (K_G) delineates accurately composite multipath/shadowing fading channels which is more applicable model for real-world environment. The study of channel capacity of K_G fading channels is presented without any limitation in the parameters of K_G fading channels in this paper. The closed form expressions for numerous adaptive transmission methods are obtained by using Probability Density Function base approach.

Keywords Adaptive transmission · Channel capacity · Fading channel ·
Generalized *K*-fading channel · Probability density function

1 Introduction

In real-world environment, wireless channels are mostly represented by composite multipath fading and the phenomenon of shadowing [1]. In such situation, generalized *K*-fading model (K_G) is suitable, as it incorporates the fading and shadowing effect occurred in wireless communication channels. The K_G fading model is a special case of *K*-distribution and it also indicates other fading models like Nakagami-m and Rayleigh-Lognormal [2] and it is mainly applicable in RADAR applications [3]. In [3], the expressions of the SNR, the average capacity, and the BER for various modulation schemes for K_G fading channels had been derived. In [4], the perfor-

L. M. Devi (✉) · N. Das · S. Goswami · A. D. Singh
Department of ECE, NERIST, Nirjuli, Arunachal Pradesh, India
e-mail: monalaishram16@gmail.com

N. Das
e-mail: nibeditadas577@gmail.com

S. Goswami
e-mail: suparnanerist@gmail.com

A. D. Singh
e-mail: ads@nerist.ac.in

297

mance of various diversity receivers over K_G fading channels have been analyzed and also investigated the expressions of Probability Density Function (pdf), Cumulative Distribution Function (CDF), Moment Generating Function (MGF) and moments output signal to noise ratio (SNR). In [5], performance had been studied using various measuring parameters of cascaded K_G fading channels. In [6], the multiple-input multiple-output (MIMO) systems capacity over K_G fading channels has been studied. In [7], the performance of multi-hop cooperative relay networks operating over K_G fading channel had been presented. In [8], the performance study of selection combining receiver over K_G fading channels had been presented. In [9], the capacity of Rayleigh fading channels had been studied under numerous adaptive transmission methods and various configurations. In [10–13], capacity of different combining receiver with numerous adaptive transmission methods over different fading channels such as TWDP and Rician had been studied and presented. The analysis of channel capacity under numerous adaptive transmission methods for K_G fading channels have presented in [14], but the expression is valid for some particular value of k, fixed value of m and the results were presented in terms of Lommel functions. In [15], the performance and capacity expressions under numerous adaptive transmission methods for K_G fading are presented in terms of hypergeometric functions with the condition that k-m is not an integer.

In this present paper, the capacity analysis of K_G fading channels under numerous adaptive transmission methods is presented. The capacity expressions are obtained with Gamma function and evaluate different values of K_G fading parameters and are valid for integer values of k-m as well.

This paper is arranged in the following way. In Sect. 2, fading model of generalized K and its pdf expression is given. In Sect. 3, the formulas of capacity for adaptive transmission methods is presented and obtained each derived expression. The results and discussion part are presented in Sect. 4. In Sect. 5, the conclusion part is done.

2 Generalized *K*-Fading Model

In wireless communication, multipath and shadowing fading are the most common disturbance during signal transmission. The generalized K (K_G) distribution fading model delineates the effect of both fast and slow fading which are occurred in wireless communication received signal [14]. The K_G fading channels have two shaping parameters, i.e., m and k, where m represents the parameters of Nakagami which is concerned with short term fading and k is the Gamma distribution parameter which is represented the received average power because of shadowing. The pdf of the output SNR of K_G fading channels is [14]

$$f_\gamma(\gamma) = \frac{a^{\beta+1}}{2^\beta \Gamma(m)\Gamma(k)} \gamma^{\frac{\beta-1}{2}} K_\alpha\big(a\sqrt{\gamma}\big), \tag{1}$$

where $\alpha = k - m$, $\beta = k + m - 1$, $K_\alpha(\cdot)$ is the modified Bessel function of the second kind and order α, and $a = \sqrt{\frac{4km}{\bar{\gamma}}}$ here $\bar{\gamma}$ is the average signal to noise ratio SNR also $\Gamma(\cdot)$ is the Complete Gamma function and γ is the instantaneous SNR as well. The $K_\alpha(\cdot)$ is expanded using [16, (8.468)] and (1) is rewritten as follows:

$$f_\gamma(\gamma) = \frac{a^{\beta+1}}{2^\beta \Gamma(m)\Gamma(k)} \sqrt{\frac{\pi}{2a}} \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma\left(\alpha - \frac{1}{2} + t + 1\right)(2a)^{-t}}{\Gamma(t+1)\Gamma\left(\alpha - \frac{1}{2} - t + 1\right)}$$
$$\times \gamma^{\frac{2\beta-2t-3}{4}} e^{-a\sqrt{\gamma}}. \tag{2}$$

3 Adaptive Transmission System Capacity

The capacity of K_G fading channels expressions is derived under adaptive transmission methods which are employed at the transmitter side to adapt the transmission power and rate so as to get higher channel capacity [11].

3.1 Channel Inversion with Fixed Rate (CIFR)

The purpose of CIFR scheme is to provide fixed received rate by adjusting the transmission power. The channel capacity in bit/second/hertz is expressed as [11]

$$C_{CIFR}(\bar{\gamma}) = B \log_2\left(1 + \frac{1}{R_{CIFR}(\bar{\gamma})}\right), \tag{3}$$

where $R_{CIFR}(\bar{\gamma})$ is the rate of CIFR method and it is given as

$$R_{CIFR}(\bar{\gamma}) = \int_0^\infty \left(\frac{1}{\gamma}\right) f_\gamma(\gamma) d\gamma. \tag{4}$$

Putting (2) into (4), $R_{CIFR}(\bar{\gamma})$ is deduced as

$$R_{CIFR}(\bar{\gamma}) = \frac{a^{\beta+1}}{2^\beta \Gamma(m)\Gamma(k)} \sqrt{\frac{\pi}{2a}} \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma\left(\alpha - \frac{1}{2} + t + 1\right)(2a)^{-t}}{\Gamma(t+1)\Gamma\left(\alpha - \frac{1}{2} - t + 1\right)}$$
$$\times \int_0^\infty \gamma^{\frac{\beta-1}{2} - \frac{t}{2} - \frac{1}{4} - 1} e^{-a\sqrt{\gamma}} d\gamma \tag{5}$$

The variable x of (5) is changed, i.e., $x = \sqrt{\gamma}$, the capacity is rewritten as

$$R_{CIFR}(\bar{\gamma}) = \frac{a^{\beta+1}}{2^{\beta-1}\Gamma(m)\Gamma(k)}\sqrt{\frac{\pi}{2a}}\sum_{t=0}^{\alpha-\frac{1}{2}}\frac{\Gamma\left(\alpha-\frac{1}{2}+t+1\right)(2a)^{-t}}{\Gamma(t+1)\Gamma\left(\alpha-\frac{1}{2}-t+1\right)}$$

$$\times \int_0^{\infty} x^{\left(\frac{2\beta-2t-5}{2}\right)}e^{-ax}dx \tag{6}$$

The integral of (6) is solved using [16, (3.351.3)] and after algebraic manipulation the final expression of $R_{CIFR}(\bar{\gamma})$ is obtained as follows:

$$R_{CIFR}(\bar{\gamma}) = \frac{\sqrt{\pi}a^2}{2^{\beta-0.5+t}\Gamma(m)\Gamma(k)}\sum_{t=0}^{\alpha-\frac{1}{2}}\frac{\Gamma\left(\alpha-\frac{1}{2}+t+1\right)}{\Gamma(t+1)\Gamma\left(\alpha-\frac{1}{2}-t+1\right)}$$

$$\times \sqrt{\frac{\pi}{2a}}\Gamma\left(\beta-t-\frac{3}{2}\right) \tag{7}$$

The final expression of CIFR capacity is achieved by putting (7) into (3).

3.2 Truncated Channel Inversion with Fixed Rate (TIFR)

In order to get constant received rate with acceptable SNR level, a large transmission power is mandatory in the methods of CIFR. However, it becomes a big challenge to transmit this massive power especially for mobile and power limited transmitters. During deep fade, all the transmitted power are noise dominated so disturbance occur in decoding process. Therefore, TIFR method is engaged so as to surmount the effect of CIFR method. In TIFR transmitter, transmission immediately stops when the receiver SNR falls below the threshold level (γ_{th}), and the channel inversion occurs if the received SNR is higher than γ_{th}. The TIFR capacity expression is [11]

$$C_{TIFR}(\bar{\gamma}) = B\,\log\left(1+\frac{1}{R_{TIFR}(\bar{\gamma})}\right)(1-P_{out}(\gamma_{th})), \tag{8}$$

where

$$R_{TIFR}(\bar{\gamma}) = \int_{\gamma_{th}}^{\infty}\left(\frac{1}{\gamma}\right)f_{\gamma}(\gamma)d\gamma \tag{9}$$

and

$$P_{out}(\gamma_{th}) = \int_0^{\gamma_{th}} f_{\gamma}(\gamma)d\gamma \tag{10}$$

is the outage probability having threshold value γ_{th}.

For solving the capacity of TIFR, it is needed to calculate $R_{TIFR}(\bar{\gamma})$ and $P_{out}(\gamma_{th})$. The expression of $R_{TIFR}(\bar{\gamma})$ is rewritten by putting (2) into (9) and change the variable $x = \sqrt{\gamma}$ as follows:

$$
\begin{aligned}
R_{TIFR}(\bar{\gamma}) = {} & \frac{\sqrt{\pi}a^{\beta+1}}{2^{\beta-1+0.5+t}a^{0.5+t}\Gamma(m)\Gamma(k)} \\
& \times \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma\left(\alpha - \frac{1}{2} + t + 1\right)}{\Gamma(t+1)\Gamma\left(\alpha - \frac{1}{2} - t + 1\right)} \int_{\sqrt{\gamma_{th}}}^{\infty} x^{\frac{2\beta-2t-5}{2}} e^{-ax} dx
\end{aligned}
\tag{11}
$$

The integral of (11) is solved by using [16, (3.381.3)]. And after simplification the final expression is

$$
\begin{aligned}
R_{TIFR}(\bar{\gamma}) = {} & \frac{\sqrt{\pi}a^2}{\Gamma(m)\Gamma(k)} \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma\left(\alpha - \frac{1}{2} + t + 1\right)}{2^{\beta-1+0.5+t}\Gamma(t+1)\Gamma\left(\alpha - \frac{1}{2} - t + 1\right)} \\
& \times \Gamma\left(\frac{2\beta - 2t - 3}{2}, a\sqrt{\gamma_{th}}\right),
\end{aligned}
\tag{12}
$$

where $\Gamma(.,.)$ is the upper incomplete Gamma function. The $P_{out}(\gamma_{th})$ expression is rewritten by putting (2) into (10) and deduced in term of infinite series and the final expression is obtained as

$$
\begin{aligned}
P_{out}(\gamma_{th}) = {} & \frac{\sqrt{\pi}}{\Gamma(m)\Gamma(k)} \sum_{t=0}^{\alpha-\frac{1}{2}} \frac{\Gamma\left(\alpha - \frac{1}{2} + t + 1\right)(2a)^{-t}}{2^{\beta+t+0.5}\Gamma(t+1)\Gamma\left(\alpha - \frac{1}{2} - t + 1\right)} \\
& \times \sum_{p=0}^{\infty} \frac{(-1)^p a^{\beta+1-0.5-t+p}}{p!} \frac{4}{2\beta - 2t + 2p + 1} \gamma^{\frac{2\beta-2t+2p+1}{4}}
\end{aligned}
\tag{13}
$$

Then, the TIFR capacity is obtained after substituting (12) and (13) in (8).

3.3 Constant Transmitting Power

The transmitter adopts Optimal Rate Adaption (ORA) when the rate of transmission varies according to the condition of the channels even though the power of transmission is fixed. The channel capacity in bit/sec/hertz of ORA method is [11]

$$
C_{ORA} = B \log_2 e \int_0^{\infty} \ln(1 + \gamma) f_{\gamma}(\gamma) d\gamma
\tag{14}
$$

The ORA capacity expression is rewritten by putting (2) into (14) and changes the variable $x = \sqrt{\gamma}$ then the expression is obtained as

$$
C_{ORA} = B \log_2 e \frac{a^{\beta+1}}{2^{\beta-1}\Gamma(m)\Gamma(k)} \sqrt{\frac{\pi}{2a}}^{\alpha-\frac{1}{2}} \sum_{t=0}^{\infty} \sum_{p=0}^{\infty} \frac{(-1)^p a^p}{p!}
$$

$$
\times \frac{\Gamma(\alpha+t+0.5)(2a)^{-t}}{\Gamma(t+1)\Gamma(\alpha-t+0.5)} \int_0^{\infty} \ln(1+x^2).x^{\frac{2\beta+2p-2t-1}{2}} dx \qquad (15)
$$

The integral of (15) is solved by using [16, (4.295.40)]. And after simplification the final expression of ORA capacity is obtained.

4 Numerical Results and Discussion

Figure 1 exhibits capacity versus average SNR $(\bar{\gamma})$ for CIFR method. The figure shows that the channel capacity is improved when the shaping parameter k value increases in fixed value of m, i.e., (m = 1). The increased in value of k indicates decrease in the severity of shadowing, resulting to a better channel. Hence, the capacity increases with the value of k. In other words, small value of k means heavy shadowing condition and larger value of k means light shadowing. Figure 2 shows that when the value of m increases with fixed value of k (k = 5), the capacity of the channel increases. This is because higher value of m indicated lesser amount of fading. It can also be concluded that value of m can take both integer and non-integer values. Figure 3 shows average SNR versus capacity for TIFR method. It is concluded that

Fig. 1 Capacity against average SNR for CIFR method and channel fading condition
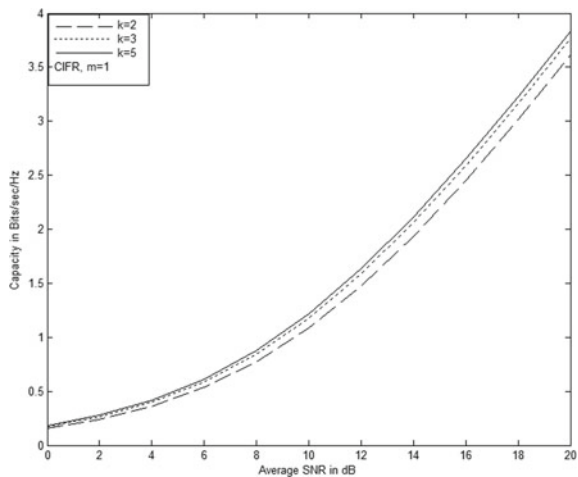
Fig. 2 CIFR capacity when k is fixed
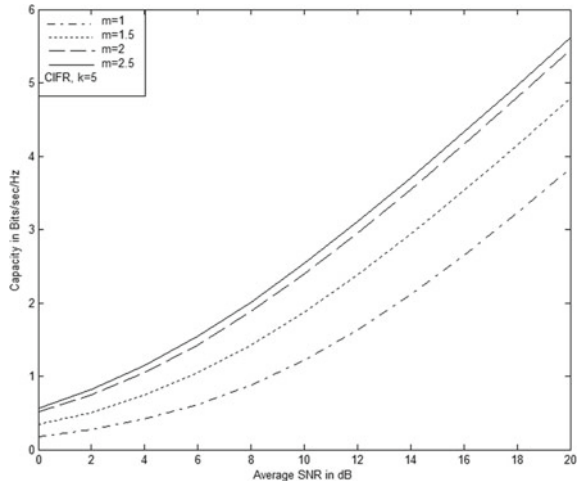
Fig. 3 Capacity for TIFR method
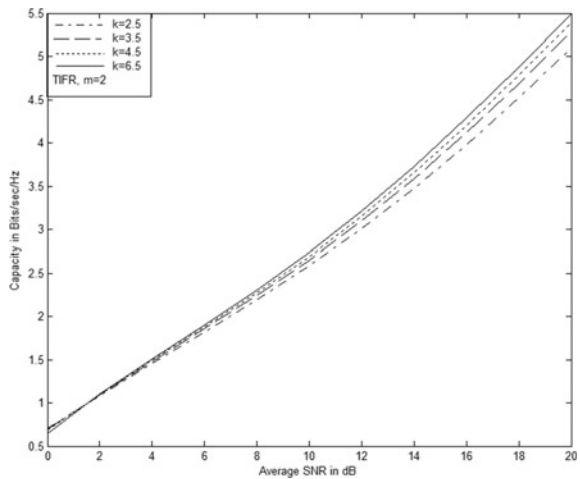
the channel capacity improves with the increase in the value of k for a fixed value of m. This is because increased in k value means the received average power increases hence the capacity increases.

Figure 4 shows the comparison between CIFR and TIFR capacity methods when the value of m is fixed. It is observed from the plot that in lower SNR capacity of TIFR is higher than the capacity of CIFR method. But in case of higher SNR the result of CIFR and TIFR capacity are almost near to each other. The main region is that for TIFR during the time of heavy shadowing the received signal SNR is below the threshold and hence TIFR transmitter immediately stops transmission.

Fig. 4 Comparison between
CIFR and TIFR capacity
methods
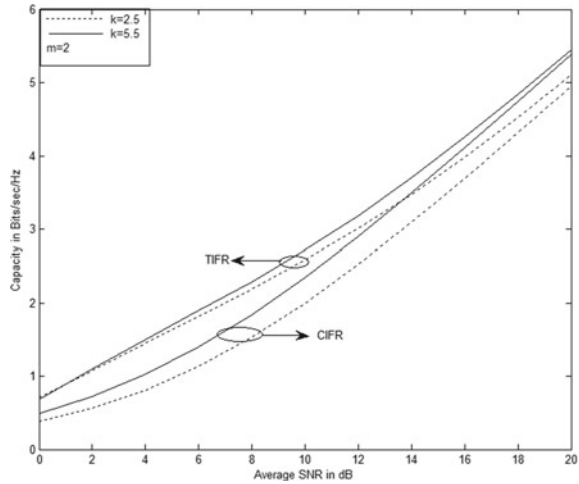
5 Conclusions

The channel capacity expressions under adaptive methods, i.e., Channel Inversion
with Fixed Rate, Truncated Channel Inversion with Fixed Rate and Constant Trans-
mitting Power in term of Gamma function over generalized K-fading channel are
derived using the pdf-based approach. The expressions are numerically evaluated
for different values of fading parameters, including integer values of k-m as well.
The obtained expressions can be used to estimate accurate mathematical model in
composite radio environment to trim down the effect of fading.

References

1. Simon MK, Alouini MS (2005) Digital communication over fading channels, a unified approach
   to performance analysis, 2nd edn. Wiley, New York
2. Abdi A, Kaveh M (1998) K distribution: an appropriate substitute for Rayleigh-lognormal
   distribution in fading-shadowing wireless channels. Electron Lett 34(9):851–852
3. Bithas PS, Sagias NC, Mathiopoulos PT, Karagiannidis GKK, Rontogiannis AA (2006) On
   the performance analysis of digital communications over generalized-K fading channels. IEEE
   Commun Lett 10(5):353–355
4. Bithas PS, Mathiopoulos PT, Kotsopoulos SA (2007) Diversity reception over generalized-K
   (KG) fading channels. IEEE Trans Wirel Commun, 4238–4243
5. Trigui I, Laourine A, Affes S, Stephenne A (2009) On the performance of cascaded generalized
   K fading channels. IEEE Commun Lett 10(5):353–355
6. Matthaiou M, Chatzidiamantis ND, Karagiannidis GK, Nossek JA (2010) On the capacity of
   generalized K-fading MIMO channels. IEEE Trans Signal Process 58(11)
7. Lateef HY, Ghogho M, McLernon DC (2011) Performance analysis of multi-hop cooperative
   relay networks over generalized-K fading channels. IEEE, 978-1-4244-9715-7/11/$26.00

8. Subadar R, Das P (2017) Performance of L-SC receiver over K-fading channels. In: 8th ICC-CNT
9. Alouini M-S, Goldsmith AJ (1999) Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques. IEEE Trans Veh Technol 48:1165–1181
10. Singh AD, Subadar R (2014) Capacity analysis of SC receiver over TWDP fading channel. (AEU) 68(2):166–17I
11. Singh AD, Subadar R (2015) Capacity analysis of MRC receiver with adaptive transmitters over TWDP fading channels. In: ISACC
12. Singh AD, Subadar R (2015) Capacity of dual-MRC receiver with adaptive transmitters over non identical TWDP fading channels. In: International conference on advances in computer engineering and applications (ICACEA). IMS Engineering College Ghaziabad India
13. Khatalin S, Fonseka JP (2006) On the channel capacity in Rician and Hoyt fading environment with MRC diversity. IEEE Trans Veh Technol 55(1):137–141
14. Laourine A, Alouni M-S, Affes S, Stephenne A (2008) On the capacity of generalized-K fading channels. IEEE Trans Wirel Commun 7(7)
15. Efthymoglou GP, Ermolova NY, Aalo VA (2010) Channel capacity and average error rates in generalised-K fading channels. IET Commun 4(11):1364–1372
16. Gradshteyn IS, Ryzhik IM (2000) Table of integrals, series, and products, 6th edn. Academic, San Diego

Speckle Reduction of Ultrasound Image via Morphological Based Edge Preserving Weighted Mean Filter

Dipannita Ghosh, Amish Kumar, Palash Ghosal and Debashis Nandi

Abstract Medical ultrasound images suffer from inherently generated speckle noise that makes difficult the radiologists to diagnose the diseases. Proper speckle reduction technique is required to improve quality of image which may help the doctor to diagnose the diseases correctly. Speckle removal is, specially, a filtering technique that reduces the amount of speckle noise. But as an effect of filtering, the edges of objects may become blur and fine details within the image may be lost. Many methods have already been proposed to achieve these two requirements. But they do not satisfy both up to the desirable level. In this paper, we propose a novel morphological operation based speckle filter which reduces the effect of speckle noise to the greater extent in one side and keeps detail information in other side. The use of morphological operators helps in extracting the structures present within the images followed by an edge preserving weighted mean filtering method that may help in despeckling of ultrasound images. The proposed technique is applied on the ultrasound simulated data as well as on the real data obtained from the ultrasound machine. The quantitative results and the output images confirm the superiority of the method when compared with some of the other contemporary speckle reduction methods.

Keywords Ultrasound imaging system · Speckle noise · Speckle filtering · Morphological operation · Weighted mean filter · Thresholding

D. Ghosh (✉) · A. Kumar (✉) · P. Ghosal · D. Nandi
Department of Computer Science Engineering, National Institute of Technology, Durgapur, Durgapur 713209, West Bengal, India
e-mail: dipannitaghosh21@gmail.com

A. Kumar
e-mail: amishkumar562@gmail.com

P. Ghosal
e-mail: ghosalpalash@gmail.com

D. Nandi
e-mail: debashisn2@gmail.com

307

1   Introduction

Ultrasound imaging systems have gained their popularity in the domain of medi-cal diagnosis due to their noninvasive nature, real-time operation, and lesser cost as compared to the other medical imaging modalities. Every year, the applications of B-mode ultrasound imaging are being expanded more and more and newer applica-tions are being introduced in medical diagnosis. However, sometimes, the accuracy of diagnosis falls due the quality of generated image. A particular type of noise pat-tern, called speckle, is formed due to coherent nature of ultrasound imaging. This speckle affects the resolution and contrast of the generated image and disrupts the accuracy of diagnosis. In many cases, it causes miss-classifications of tumors and other diseases. Therefore, suppression of speckle is one of the most important tasks as a preprocessing operation for improving the image quality and thereby, improving the accuracy of segmentation for anatomical structures in ultrasound images.

Speckle reduction is a task of filtering which improves the smoothness of the image. But this improvement of smoothness may decrease the contrast and make the object boundary blur which indicates the loss of fine details from the image. Moreover, sometimes smaller objects may be removed from the image due to filtering effect. Hence, it is a challenging task to suppress speckle noise by protecting the objects' boundaries in the image. Various speckle filtering algorithms have been developed so far to meet such requirements but being having their own merits and demerits. Some of the frameworks provide better smoothing while other gives better visual interpretation. The major challenge in speckle reduction filters is to suppress noise without degrading the sharp transitions in intensity value of the image.

With this objective, the present article proposes a speckle reduction method by developing a weighted mean filter based on morphological operator and consequently followed by thresholding based artifact removal procedure. Detailed explanation of each operation is given in Sect. 3. Section 2 is discussed with the review of previous works. The output images, experimental results and their analysis have been devoted in Sect. 4. Finally, the efficiency of the proposed algorithm is concluded in Sect. 5.

2   Review of Previous Works

In this section, we present a brief review of some previous works. In the decade of 1980s and 1990s, a number of adaptive linear and nonlinear filters such as Lee [1], Kuan [2], Bamber and Daft [3], and Frost [4] were proposed to suppress speckle noise from SAR and ultrasound images. The depth of filtering of all these filters is determined from the local statistics of filtering window. Though this process is capable of reducing speckle in homogeneous region, it passes noise in the vicinity of the objects' edges. If we try to reduce the noise with large extent around the edges, the edges become blur. In 1990, Perona and Malik [5] have proposed anisotropic diffusion based noise reducing algorithm which provides a significant improvement

over the other contemporary methods. This conventional ultrasonic imaging is then tailored to ultrasonic imaging known as speckle reducing anisotropic diffusion filter [6]. In parallel with this, some other different techniques based on wavelet filters [7, 8] had been proposed. During 2005, filters based on nonlocal means [9] has been developed and later extended with many other modifications [10]. Later, a trend of hybrid technique came in the literature of ultrasound denoising filters where the conventional methods are merged accordingly to develop the new method of speckle reduction [11]. However, the method developed in this paper has been compared with some of the recent speckle reduction methods developed in the last decades like NLM [9], OBNLM [10], SRAD [6], DWT [8], and HMM [11].

3 Proposed Method of Speckle Reduction

3.1 Morphological Operation

The use of mathematical morphological operation is being found in the reduction of speckle noise from medical ultrasound imaging [12]. Since the presence of speckle noise primarily obscures the object boundaries of ultrasound image, the present article involves morphological operation for retention of the shapes and boundaries of the objects. Morphological closing tends to smooth the contour regions removing small holes and filling gaps in the contour. With this consideration, the present article has applied morphological closing on the speckled ultrasound image I for any pixel (x, y). M is the morphologically operated image given by

$$M(x, y) = \text{closing}(I(x, y)) = \text{erosion}(\text{dilation}(I(x, y)) \tag{3.1}$$

The erosion and dilation using the structuring element $s(a, b)$ can be mathematically represented as follows:

$$\text{erosion}(I(x, y)) = \min\{I(x + a, y + b) - s(a, b)\} \tag{3.2}$$

$$\text{dilation}(I(x, y)) = \max\{I(x + a, y + b) - s(a, b)\} \tag{3.3}$$

Further, an edge preserving weighted mean filtering operation is performed, which is eventually giving rise to certain artifacts. To eliminate these artifacts, an artifact removal operation is employed thereafter to produce the edge preserved noise reduced ultrasound image. The flow diagram of the proposed speckle reduction algorithm is given in Fig. 1.

Fig. 1 Flow diagram of the
proposed method
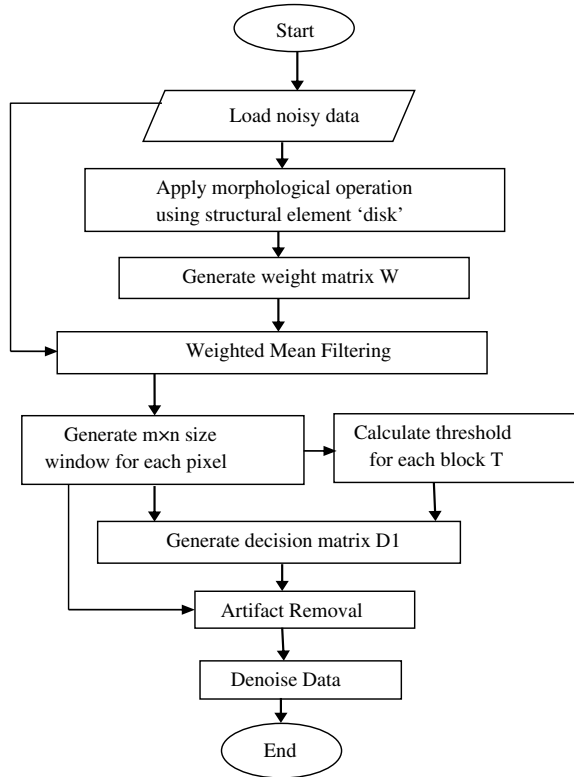
3.2   Weight Estimation

The morphological processed image obtained from the previous step retains the
shape and pattern of the objects contained within the images. But, our objective
is to generate a noise reduced image with the maximum protection of edge. Thus,
to preserve the edge of the image, in this step, an edge preserving weighted mean
filtering is performed on the speckle image using a weight matrix W that gives
rise to another processed matrix Z. W has been calculated from the morphological
processed image M given by

$$W(x, y) = \frac{1}{\sqrt{2\pi \times \text{var}(P(x, y) - \text{mean}(P))}} e^{-\left(\frac{(P(x, y) - \text{mean}(P))^2}{2 \times \text{var}(P(x, y) - \text{mean}(P))}\right)} \qquad (3.4)$$

where P is the $m \times n$ window of M and var (.) function computes the variance over
the window. The processed matrix Z can be computed as

$$Z(i, j) = \sum_{x=-k}^{k} \sum_{y=-l}^{l} I(i + x, j + y) \times W(k + x, l + y), k = \left\lfloor \frac{m}{2} \right\rfloor, l = \left\lfloor \frac{n}{2} \right\rfloor \quad (3.5)$$

Z is the image that may preserve edge reduce noise but contain some artifacts which is to be removed by an artifact removal technique. The next step describes the artifact removal technique applied on Z.

3.3 Artifact Removal

To remove the artifacts from Z, the following operations are performed:

(a) A threshold value is selected as

$$T = K \times \text{mean}(w)$$

where w is the $m \times n$ square window of Z and K is the constant taken as 0.5.

(b) Again, a distance matrix D has been created which is the difference between the center pixel value (w_c) of the window of Z and its neighborhood. The matrix D is represented as

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots\cdots & d_{1,n} \\ d_{2,1} & d_{2,2} & \cdots\cdots & d_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{m,1} & d_{m,2} & \cdots\cdots & d_{m,n} \end{bmatrix} \quad (3.6)$$

where $d_{x,y} = |w_c - w_{x,y}|$; $x = 1, 2, \ldots m$ and $y = 1, 2, \ldots n$

(c) Based on the threshold value T, a decision matrix is formed as

$$D1 = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots\cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots\cdots & b_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & b_{m,2} & \cdots\cdots & b_{m,n} \end{bmatrix} \quad (3.7)$$

The elements of the matrix $D1$ are evaluated by using the following rule:

$$b_{i,j} = \begin{cases} 0, \; if \; d_{x,y} > T \\ 1, \; if \; d_{x,y} < T \end{cases}$$

(d)   Finally, the processed image $Y(i, j)$ is obtained as,

$$Y(i, j) = \sum_{x=-k}^{k} \sum_{y=-l}^{l} Z(i + x, j + y) \times D1(k + x, l + y), k = \left\lfloor \frac{m}{2} \right\rfloor, l = \left\lfloor \frac{n}{2} \right\rfloor$$

(3.8)

4   Experimental Results and Analysis

The performance of the proposed speckle reduction algorithm is verified by comparing it with some of the recent speckle reduction methods like NLM, OBNLM, SRAD, DWT, and HMM. In this experiment, the synthetic image of a cyst (created using Matlab), a kidney (obtained from Field II simulation program library) [13] and a real ultrasound image of benign tumor (available at Ultrasound Imaging Laboratory at the University of Illinois) [14] are taken to quantify the robustness of the other filters with that of the proposed method. The synthetic images have been embedded with Rayleigh distributed speckle noise using MATLAB software for different $\sigma = 0.8$, 0.9 and 1. The output images of the proposed method are shown in Figs. 2 and 3 for $\sigma = 0.8$. For cyst image (Fig. 2), it is found that NLM and OBNLM filter smooth the noise pixels which is desirable in the homogeneous region but hampers the detection of the edge prone area. For the edge region, the boundary of the fine objects gets dispersed and fused with its background pixels. SRAD filter can eliminate the noise to certain extent but at the same time makes the edge region blur. HMM though takes the properties of both noise removing and edge preserving filters, it removes noise and preserves the prominent edges but fails in retaining the objects with weak edges. As seen from the output images, the proposed algorithm outperforms in terms of preservation of edges which makes the minute details visible and distinguishable. It also shows that the proposed algorithm is better in reducing noise as compared with the other techniques. The same observation holds true for the kidney image as shown in Fig. 3 and also worked for the real image of ultrasound benign tumor (Fig. 4). The plots of the quality metrics for both the cyst and kidney image are shown in Fig. 5.

4.1   Quantitative Analysis

For any quality enhancement algorithm of images, the capability of denoising is measured using quantitative evaluation index. The evaluation of image enhancement can be done by computing the quality metrics like PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index) [15], Beta [16], and EKI (Edge keeping Index) [17]. The quality metrics of Tables 1 and 2 show that the proposed method is giving considerably better result with respect to the other speckle reduction algorithms listed.
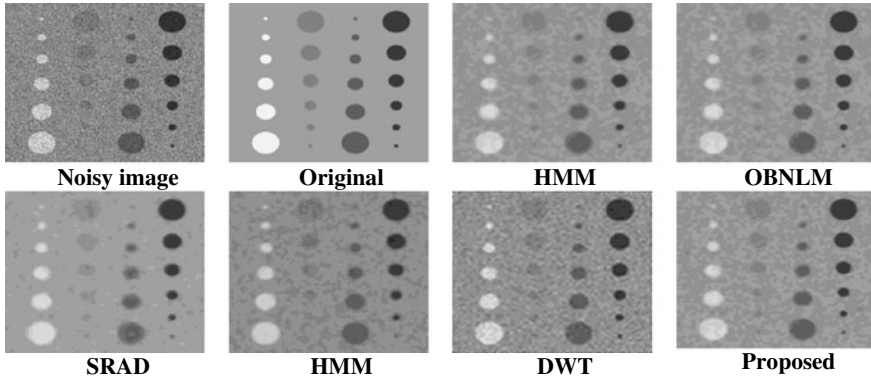
Fig. 2 Output images for speckle noise $\sigma = 0.8$
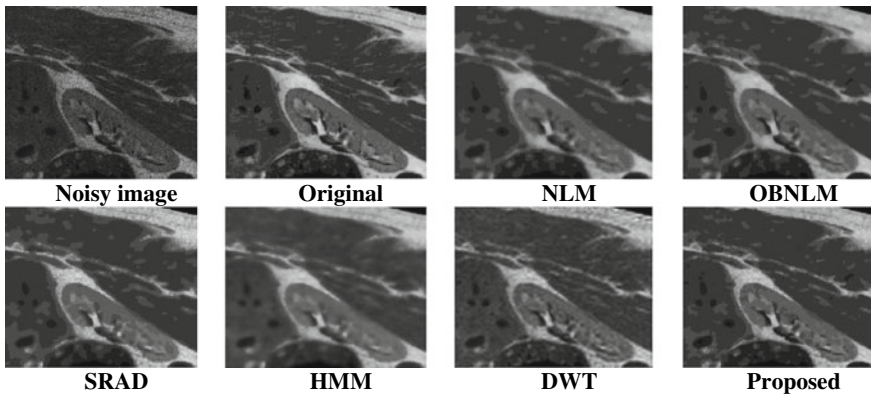
Fig. 3 Output images for speckle noise, $\sigma = 0.8$
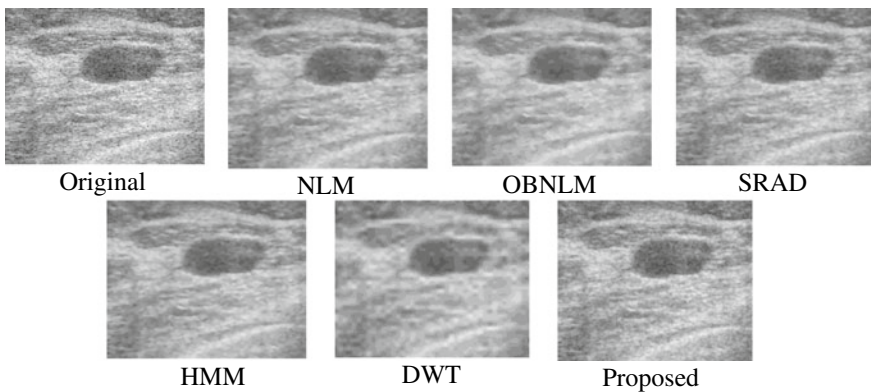
Fig. 4 Output images of real ultrasound of benign tumor
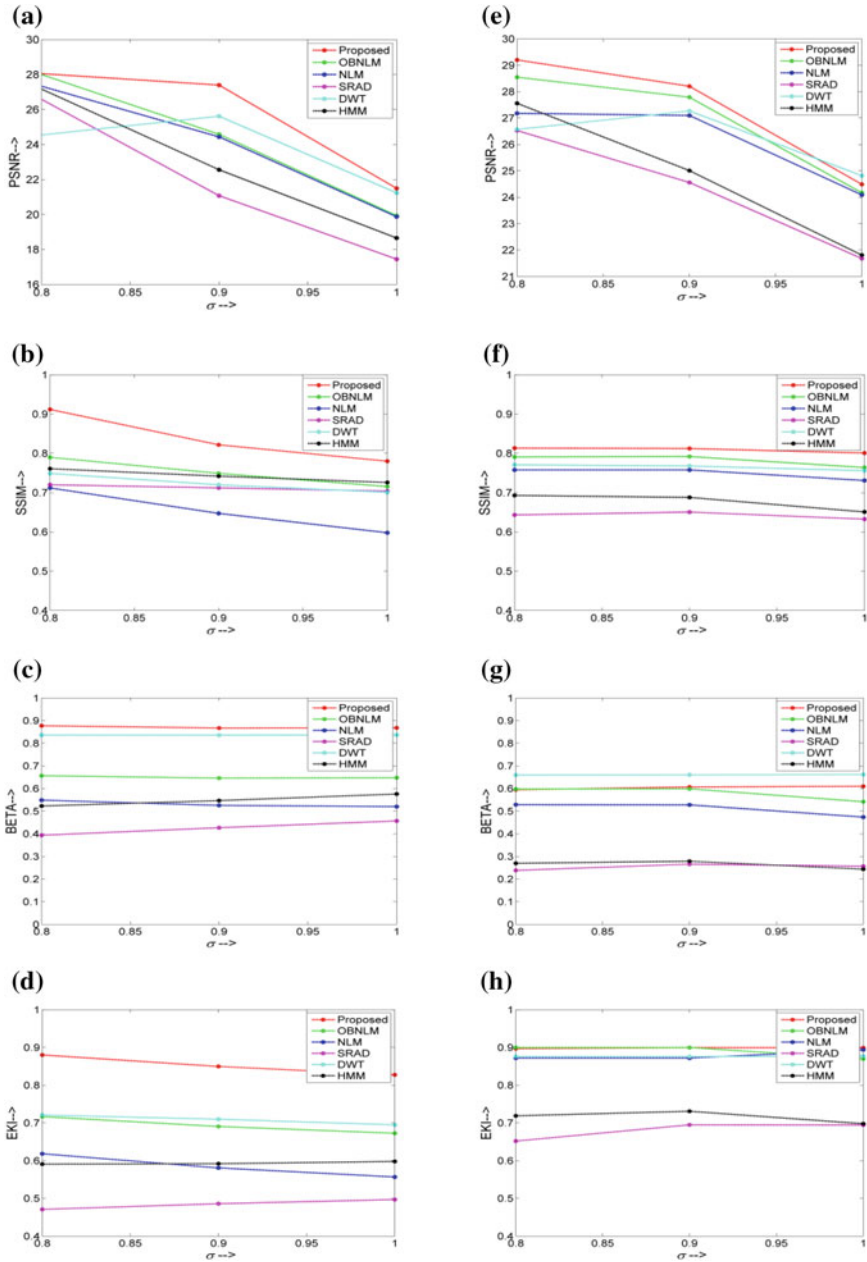
Fig. 5 Plot of quality metrics for cyst (**a**, **b**, **c**, **d**) and Kidney cut (**e**, **f**, **g**, **h**) image for different level of noise in terms of PSNR, SSIM, BETA, and EKI

Table 1 Quality metrics of cyst image

	NLM	OBNLM	SRAD	DWT	HMM	Proposed
$\sigma = 0.8$						
PSNR	27.33	28.00	26.58	24.55	27.17	28.05
SSIM	0.71	0.79	0.72	0.75	0.76	0.91
BETA	0.55	0.66	0.39	0.84	0.52	0.88
EKI	0.62	0.72	0.47	0.72	0.59	0.88
$\sigma = 0.9$						
PSNR	24.44	24.58	21.08	25.62	22.56	27.39
SSIM	0.65	0.75	0.71	0.72	0.74	0.82
BETA	0.53	0.65	0.43	0.84	0.55	0.88
EKI	0.58	0.69	0.49	0.71	0.59	0.85
$\sigma = 1$						
PSNR	19.88	19.94	17.45	21.24	18.66	21.49
SSIM	0.60	0.72	0.70	0.70	0.73	0.78
BETA	0.52	0.65	0.46	0.84	0.58	0.87
EKI	0.56	0.67	0.50	0.69	0.60	0.83

Table 2 Quality metrics of kidney cut image

	NLM	OBNLM	SRAD	DWT	HMM	Proposed
$\sigma = 0.8$						
PSNR	27.18	28.55	26.53	26.57	27.56	29.20
SSIM	0.76	0.79	0.64	0.77	0.69	0.81
BETA	0.53	0.60	0.24	0.66	0.27	0.59
EKI	0.87	0.90	0.65	0.88	0.72	0.90
$\sigma = 0.9$						
PSNR	27.10	27.79	24.56	27.27	25.01	28.21
SSIM	0.76	0.79	0.65	0.77	0.69	0.81
BETA	0.53	0.60	0.27	0.66	0.28	0.61
EKI	0.87	0.90	0.69	0.87	0.73	0.90
$\sigma = 1$						
PSNR	24.09	24.17	21.68	24.82	21.81	24.49
SSIM	0.73	0.76	0.63	0.75	0.65	0.80
BETA	0.47	0.54	0.26	0.66	0.24	0.61
EKI	0.89	0.87	0.69	0.88	0.70	0.89

5   Conclusion

In this paper, a morphological based despeckling algorithm for ultrasound image has been presented. The algorithm performs weighted neighborhood operations with the morphological processed image to obtain edge protected artifact free speckle reduced ultrasound images. The proposed method has been carried out on both synthetic and real ultrasound images and its performance has been compared with some of the recently developed speckle noise removal algorithm in medical ultrasound image. Visual and quantitative analysis substantially proves the overall enhancement of ultrasound image through the proposed method over the other recent methods.

References

1. Lee JS (1980) Digital image enhancement and noise filtering by use of local statistics. IEEE Trans Pattern Anal Mach Intell 2(2):165–168
2. Kuan DT, Sawchuk AA, Strand TC, Chavel P (1987) Adaptive restoration of images with speckle. IEEE Trans Acoust Speech Signal Process 35(3):373–383
3. Bamber JC, Daft C (1986) Adaptive filtering for reduction of speckle in ultrasonic pulse-echo images. Ultrasonics 24(1):41–43
4. Frost VS, Stiles JA, Shanmugan KS, Holtzman JC (1982) A model for radar images and its application to adaptive digital filtering for multiplicative noise. IEEE Trans Pattern Anal Mach Intell 4(2):157–166
5. Perona P, Malik J (1990) Scale-space and edge detection using anisotropic diffusion. IEEE Trans Pattern Anal Mach Intell 12(7):629–639
6. Yu Y, Acton ST (2002) Speckle reducing anisotropic diffusion. IEEE Trans Image Process 11(1):1260–1270
7. Donoho D, Johnstone I (1994) Ideal spatial adaptation via wavelet shrinkage. Biometrika 81:425–455
8. Michailovich OV, Tannenbaum A (2006) Despeckling of medical ultrasound images. IEEE Trans Autom Control AC 53(1):64–78
9. Buades A, Coll B, Morel JM (2005) A non-local algorithm for image denoising. Proc CVPR 2:60–65
10. Coupe P, Hellier P, Kervrann C, Barillot C (2009) Nonlocal means-based speckle filtering for ultrasound images. IEEE Trans Image Process 18:2221–2229
11. Gungor MA, Karagoz I (2015) The homogeneity map method for speckle reduction in diagnostic ultrasound images. Measurement 68:100–110
12. Kher A, Mitra S (1994) An efficient speckle removing algorithm based on new morphological operators. Circuits and systems. In: Proceedings of the 37th Midwest symposium, vol 2, pp 852–855 (1994)
13. Field II simulation program library. https://field-ii.dk/examples/ftp_files/kidney/kidney_cut.bmp
14. Ultrasound Imaging Laboratory at the University of Illinois. http://ultrasonics.bioengineering.illinois.edu/data_patient.asp
15. Wang Z, Bovik AC (2004) Image quality assessment: from error visibility to structure similarity. IEEE Trans Image Process 13(4):600–612
16. Sattar F, Floreby L, Salomonsson G, Lovstrom B (1997) Image enhancement based on a nonlinear multiscale method. IEEE Trans Image Process 6(6):888–895
17. Kini M, Pandey R, Das A, Malani SK (2014) Comprehensive image processing for automated detection of hypertrophic cardiomyopathy. IJCER 3(2):56–62

Autonomous Vehicle—A Miniaturized Prototype Development

Jayanta K. Baruah, Ashutosh Kumar, Rabindranath Bera and Sourav Dhar

Abstract Driverless vehicles will become the reality in upcoming days and will redefine the future mobility of travelers and goods. According to protagonists of autonomous vehicle (AV), the accidents caused by the manual error, which is regarded as the major cause of road accidents, would be eliminated completely. Thus, almost all automobile giants are working toward the improvement of AV. In this work, first, a case study on Google driverless car is presented. Motivated by this case study, the development of a miniaturized prototype AV has been presented in this work. This prototype AV is equipped with two ultrasonic sensors and featured with automatic maneuvering and parking.

Keywords ADAS · AV · Ultrasonic sensors · Parking assistance

1 Introduction

A driverless car alternately termed as autonomous vehicle (AV) is basically a robotic automobile that is designed to travel between places without a human operator. The mandatory feature of an AV is the ability of navigation without human intervention to a scheduled terminus over streets that have not been adapted for its use. Maneuvering being a prime feature of driverless car in obstacle detection and path searching has a huge impact on perception and reaction of people toward an autonomous system. AVs continuously monitor their surroundings using range sensors (radar,

J. K. Baruah (✉) · A. Kumar · R. Bera · S. Dhar (✉)
Sikkim Manipal Institute of Technology, Majitar, Rangpo 737136, Sikkim, India
e-mail: jkbaruah01@yahoo.com

S. Dhar
e-mail: sourav.d@smit.smu.edu.in; sourav.dhar80@gmail.com

A. Kumar
e-mail: ashutoshsmit@gmail.com

R. Bera
e-mail: rbera@smit.smu.edu.in; rbera50@gmail.com

317

sonar, LIDAR, GPS) and computer vision. However, computer vision is a more complex and expensive obstacle recognition technique compared to the range sensors based method. Thus, the majority of commercial self-governing robots use range sensor to identify obstacles. The use of radar, IR sensors, and ultrasonic sensors for evolving obstacle discovery system is in use since 1980s [1]. Use of charge-coupled device (CCD) camera for vehicular safety is presented in [2, 3]. The sensors are not only limited to obstacle detection. Other sensors could be used to extract different features. Sonar is also another good obstacle detection range sensor and systems based on sonar have been presented in [4, 5]. The ultrasonic sensors are equally effective for obstacle avoidance [6]. The use of optical sensors (laser scanner) has been presented in [7] for obstacle avoidance. The concept of support vehicle machine (SVM) has been introduced in [8] for making local track for unmanned ground vehicle. How can unmanned ground vehicle be utilized for remote-controlled surveillance is shown in [9]. Finally, an overall ranking of different sensors used in advanced driver assistance system (ADAS) has been presented in [10]. Thus, the objective of this work is to design a simple, cost-effective obstacle avoidance AV using two ultrasonic sensors.

1.1 Case Study: Google Autonomous Car

The control unit of the car computes the most appropriate route toward the predefined destination and starts the car on its way. The 3D dynamic mapping for the vehicles current environment is created by roof-mounted LIDAR monitoring over a range of 60 m around the car. There are radar systems mounted in front of the car as well as rear bumpers, and this radar assembly is used to calculate the distance of targets/obstacles. There is a sensor system mounted on the left rear wheel which is used to monitor the sideward movement of the car as well as its relative position on the 3-D map. All the sensors and video cameras in the car are connected to an artificial intelligence (AI) driven control unit in the car. The control unit also takes input from Google Street View. The control unit simulates decision-making other cognitive capabilities of human to control steering, brakes, and other driver activities. The Google Maps are continuously consulted for prior information about landmarks, traffic signs, etc. However, there is also a manual mode activation facility available to take manual control of the vehicle [11].

1.2 ADAS and AV

As shown in Fig. 1, perception, mapping, and localization, the integral part of an ADAS, are the major concern of the automobile companies for the protection of AVs [12]. Fully autonomous cars must be able to take decisions precisely and continuously during travel. For taking decisions precisely according to the situations,
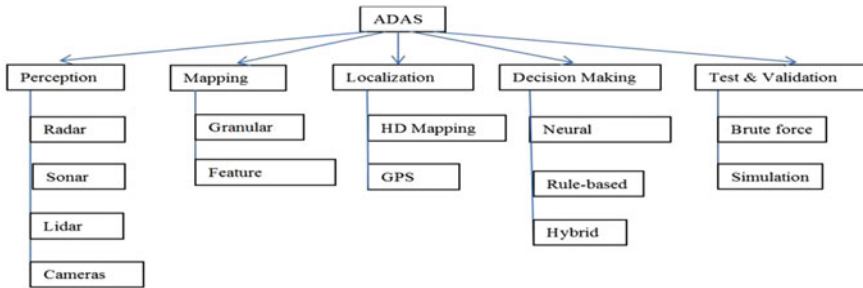
Fig. 1 Functional block diagram of ADAS

today's decision-making systems employ multi-criteria decision-making (MCDM) techniques in addition to neural networks [10, 13]. The automobile industry is performing repeated testing and validation before the mass production of the AVs [12, 14, 15]. The commercial availability of fully autonomous cars might take a little more time, but the corporates are already concerned about what the final AV archetype will appear.

2 Problem Definition

During the development of this prototype, two most important problems related to AVs have been addressed:

1. Parking assistance

   It deals with precise guidance (either autonomously or semi-autonomously) to the vehicle during parking within a limited available space.

2. Maneuvering and obstacle avoidance

   One of the most important capabilities expected from an AV is avoiding collision with obstacles in its path. For this purpose, AV should be able to carry out sudden maneuver as soon as the obstacle is detected.

3 Prototype Development

3.1 Hardware Development

The system consists of power supply unit, Arduino Nano, LM298 H bridge module, Ultrasonic Sensor modules, Geared DC Motors, Micro Servo motor. The block diagram of the system is shown in Fig. 2.

Fig. 2 Block diagram of the system
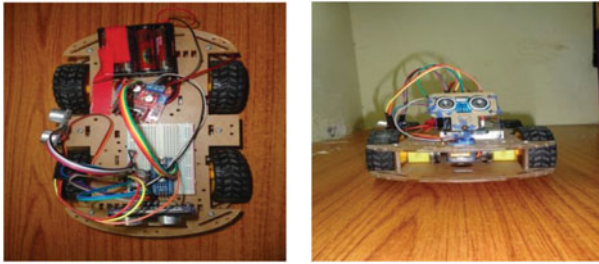
Fig. 3 Driverless car prototype

In this work, the distance of the object is measured by two ultrasonic sensors, and the sensor output is connected to signal conditioning unit which is connected to Arduino Nano board. The measured results are displayed on the computer monitor. The sensor is attached to servo motor to find the polar distance around the sensor up to 180° rotations (Fig. 3).

This application is also used to find the obstacles detection and the exact distance can also be obtained. Basically, in our car, there are two features: one is for parking assistance system and another is for obstacle detection and pathfinding. The measured distance is displayed on the computer monitor. The hardware components of the system are, viz., two ultrasonic sensor modules, Servo motor, Arduino Nano board with ATmega 328 Microcontroller and computer respectively. If any obstacle is detected at the range of less than or equal to 20 cm, the servo will rotate at 180°, and the ultrasonic sensor1 is situated on the top of it and accordingly takes the decision to move where the path is clear. The ultrasonic sensor2 which is located below the chassis of the car detects any obstacle under the car and displays the distance in the serial monitor of the computer.

3.2 Software Development

Figure 4 shows the activity chart of the AV at initialization of the sensors which becomes active when the motor is actuated to move in the forward direction simultaneously. The ultrasonic sensor transmits a sound at 37 kHz and then waits to receive

Fig. 4 Activity chart of the
developed AV system

a corresponding echo from the sent signal. The system calculates the distance ahead
of the obstacle once the time is estimated ahead given by the Eq. (3.1).

$$D = \frac{t_{IN} \cdot V}{2} \tag{3.1}$$

where

D   Distance between the sensor and the detected object.
t_{IN} Time between transmitted and received reflected wave.
V   Ultrasonic wave propagation speed in air at normal speed 344 m/s.

   As illustrated in Fig. 5, if the distance ahead is less than or equal to 20 cm, the
controller prompts the motor to turn at 90° angle and move in the forward direction
where the path is clear. The ultrasonic sensor2 sends out its signal and if it detects
an obstacle, the motor remains in the forward direction and displays the measured
distance from the obstacle under the chassis of the car in the serial monitor of the
computer.

3.3   Data Acquisition

Data collection is possible through the use of two ultrasonic sensors which were
placed just on top of the robot car to read data from the environment and send
digital information to the microcontroller which then reads the data and carries out
the necessary instructions as designed by the users. The sensors for collecting data
include two ultrasonic sensors. Ultrasonic sensor1 and sensor2 will send the data to
the control unit (computer) of the smart car (Fig. 6).

Fig. 5 Obstacle detection
and avoidance using
ultrasonic sensor
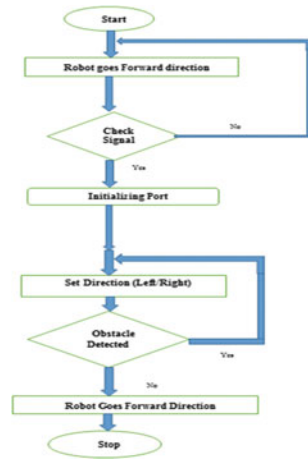
Fig. 6 Sensor data
acquisition

```
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
16
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
250
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
23
distance of ultrasonic sensor 2
100
distance of ultrasonic sensor 1
25
distance of ultrasonic sensor 2
250
distance of ultrasonic sensor 1
```

3.4 Signal Processing

The outputs from all sensors used on the board are all connected to the Arduino Nano
board. The digital data obtained from the sensors are fed to the Arduino Nano board.
After receiving the measured data from the sensors, the microcontroller is able to
make a decision with the based on the algorithm mentioned in Fig. 5. According
to the decision, the dc motors rotate and hence the car moves toward the desired
direction avoiding the obstacle.

4   Discussion and Conclusion

The developed system was tested by placing obstacle at several distances across its path. The responses of sensors were evaluated individually, since they were positioned on different part of autonomous car.

This paper presented a prototype development of a simple, cost-effective unmanned vehicle. Two ultrasonic sensors were employed to discover obstacles along the path of the car. The evaluation of the autonomous system shows that it is able to avoid obstacles, collision and has the capability to change its position. It is obvious that more functionality could be included in this design to accomplish various functions with no intervention of humans. Driverless vehicles will be a reality in the few years and will reshape the future mobility of people and goods. Currently, many research centers are working on defining the sensing suite, perception algorithms, onboard intelligence, control architectures, and communication layers that will likely constitute the basics of our future vehicles. This prototype AV will be helpful in hostile environment, defense, and security zones of the country.

References

1. Daniel Wang E (2018) Obstacle avoidance algorithms and sensors for Autonomous-Robots. www2.ece.gatech.edu/academic/courses/ece4007/10fall/ECE4007L03/…/ewang9.doc. Accessed 28 April 2018
2. Navarro-Serment L, Paredis C, Khosla P (1999) A beacon system for the localization of distributed robotic teams. In: Proceedings of the international conference on field and service robotics, pp 232–237
3. Soto A, Saptharishi M, Ollennu A, Dolan J, Khosla P (1999) Cyber-ATVS: dynamic and distributed reconnaissance and surveillance using all terrain UGVS. In: Proceedings of the international conference on field and service robotics, pp 329–334
4. Chatila R, Andrade G, Lacroix S, Mallet A (1999) Motion control for a planetary rover. In: Proceedings of the international conference on field and service robotics, pp 381–388
5. Ulrich I, Borenstein J (1998) VFH+: reliable obstacle avoidance for fast mobile robots. In: Proceedings of the IEEE conference on robotics and automation, pp 1572–1577
6. Bhagat K, Deshmukh S, Dhonde S, Ghag S (2016) Obstacle avoidance robot. Int J Sci Eng Technol Res (IJSETR) 5(2)
7. Moon HC, Lee HC, Kim JH (2006) Obstacle detecting system of unmanned ground vehicle. In: International joint conference on SICE-ICASE
8. Qingyang C, Zhenping S, Daxue L, Yugiang F, Xiaohui L (2012) Local path planning for an unmanned ground vehicle based on SVM. Int J Adv Rob Syst 9:246–255
9. Fofilos P, Xanthopoulos KI, Romanos EA, Zikidis K, Kanellopouls N (2014) An unmanned ground vehicle for remote-controlled surveillance. J Comput Model 4(1):223–236
10. Baruah JK, Bera R, Dhar S (2017) Ranking of sensors for ADAS—an MCDM based approach. In: ICCDN 2017, SMIT, Sikkim, India, June 2017
11. Iozzio C (2015) 4 driverless car features going standard-today's safety features foreshadow the robotic cars of tomorrow. Sci Amer
12. Litman T (2018) Autonomous vehicle implementation predictions-implications for transport planning. Victoria Transport Policy Institute
13. Dhar S, Ray A, Bera R (2013) Cognitive vertical handover engine for vehicular communication. Peer Peer Netw Appl (Springer) 6:305. https://doi.org/10.1007/s12083-012-0171-5

14. Alli KS, Onibonoje MO, Oluwole AS, Ogunlade MA, Mmonyi AC, Ayamolowo O, Dada SO (2018) Development of an Arduino-based obstacle avoidance robotic system for an unmanned vehicle. ARPN J Eng Appl Sci 13(3)
15. Jiang T, Petrovic S, Ayyer U, Tolani A, Husain S (2015) Self-driving cars: disruptive or incremental? Appl Innov Rev (1)

BER Analysis for Lattice Reduction Aided MIMO Receiver

Ziree Ziree Daimary and Samarendra Nath Sur

Abstract This paper demonstrates the performance of lattice reduction algorithm for large-scale MIMO. For a device that uses tens and thousands of antennas offering more signal paths and better data link reliability has a major disadvantage. There is increased complexity of both hardware and signal processing. The lattice reduction algorithm has been deemed to be the algorithm to solve receiver complexity in large-scale MIMO. The simulation result in BER versus average SNR will exhibit the comparative performance of various algorithms.

1 Introduction

Multiple-Input Multiple-Output (MIMO) is a method of enhancing the capacity of the radio link using multiple transmit and receive antennas to exploit multipath propagation. Several different diversity methods are used to achieve robustness in varying channel such as time diversity, frequency diversity and spatial diversity [1].

To enhance the system performance, MIMO exploits the precoding gain, multiplexing gain and diversity gain. Large-scale MIMO exploits these gains to fulfil the demand of the future generation communication system.

Z. Z. Daimary · S. N. Sur (✉)
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, East Sikkim 737136, India
e-mail: samarendra.s@smit.smu.edu.in; samar.sur@gmail.com

Large-scale MIMO (massive MIMO, very large MIMO, hyper-MIMO, fill-dimension MIMO and ARGOS) as the name suggests use large number of antennas. It includes all the benefits of MIMO with huge improvements in throughput and energy efficiency. And also, user tracking provides robustness against intentional jamming [1, 2]. Therefore, as large-scale MIMO become inhabitable for modern communication system, design of proper receiver algorithm is very important.

The optimal MIMO receiver uses maximum-likelihood detector (MLD) offers optimal error rate performance but it is difficult to implement in four antennas or more as the complexity increases exponentially with the increase in the number of antennas making it impossible to implement in large-scale MIMO antenna system that possibly uses antennas ranging from tens to thousands.

However, there are linear-based detectors available like Zero Forcing (ZF) or Minimum Mean Square Estimation (MMSE) or non-linear Successive Interference Cancellation (SIC) that offer manageable receiver complexity at the expense of highly suboptimal error rate performance due to the susceptible noise amplification, particularly when the channel matrix is ill-conditioned. Therefore, in this paper, lattice reduction-aided receiver has been considered, and performance has been analysed. As in [3–10], lots of lattice reduction-aided receiver architectures have been proposed for the betterment of the MIMO system. Here, in this paper, the authors have considered complex lattice reduction algorithm.

2   System Model

For a flat fading channel matrix H of the size $N_R \times N_T$

$$y = Hx + n \tag{1}$$

where n represents the $N_{R \times 1}$ Additive White Gaussian Noise (AWGN) with elements having zero mean and σ_n^2 variance. The complex valued system model can also be transformed as an equivalent real-valued system model. By separating the real and imaginary part of N_T and N_R transmit and receive antennas, respectively, for the complex baseband channel [11] leads to an equivalent real channel with $2 N_T$ real inputs and $2 N_R$ real outputs.

Thus, the channel matrix can be written as

$$H = \begin{bmatrix} R(H) & -I(H) \\ I(H) & R(H) \end{bmatrix} \tag{2}$$

where $y = \begin{bmatrix} R(y) \\ I(y) \end{bmatrix}$, $x = \begin{bmatrix} R(x) \\ I(x) \end{bmatrix}$, $n = \begin{bmatrix} R(n) \\ I(n) \end{bmatrix}$, where R and I represent the real and imaginary part, respectively, and H $R^{2N_R \times 2N_T}$, n R^{2N_R} and x R^{2N_T}.

2.1 Maximum-Likelihood Estimation

The MLD was seen that optimum detectors offered significant gain in performance over conventional symbol by symbol detectors. The ML detection problem corresponds to the simultaneous detection of several users' symbols in the presence of multiple access interference [5].

Equation for detection via maximum-likelihood detector

$$\hat{x}_{ML} = \frac{\text{arg } min}{x} \bigg| ||y - Hx||^2 \tag{3}$$

ML obtains the highest complexity and the best performance, which makes it the optimal method in case of performance.

2.2 Zero Forcing

Here, the interference caused by H is forced to zero by multiplying the Moore–Penrose pseudoinverse of the channel matrix [4, 5].

$$\tilde{x}_{ZF} = H^{\dagger} y \tag{4}$$

$$\hat{x}_{ZF} = \underset{x \in \Omega^{N_T}}{argmin} |\tilde{x}_{ZF} - x| \tag{5}$$

Here, $H^{\dagger}$ is the Moore–Penrose pseudoinverse of the channel matrix written as

$$H^{\dagger} = \left(H^H H\right)^{-1} H^H \tag{6}$$

2.3 Minimum Mean Square Estimation

$$\hat{x}_{MMSE} = (H^T H + \sigma_n^2 I_{m_{real}})^{-1} H^T x \tag{7}$$

Here, $\hat{x}_{MMSE}$ is obtained by finding out each of $\tilde{x}$ to the nearest constellation points. MMS performs better than ZF in noisy environment and has considerably low computational complexity and low performance due to the susceptible noise amplification, particularly when the channel matrix is ill-conditioned. Better performance compared with ZF, MMSE keeps good balance between complexity and BER [4, 5].

3 Lenstra–Lenstra–Lovász (LLL) Lattice Basis Reduction

The Lenstra–Lenstra–Lovász (LLL) lattice basis reduction algorithm is a polynomial time lattice reduction algorithm invented by Arjen Lenstra, Hendrik Lenstra and Laszlo Lovász in 1982 [11, 12].

The goal of the lattice basis reduction is finding a basis vectors that are short and almost orthogonal, which is known as lattice basis reduction [4]. For MIMO detection, the lattice point search approach is proven to be one of the best solutions. It is a powerful technique to build up better conditioned channel matrix by factorizing the channel matrix into the product of well-conditioned matrix and a unimodular matrix [5].

To mitigate noise enhancement and obtain similar diversity order to ML detectors, lattice basis reduction (LR) is used get the shortest and the closest vectors. For lower BER and complexity LR algorithms are used with ZF and MMSE.

The real-valued system with reduced basis is obtained as

$$\dot{y} = \tilde{H}x + n = HTT^{-1}x + n = \tilde{H}x + \dot{n} \tag{8}$$

where $\tilde{H} = HT, TT^{-1} = 1, z = T^{-1}x$

$\tilde{H}$ is the lattice-reduced channel matrix obtained from the and H and T is the unimodular matrix. The different steps in LLL algorithm include Gram–Schmidt orthogonalization, size reduction, Lovász condition and swap process.

4 LLL Algorithm Analysis

4.1 Algorithm

Algorithm 1 : LLL Algorithm

Input: H

Output: $\tilde{Q}, \tilde{R}, T$

Initialise:[Q,R]=QRD(H)

$\tilde{Q}=Q$, $\tilde{R}=R$, $T=2I_{N_T}$, k=2

1:   while $k \leq N_T$

2:     for n=k-1: -1:1

3:         μ=round($\tilde{R}(n,k)/\tilde{R}(n,n)$)

4:             if $\mu \neq 0$

5:                 $\tilde{\mathbf{R}}$(1:n,k)=$\tilde{\mathbf{R}}$(1:n,k)-$\mu\tilde{\mathbf{R}}$(1:n,n)

6:                 $\mathbf{T}$(:,k)=$\mathbf{T}$(:,k)-$\mu\mathbf{T}$(:,n)

7:             end

8:     end

9:     if $\delta\,\tilde{\mathbf{R}}$(k-1,k-1)^2 > $\tilde{\mathbf{R}}$(k,k)^2+$\tilde{\mathbf{R}}$(k-1,k)^2

10:            Swap columns k-1 and k in $\tilde{\mathbf{R}}$ and $\mathbf{T}$

11   Θ=[α', β; -β, α]   with $\alpha = \dfrac{\tilde{R}(k-1,k-1)}{\|\tilde{R}(k-1:k,k-1)\|}$

$$\beta = \frac{\tilde{R}(k-1,k-1)}{\|\tilde{R}(k-1:1,k-1)\|}$$

12            $\tilde{\mathbf{R}}$(k-1:k,k-1:N_T)=$\Theta\tilde{\mathbf{R}}$(k-1:k,k-1: N_T)

13            $\tilde{\mathbf{Q}}$(:,k-1:k)=$\tilde{\mathbf{Q}}$(:,k-1:k)Θ^{H}

14:            k= max(k-1,2)

15:     else

16:            k=k+1

17:     end

18: end

4.2   Gram–Schmidt Orthogonalization (GSO)

The main reason why Gram–Schmidt orthogonalization is widely used in LR is because it allows triangularization of the basis [10].

Let $b_1, b_2, \ldots, b_d$ be linearly independent vectors in R_n.

Then, the GSO is defined as $b_1^*, b_2^*, \ldots, b_d^*$.

Then, we get

$$b_i^* = b_i - \sum_{j=1}^{i-1} \mu_{i,j} b_j^* \tag{9}$$

where

$$\mu_{i,j} = \frac{\left\langle b_i, b_j^* \right\rangle}{\left\| b_j^* \right\|^2} \tag{10}$$

The orthonormal basis is defined as $\frac{b_1^*}{\|b_1^*\|}, \frac{b_2^*}{\|b_2^*\|}, \ldots, \frac{b_d^*}{\|b_d^*\|}$.

4.3 Size Reduction

Size reduction is performed on the basis of Gram–Schmidt coefficient where $|\mu_{i,j}|$ ≤ 0.5 holds. The size reduction algorithm represented by line 2–8.

4.4 Lovász Condition

Line 9 exhibits the Lovász condition. If the Lovász condition holds, we perform k = k+1 which can be seen in the line 16 of the algorithm meaning that we need to move on to the next column index, but if the condition fails we perform swapping and update the value of μ, or else we again perform the size reduction process for k from k − 1 to k and update μ. When k becomes greater than 2, we perform k = k − 1.

4.5 Swapping

If the Lovász condition does not hold then the swapping of the two vectors as can be seen on line 10–13 is performed so that the norm of the basis vector will not decrease too much [3]. This will continue until the Lovász condition is fulfilled after which the coefficient will be updated. Finally, the reduced basis is obtained, i.e. shorter and more orthogonal.

5 Linear Detection Combined with Lattice Reduction

ZF or MMSE detectors cannot give optimal performance if the channel matrix is ill-conditioned. Thus, LR-aided detectors are performed to obtain a more orthogonal channel matrix before detection using low complexity (suboptimal receiver) [4].

6 LR-Aided ZF Detection

In this process, the MIMO channel effect is reversed on the transmitted symbols. Also known as least square estimation the least square solution of Eq. (1) interference caused by H is forced to zero [8, 9].

Pre-filter of ZF is given by

$$Z_{ZF\_LR} = T^{-1}x_{ZF} = T^{-1}\left(x + \left(\tilde{H}^T\tilde{H}\right)^{-1}\tilde{H}^T n\right)x \tag{11}$$

Desired signal

$$\hat{x}_{ZF\_LR} = T Z_{ZF\_LR} \tag{12}$$

Here, the matrix T is a unimodular matrix. Through ZF filter we receive Z_{ZF_LR} which is rounded to the nearest constellation points.

7 LR-Aided MMSE Detection

$$E(zz^T) = E\left[T^{-1}y(T^{-1}y)^T\right] = T^{-1}E(yy^T)(T^{-1})^T \tag{13}$$

$$E(yy^T) = E\left[\left(\tilde{H}z + n\right)\left(\tilde{H}z + n\right)^T\right]$$

$$= \tilde{H}E(zz^T)\tilde{H}^T + \frac{N_0}{2}I_{N_{T\_real}}$$

$$= \tilde{H}T^{-1}(T^{-1})^T\tilde{H} + \frac{N_0}{2}I_{N_{T\_real}}$$

$$E(yy^T) = \bar{H}\bar{H}^T + \frac{N_0}{2}I_{N_{T\_real}} \tag{14}$$

Pre-filter of MMSE is given by

$$Z_{MMSE\_LR} = T^{-1}x_{MMSE} = T^{-1}\left(H^T H + \sigma_n^2 I_{N_{T_{real}}}\right)^{-1}\tilde{H}^T y$$

Table 1 BER comparison between different modulation schemes

Modulation index	SNR	BER	SNR	BER
QPSK	0.0001	0.4954	31	1.55×10^{-5}
16QAM	0.0001	0.4963	31	0.001845
64QAM	0.0001	0.4977	31	0.05413

$$Z_{MMSE\_LR} = \left(\tilde{H}^T \tilde{H} + T T^T \sigma_n^2 I_{N_{T_{real}}} \right)^{-1} \tilde{H}^T y \qquad (15)$$

Desired signal

$$\hat{x}_{MMSE\_LR} = T Z_{MMSE\_LR}$$

$$\hat{x}_{MMSE\_LR} = T \left(\tilde{H}^T \tilde{H} + T T^T \sigma_n^2 \right)^{-1} \tilde{H}^T y \qquad (16)$$

Z_{MMSE_LR} must be decoded to get the exact sent symbols and bits for which demodulation is done and the estimation received bits $\hat{x}_{MMSE-LR}$ [8, 9].

The estimation $\hat{x}$ is obtained by independently rounding each of the $\hat{x}$ to the nearest constellation points. MMSE performs better in noisy environment than ZF as it balances the reduction of interference instead of removing or reversing it completely.

8 Simulated Result

Here, in this paper all the results have been taken based on the MATLAB simulation. Here, 10×10 MIMO system with Nakagami-m channel has been considered for the simulation. The Nakagami-m fading model will give a diversity order of m. Although a number of channel models exist to describe amplitude and phase of multipath propagation, Nakagami has certain advantages. It is dynamic in nature and includes different distributions as special cases and is equally applicable to any of the fading environment, i.e. urban, suburban, etc. Basically for m > 1, it exhibits half Gaussian, for m = 1 Rayleigh, for m < 1 Rician and for very large values of m it exhibits full Gaussian distribution.

Figure 1 and Table 1 show the BER comparison for LLR-ZF-aided MIMO system with different modulation schemes. For the simulation purpose, 10×10 MIMO system with Nakagami-m channel has been considered with m = 1. The QPSK modulation exhibits the best performance with 64 QAM as worst throughout the BER versus SNR performance.

Figure 2 and Table 2 show the BER comparison for MIMO system with different receiver systems. For the simulation purpose, 10×10 MIMO system with Nakagami-m channel has been considered with m = 1. As indicated in the figure, LR-aided
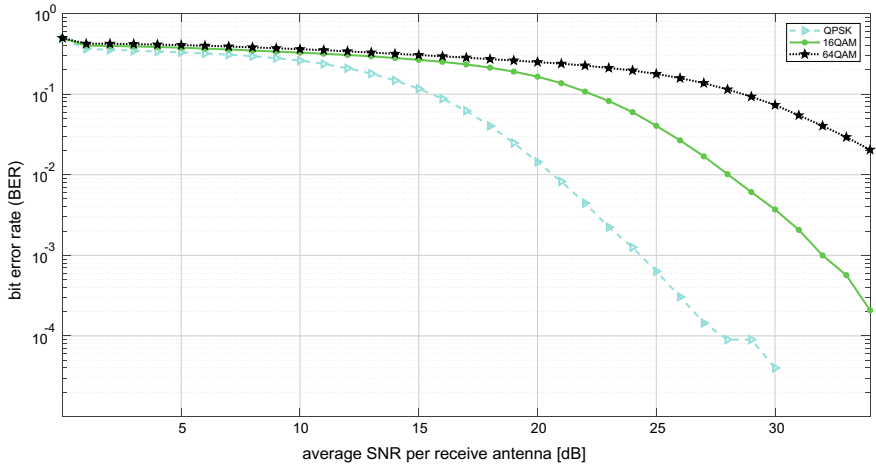
Fig. 1 BER comparison between different modulation schemes
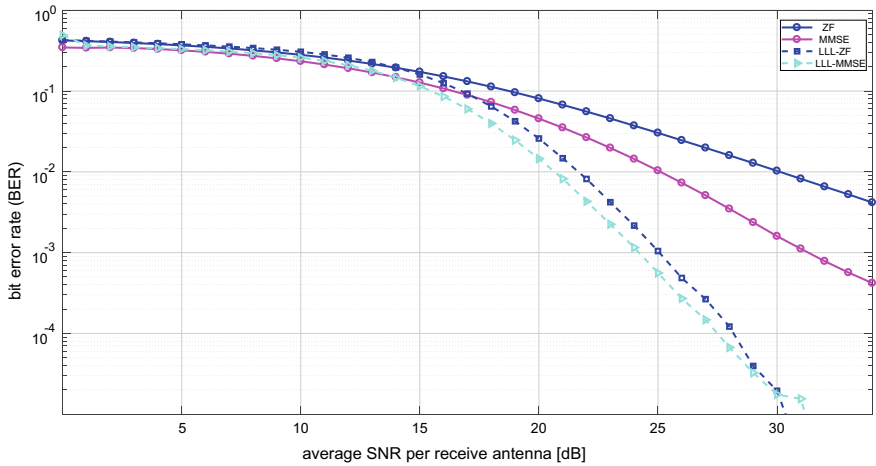
Fig. 2 BER comparison between different detectors

MMSE provides the best performance in comparison to other receiver (Tables 3 and 4).

As in Figs. 3 and 4, MIMO system performance improves with the increase in m values. This is very true, as increasing m value improves the channel condition.

Table 2 BER comparison between different detectors

Receiver	SNR	BER	SNR	BER
ZF	0.0001	0.4236	30	0.01048
MMSE	0.0001	0.3461	30	0.001597
LR-aided ZF	0.0001	0.4267	30	1.95×10^{-5}
LR-aided MMSE	0.0001	0.4923	30	1.75×10^{-5}

Table 3 BER comparison of LLR-MMSE-aided MIMO system with different channel conditions

m	SNR	BER	SNR	BER
0.5	0.0001	0.4847	27	3.77×10^{-5}
1	0.0001	0.4923	27	1.48×10^{-4}
5	0.0001	0.4961	27	1.67×10^{-5}

Table 4 BER comparison of LLR-ZF-aided MIMO system with different channel conditions

m	SNR	BER	SNR	BER
0.5	0.0001	0.4858	34	0.004608
1	0.0001	0.4858	34	0.002514
5	0.0001	0.4765	34	0.001571
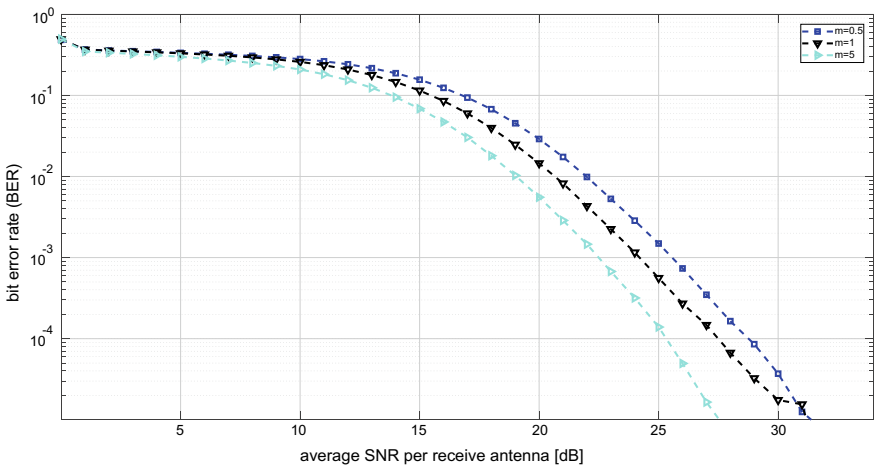
Fig. 3 BER comparison of LLR-MMSE-aided MIMO system with different channel conditions
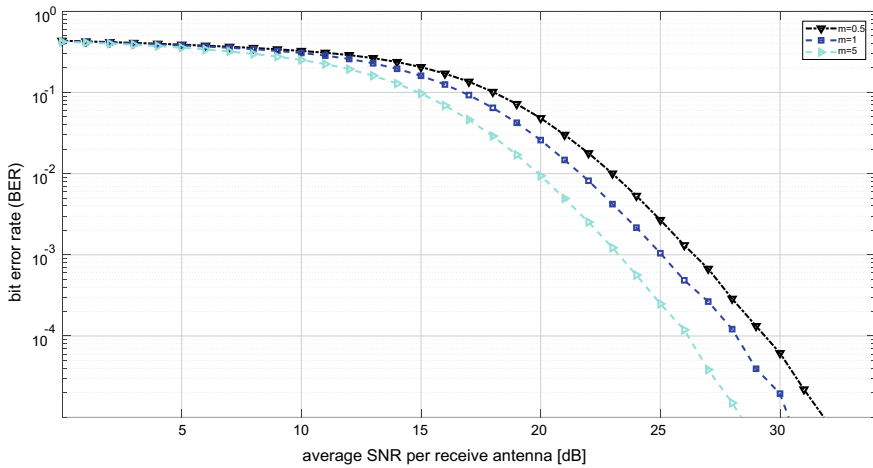
Fig. 4 BER comparison of LLR-ZF-aided MIMO system with different channel conditions

9 Conclusion

In this paper, authors have considered Nakagami channel with 10×10 antennas to represent massive MIMO. Nakagami channel is a dynamic channel meaning it gives different characteristics for different values of m. In LLL algorithm, LR-aided MMSE's BER performance is superior in comparison to all other kinds of methods, including ZF, MMSE and LR-aided ZF. The ZF continues to exhibit the worst BER performance of all time. And also higher the modulation index gives the worst BER performance, i.e. observed in both LLL algorithms.

References

1. Gupta A, Jha RK (2015) A survey of 5G network: architecture and emerging technologies. IEEE Access 3:1206–1232
2. https://resources.altium.com/pcb-design-blog/advantages-of-massive-mimo-for-5g-cellular-infrastructure-2
3. Wang L (2016) A study of Lattice reduction detection techniques for LTE systems. MS thesis, Dept. Elect. Inst. Tech., LTH, LU
4. Murray AT, Weller SR (2013) A digital signal processing architecture for soft-output MIMO Lattice reduction aided detection. In: Design and architectures for digital signal processing, pp 231–258
5. Yao H, Wornell GW (2002) Lattice-reduction-aided detectors for MIMO communication systems. In: Global telecommunications conference, pp 424–428
6. Lenstra AK et al (1982) Factoring polynomials with rational coefficients. Math Ann 261:515–534
7. An HS et al (2011) Lattice reduction aided precoding for multiuser MIMO using Seysen's algorithm. Inha University

8. Kim H et al (2017) Low-complexity lattice reduction algorithm for MIMO detectors with tree Searching. EURASIP J Wirel Commun Netw. https://doi.org/10.1186/s13638-016-0756-4

9. Milliner DL, Barry JR (2006) A Lattice-reduction-aided soft detector for multiple-input multiple-output channels. In: IEEE GLOBECOM 2006, Dept. ECE, GIT, Ga.,USA

10. Nguyen PQ, Vallée B (2009) The LLL algorithm: survey and applications. Springer Science & Business Media

11. https://en.m.wikipedia.org/wiki/MIMO

12. https://crypto.stackexchange.com/questions/39532/why-is-thelov%C3%A1sz-condition-used-in-the-lll-algorithm

Sparse Coding of Bottleneck Features-Based i-Vector Representation for Language Recognition

Om Prakash Singh and Rohit Sinha

Abstract In recent years, the deep neural network (DNN) and the sparse coding paradigms have received a lot of attention in many pattern recognition domains including language recognition (LR). Motivated by their success, we have explored sparse coding of DNN-based i-vectors for LR task in this work. For i-vector representation, DNN-based bottleneck (BN) front-end features are used unlike the conventional mel frequency cepstral coefficient (MFCC) features. Different kinds of KSVD-based dictionary learning approaches have been explored for LR task. The effectiveness of the proposed approaches has been evaluated on AP17-oriental language recognition datasets in closed-set test conditions. On comparing with the baseline method (BN feature-based i-vector with cosine distance scoring), the proposed LR approaches are found to yield competitive performances.

Keywords Language recognition · Deep neural network · Bottleneck features · Sparse coding · Learned dictionary

1 Introduction

The success of deep neural networks (DNNs) in automatic speech recognition (ASR) [5] has inspired many researchers to explore the DNNs in other pattern recognition domains. In the recent past, DNNs have also been successfully explored in speaker recognition (SR) [8, 9, 13] and language recognition (LR) [10, 11, 13, 17]. The DNNs have been used for classification task as well as extracting the bottleneck features (BNFs). The state-of-the-art LR approach is based on the i-vector representation using BNFs [13].

O. P. Singh (✉) · R. Sinha
Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati, India
e-mail: o.singh@iitg.ac.in; opsing@gmail.com

R. Sinha
e-mail: rsinha@iitg.ac.in

337

In the recent past, the sparse coding is another popular approach which has got lot of attention in many signal processing areas including language recognition [4, 14–16]. The sparse coding technique relies on the underlying assumption that a signal can have a compact representation as a linear combination of few columns (atoms) of an overcomplete dictionary. In an ideal condition, the selected dictionary atoms belong to the same class label as that of the target signal.

Motivated by the success of DNNs and sparse coding, we explore the sparse coding of the i-vector representation of utterances on BNFs for LR task. The KSVD [1] learned-exemplar, discriminative KSVD (D-KSVD) [20], and label consistent-KSVD dictionary (LC-KSVD) [6] based dictionaries have been created and used for sparse coding.

2 Bottleneck Feature-Based i-Vector for LR

In this section, we first present the structure of DNN employed for extracting the features. The frame level outputs extracted from the bottleneck layer corresponding to input acoustic features are referred to as the BNFs. The BNFs are used as the front features for conventional i-vector representation employing Gaussian posteriors. An alternate approach to extracting statistics in i-vector framework is also described, which is based on posteriors from DNN instead of Gaussian mixture model (GMM) component posteriors.

2.1 *DNN Bottleneck Feature*

The BNFs can be generated from one of the hidden layer which is having less number of units or nodes as compared to hidden layer on either sides. The BN layer (narrowest hidden layer) produces the output which is a compact version of original inputs as it forces to represent information in low dimension. Figure 1 shows the structure of deep bottleneck network with one input layer, multiple hidden layers, and an output layer. Among the hidden layers, the layer having less number of nodes is known as bottleneck layer. In this work, $\pm N$ frames (left and right context $= N$) of spectral features (MFCCs) are spliced together, and the dimensionality is reduced by linear discriminant analysis (LDA) using the acoustic states as classes. The reduced dimensional features are further spliced together with $\pm M$ frames (left and right context $= M$) and applied to the input layer of DNN. The BNFs are extracted from a DNN trained for ASR using Kaldi [12] framework. In ASR, the pronunciations of all words are represented by a sequence of senones (e.g., the tied-triphone states), trained on top of LDA followed by maximum likelihood linear transform (MLLT) features using particular transcribed speech corpus.
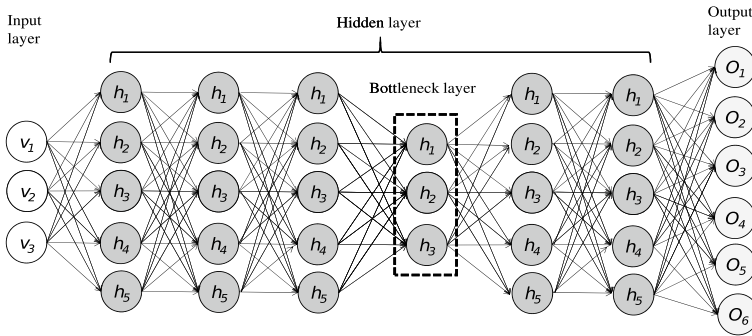
Fig. 1 Structure of deep neural network employed for extracting the bottleneck features

2.2 i-Vector Representation

The i-vector [3] is a low-dimensional representation of GMM mean supervectors derived using factor analysis with a low-rank projection matrix called the total variability matrix. It captures both the language and the session/channel variabilities, and thus, appropriate session/channel compensation is required to minimize the session/channel effect. In this approach, language-specific GMM supervectors are modeled as

$$s = m + Tw \tag{1}$$

where m is the language-independent universal background model (UBM) based mean supervector, T is the total variability matrix, and w is known as the identity vector or i-vector.

2.3 GMM Posteriors

The GMM posterior is the per-mixture posterior probability of each feature vector computed using UBM, which is essentially a speaker/language-independent GMM. The GMM posteriors along with utterance-specific feature vectors are used to accumulate zeroth-, first-, and second-order sufficient statistics. Finally, the i-vector modeling is done using accumulated statistics.

2.4 DNN Posteriors

In [9], an alternate approach to extracting statistics based on DNN has been proposed. The method uses class posteriors from the DNN instead of GMM component

posteriors for accumulating the statistics. The i-vector extraction is done in the same way as it is done in the case of GMM-based statistics.

3   Proposed LR Approaches

3.1   Learned-Exemplar Dictionary-Based LR Approach

Given the training data for the lth language containing n^l numbers of m-dimensional i-vectors $\boldsymbol{Y}^l = \{\boldsymbol{y}_i^l\}_{i=1}^{n^l}$ and the l_1 regularization coefficient λ_1' for sparsity constraint, the class-based dictionary learning problem is formulated as

$$\min_{\boldsymbol{D}^l, \boldsymbol{S}^l} \parallel \boldsymbol{Y}^l - \boldsymbol{D}^l \boldsymbol{S}^l \parallel_F^2 + \lambda_1' \parallel \boldsymbol{S}^l \parallel_1 \tag{2}$$

where $\boldsymbol{D}^l$ is the language-specific learned dictionary having k columns and $l = 1, \ldots, L$. The matrix $\boldsymbol{S}^l$ denotes a set of sparse vectors corresponding to $\boldsymbol{Y}^l$ with $\boldsymbol{s}_i^l$ being the sparse vector for the ith i-vector of the lth language. All language-specific dictionaries are estimated and then combined to form a single learned-exemplar dictionary as

$$\boldsymbol{D} = [\boldsymbol{D}^1 \mid \boldsymbol{D}^2 \mid \ldots \mid \boldsymbol{D}^L] \tag{3}$$

where "|" denotes a horizontal concatenation operator.

Once the dictionary $\boldsymbol{D}$ is obtained, the test i-vector $\boldsymbol{x}$ for an utterance is sparse coded using the least angle regression (LAR) algorithm. The Lasso-based sparse coding problem is formulated as

$$\hat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \frac{1}{2} \parallel \boldsymbol{D}\boldsymbol{\gamma} - \boldsymbol{x} \parallel_2 + \lambda_1 \parallel \boldsymbol{\gamma} \parallel_1 \tag{4}$$

The coefficients of sparse vector $\hat{\boldsymbol{\gamma}}$ can be used to obtain the class labels of the unknown utterance, and known as sparse representation-based classification (SRC).

3.2   Language Recognition Using D-KSVD

The D-KSVD [20] is a dictionary learning approach which incorporates classification error term in addition to reconstruction error used in classical KSVD. The idea is to simultaneously learn a single dictionary $\hat{\boldsymbol{D}}$ and a linear classifier $\hat{\boldsymbol{W}}$ by solving the joint optimization problem which is formulated as

$$\langle \hat{\boldsymbol{D}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{S}} \rangle = \arg\min_{\boldsymbol{D}, \boldsymbol{W}, \boldsymbol{S}} \parallel \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{S} \parallel_F^2 + \beta \parallel \boldsymbol{H} - \boldsymbol{W}\boldsymbol{S} \parallel_F^2 + \lambda_1' \parallel \boldsymbol{S} \parallel_1 \tag{5}$$

where $Y \in \mathbb{R}^{m \times N}$ is the set of training data vectors, $D \in \mathbb{R}^{m \times k}$ is the learned dictionary, $S \in \mathbb{R}^{k \times N}$ is the set of sparse codes corresponding to Y, β is the regularization parameter, the matrix $H \in \mathbb{R}^{L \times N}$ contains the class label of training data vectors, $W \in \mathbb{R}^{L \times k}$ is a linear classifier, and λ_1' is the l_1-norm regularization coefficient which controls the number of nonzero elements in the sparse vector.

3.3 Language Recognition Using LC-KSVD

In LC-KSVD [6] dictionary learning approach, discriminative sparse code error term enforcing label consistency was incorporated in addition to the terms (i.e., reconstruction and classification error term) used in D-KSVD formulation. The LC-KSVD dictionary learning problem is formulated as

$$\langle \hat{D}, \hat{W}, \hat{A}, \hat{S} \rangle = \underset{D,W,A,S}{\arg\min} \parallel Y - DS \parallel_F^2 + \alpha \parallel Q - AS \parallel_F^2$$
$$+ \beta \parallel H - WS \parallel_F^2 + \lambda_1' \parallel S \parallel_1 \tag{6}$$

where $Y \in \mathbb{R}^{m \times N}$ is the set of training data vectors, $D \in \mathbb{R}^{m \times k}$ is the learned dictionary, $S \in \mathbb{R}^{k \times N}$ is the set of sparse codes corresponding to Y, matrix $H \in \mathbb{R}^{L \times N}$ contains the class label of training data vectors, and $W \in \mathbb{R}^{L \times k}$ is a linear classifier. The matrix $Q \in \mathbb{R}^{k \times N}$ is the discriminative sparse codes matrix promoting label consistency, $A \in \mathbb{R}^{k \times k}$ is a linear transformation, α and β are the regularization parameters used to balance the discriminative sparse code errors and classification error to overall objective function, respectively.

Equations (5) and (6) can be solved using classical KSVD algorithm. Once the dictionary D is obtained, the sparse coefficient vector s' corresponding to the test vector can be computed by using any suitable sparse coding algorithm such as LAR. Now, on applying the predicted linear classifier $\hat{W}$ to the sparse vector s', we obtain a class-similarity vector l as $l = \hat{W}s'$. Finally, the class label of the test vector is decided based on the class label associated with the index in vector l attaining the maximum value.

The special case of LC-KSVD problem with $\beta = 0$ is referred to as LC-KSVD1 while with nonzero values of regularization parameters α and β, it is referred to as LC-KSVD2. The learning procedure of LC-KSVD1 is same as of LC-KSVD2; however, the classifier W for LC-KSVD1 is trained separately after the computation of D, A, S. The estimate of linear classifier $\hat{W}$ is obtained by solving the equation

$$\hat{W} = H\hat{S}(\hat{S}\hat{S}^T + \lambda I)^{-1} \tag{7}$$

where λ is the l_2 norm regularization coefficient and I is the identity matrix. The dictionary initialization step of LC-KSVD differs from D-KSVD. The LC-KSVD learns a single learned-exemplar dictionary which is concatenation of class-specific dictionaries whereas D-KSVD learns a single dictionary using pooled data from all

classes. In our implementation, a single learned-exemplar dictionary-based initialization is considered in both D-KSVD- and LC-KSVD-based approaches. This has been done to know the actual performance gains of the algorithms.

4 Experimental Setup

4.1 Database

In this work, THCHS30 Chinese transcribed speech database [19] is used for training the ASR-based DNN system. The BNFs corresponding to training and test utterances of multilingual speech corpus for oriental languages, collected from AP17-OL3 and AP16-OL7 as per AP17-OLR Challenge [18], are extracted for this study. The AP17-OL3 includes three languages: Kazakh, Tibetan, and Uyghur whereas the AP16-OL7 includes seven languages: Cantonese, Mandarin, Indonesian, Japanese, Russian, Korean, and Vietnamese. Thus, a total of 10 languages are considered and the evaluations are done in closed-set condition on three different duration test datasets corresponding to 1 second, 3 seconds, and full-length utterances. The details of the data distribution can be found in [18].

4.2 Bottleneck Features

For extracting the BNFs, a DNN with five hidden layers and one BN layer has been trained using THCHS30 Chinese speech corpus. The hidden layers consist of 1024 nodes, and the BN layer has 60 nodes. The network uses the *tanh* as the nonlinearity function. The BN layer is placed after the third hidden layer. The DNN is trained with keeping the number of epochs as 20 and the mini-batch size as 128. The 91-dimensional feature vectors are obtained by concatenating 13-dimensional MFCCs spliced across ± 3 frames. The dimensionality of the resulting feature vectors is reduced to 40 by applying LDA. These 40-dimensional feature vectors are further spliced across ± 10 frames and used for training the DNN. Once the DNN is trained, the speech data for which the BNFs are to be extracted undergo the similar feature processing and then passed through this trained network. So the obtained 60-dimensional features at the output of BN layer are referred to as the BNFs.

4.3 The i-Vector Extraction

We build a language-independent UBM consisting of 2048 Gaussian mixtures on a set of BNFs pooled from each of the 10 languages of AP17-OLR. For computing the i-vectors, a total variability matrix T of 400 columns is trained on the

training data following the procedure as described in [7]. Both GMM posteriors- and DNN posteriors-based approaches are used for computing statistics corresponding to each of the training and test segments. In this work, we referred *BNF–GMM Posteriors–ivector* for BNFs-based i-vector representation of utterances employing GMM posteriors while BNFs-based i-vector representation of utterances employing DNN posteriors is referred to as *BNF-DNN Posteriors–ivector*.

4.4 Classification and Score Calibration

The CDS classifier is used for the baseline i-vector system employing front-end bottleneck features. The learned-exemplar dictionary-based system uses SRC. In SRC, the coefficients of nonzero locations are used to determine the true class of the test sample. However, the proposed D-KSVD and LC-KSVD use the weighted similarity l of the test vector as discussed in Sects. 3.2 and 3.3. The language scores computed using different classifiers may not be well calibrated. This may be due to limited number of training examples used for target language modeling. To calibrate the language score, multi-class logistic regression employing the FoCal toolkit [2] is used prior to the final decision. The calibration was trained in the "cheating" way, i.e., using the evaluation scores [10].

5 Results and Discussion

The performance of the proposed and contrast systems is reported on an AP17-OLR dataset in closed-set condition for three different test durations. The LR is broadly divided into two tasks: the language detection and language identification, and their performances are measured in terms of average cost detection function (C_{avg}) and identification rate (% IDR), respectively. Table 1 shows the performance

Table 1 Performance of BNF-DNN Posteriors–ivector with session/channel compensation techniques. The scores are obtained using CDS classifier and are calibrated using Gaussian backend followed by MLR

Session/Channel comp.	Full-length		3 s		1 s	
	$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR
LDA	1.38	96.12	2.28	93.76	8.52	77.73
WCCN	**1.33**	**96.26**	**2.17**	**94.12**	**8.30**	**78.11**
LDA+WCCN	1.38	96.08	2.31	93.58	8.58	77.22

Table 2 Performance of BNF–GMM Posteriors–ivector with session/channel compensation techniques. The scores are obtained using CDS classifier and are calibrated using Gaussian backend followed by MLR

Session/Channel comp.	Full-length		3 s		1 s	
	$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR
LDA	1.07	**96.84**	1.86	94.91	7.27	80.88
WCCN	**1.03**	96.80	**1.77**	**94.92**	**7.06**	**81.09**
LDA+WCCN	1.10	96.80	1.88	94.81	7.41	80.84

of BNFs-based i-vector system with three session/channel compensation techniques. The accumulation of statistics for i-vector modeling is based on class posteriors from the DNN. The LR system employing WCCN session/channel compensation is found to give satisfactory performance compared to the LDA and LDA followed by WCCN. The similar trend is noted with i-vector modeling based on GMM posteriors, shown in Table 2. In Table 3, we have shown the performances of the proposed learned dictionary-based LR approaches employing learned-exemplar, D-KSVD, and LC-KSVD along with state-of-the-art *BNF–GMM Posteriors–ivector*-based LR for contrast purpose. The *BNF–DNN Posteriors–ivector*-based LR system is also taken into consideration. The degraded performance of *BNF–DNN Posteriors–ivector*-based LR system is noted compared to *BNF–GMM Posteriors–ivector*-based system. On comparing with the baseline method, the LC-KSVD2 dictionary-based LR approach is found to give satisfactory performance, whereas a degraded performance is observed in case of learned-exemplar-based LR system. The performance of the LC-KSVD1-based approach is comparable to baseline on full-length test segments. However, little improvement in the system performances is noted on shorter duration segments of 3 seconds and 1 second. In dictionary learning approaches, the parameters are tuned and best values are selected. The dictionary sparsity λ'_1 is varied for values from 0.005, 0.01, 0.015, 0.02, to 0.1 with increment of 0.01; decomposition sparsity λ_1 is varied for the values of 0.001, 0.005, 0.01, 0.015, and $0.02; \sqrt{\alpha}$ and $\sqrt{\beta}$ vary between 0.1 and 0.5 with increment of 0.1. The best performances with LC-KSVD2 on full-length segments are noted with following tuning parameters: dictionary sparsity $\lambda'_1 = 0.05$, decomposition sparsity $\lambda'_1 = 0.01$, $\sqrt{\alpha} = 0.4$ and $\sqrt{\beta} = 0.4$. Similarly, for the test segment duration of 3 seconds, the tuning parameters ($\lambda'_1 = 0.08$, $\lambda'_1 = 0.01$, $\sqrt{\alpha} = 0.5$ and $\sqrt{\beta} = 0.4$) are selected to achieve better performances. On 1-second test segments, the best tuning parameters noted are $\lambda'_1 = 0.005$, $\lambda'_1 = 0.001$, $\sqrt{\alpha} = 0.1$, and $\sqrt{\beta} = 0.1$. The performances of LC-KSVD2 and D-KSVD-based LR approaches are found comparatively.

Table 3 Performances of proposed LR systems employing i-vector utterance representation based on bottleneck features. The sparse coding employing l_1 norm regularization with three different learned dictionaries: learned-exemplar, D-KSVD, and LC-KSVD are considered. The scores are calibrated using Gaussian backend followed by MLR

Utterance representation	Dictionary	Classifier	Full-length		3 s		1 s	
			$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR	$100 \times C_{avg}$	% IDR
Contrast LR								
BNF–GMM Posteriors– ivector	–	CDS	1.03	96.80	1.77	94.92	7.06	**81.09**
BNF–DNN Posteriors– ivector			1.33	96.26	2.17	94.12	8.30	78.11
Proposed LR								
BNF–GMM Posteriors– ivector	learned-exemplar	SRC	1.25	96.29	1.77	94.95	7.04	80.67
	D-KSVD	$I = \hat{W}_{S'}$	**0.96**	97.08	**1.72**	94.99	7.02	81.06
	LC-KSVD1		1.03	96.78	1.73	95.00	**6.99**	81.06
	LC-KSVD2		**0.96**	**97.16**	**1.72**	**95.14**	7.03	81.00

6 Conclusion

In this work, we have explored three learned dictionaries (learned-exemplar, D-KSVD, and LC-KSVD) based LR systems employing i-vector representation based on bottleneck front-end features. The proposed D-KSVD- and LC-KSVD-based LR systems have been found to perform much better than i-vector employing bottleneck features with CDS classifier, except for a short segment of 1 second for language identification task. However, the slight degradation has been noted with learned-exemplar dictionary-based approach. The performances of the D-KSVD and LC-KSVD have been found comparable, considering the same dictionary initialization procedure.

References

1. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
2. Brummer N (2007) Focal multi-class: toolkit for evaluation, fusion and calibration of multi-class recognition scores-tutorial and user manual. http://sites.google.com/site/nikobrummer/focalmulticlass
3. Dehak N, Kenny P, Dehak R, Dumouchel P, Ouellet P (2011) Front-end factor analysis for speaker verification. IEEE Trans Audio Speech Lang Process **19**(4), 788–798 (2011)
4. Gwon YL, Campbell WM, Sturim DE, Kung H (2016) Language recognition via sparse coding. In: Proceedings of the Interspeech, pp 2920–2924
5. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 29(6):82–97
6. Jiang Z, Lin Z, Davis LS (2011) Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: Proceedings of the CVPR, pp 1697–1704
7. Kenny P, Boulianne G, Dumouchel P (2005) Eigenvoice modeling with sparse training data. IEEE Trans Speech Audio Process 13(3):345–354
8. Kenny P, Gupta V, Stafylakis T, Ouellet P, Alam J (2014) Deep neural networks for extracting Baum-Welch statistics for speaker recognition. In: Proceedings of the Odyssey, pp 293–298
9. Lei Y, Scheffer N, Ferrer L, McLaren M (2014) A novel scheme for speaker recognition using a phonetically-aware deep neural network. In: Proceedings of the ICASSP, pp 1695–1699
10. Lopez-Moreno I, Gonzalez-Dominguez J, Plchot O, Martinez D, Gonzalez-Rodriguez J, Moreno P (2014) Automatic language identification using deep neural networks. In: Proceedings of the ICASSP, pp 5337–5341
11. Matejka P, Zhang L, Ng T, Mallidi HS, Glembek O, Ma J, Zhang B (2014) Neural network bottleneck features for language identification. In: Proceedings of the Odyssey, pp 299–304
12. Povey D, Ghoshal A, Boulianne G, Burget L, Glembek O, Goel N, Hannemann M, Motlicek P, Qian Y, Schwarz P, Silovsky J, Stemmer G, Veselothers K (2011) The Kaldi speech recognition toolkit. In: Proceedings of the ASRU
13. Richardson F, Reynolds D, Dehak N (2015) Deep neural network approaches to speaker and language recognition. IEEE Signal Process Lett 22(10):1671–1675
14. Singh OP, Haris BC, Sinha R (2013) Language identification using sparse representation: a comparison between GMM supervector and i-vector based approaches. In: Proceedings of the INDICON, pp 1–4
15. Singh OP, Sinha R (2017) Sparse representation classification over discriminatively learned dictionary for language recognition. In: Proceedings of the TENCON, pp 2632–2636

16. Song Y, Hong X, Jiang B, Cui R, McLoughlin I, Dai LR (2015) Deep bottleneck network based i-vector representation for language identification. In: Proceedings of the Interspeech
17. Song Y, Jiang B, Bao Y, Wei S, Dai LR (2013) I-vector representation based on bottleneck features for language identification. Electron Lett 49(24):1569–1570
18. Tang Z, Wang D, Chen Y, Chen Q (2017) AP17-OLR Challenge: data, plan, and baseline. http://arxiv.org/abs/1706.09742
19. Wang D, Zhang X (2015) THCHS-30: a free Chinese speech corpus. http://arxiv.org/abs/1512.01882
20. Zhang Q, Li B (2010) Discriminative K-SVD for dictionary learning in face recognition. In: Proceedings of the CVPR, pp 2691–2698

Review on Indoor Channel Characterization for Future Generation Wireless Communications

Soumyasree Bera and Subir Kumar Sarkar

Abstract The market demand for high quality-of-service Internet/data access in indoors is increasing drastically because of the increase in the smart device users, thus leading to the requirement of high data rate support. The current technologies serve mainly in the outdoor scenarios where the presence of clutter, multipath is comparatively lesser than in indoor. Therefore, huge challenge is to satisfy users demand even in indoors, thus fulfilling the motto "anywhere, anytime" which is only possible by properly studying the nature of the propagation channel so as to mitigate the channel effect to the maximum extent with the adaptation of proper equalization technique. The aim of this paper is to access the various possibilities adapted worldwide to achieve the above goal.

Keywords Wideband system · Millimeter wave (mmW) · Future generation communication · Indoor EM propagation

1 Introduction

Demand of wireless communication system is increasing at a very faster rate due to the advancement of smart devices like mobile phones, tablets, etc., and its usage in indoors (like home, office, classrooms, etc.) is expected to be around 90% in the coming year [1, 2]. Users worldwide are interested in applications like high-quality video chats, UHD quality video streaming, on-demand HD video, instantaneous cloud access for IOT-based M2M applications, etc., which leads to the requirement

S. Bera (✉)
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: soumyasree.bera@gmail.com

S. K. Sarkar
Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: sksarkar@etce.jdvu.ac.in; su_sircir@yahoo.co.in

of very high throughput links in household scenarios and office scenarios [3]. Additionally, demand for always on data and connectivity to the cloud while not being tethered to Ethernet is also rising day by day. Currently, all indoor needs are served by the traditional macrocells, thus having poor coverage due to low penetration and low bandwidth leading to the dissatisfaction of users. Therefore, in order to improve the coverage, the concept of femto cell or microcell is one of the solutions taken into consideration. But the linear improvement of performance is not achievable as expected due to the intercell interference. Hence, currently, researchers are pondering the addition of the millimeter wave (mmW) bands to smart devices because of its much larger bandwidths (1.3 GHz at 28 GHz band, 1.4 GHz at 39 GHz band and over 7 GHz at 60 GHz band) thus fulfilling the demand of users [2]. Also, the beamforming technique further helps by reducing the intercell interference. Though mmW proves to be a very lucrative approach, it poses lots of challenges in terms of the cost and propagation phenomenon in indoor scenarios unlike the low-frequency applications [4]. Hence, channel characterization (electromagnetic propagation study) is highly required so as to mitigate the losses incurred by the presence of various obstacles in indoors [5–7]. This paper deals with the survey of the challenges observed and probable solutions of indoor wireless communication and thus evolving different techniques used worldwide to characterize indoor channel and the respective limitations.

2   Environments

Indoor propagation channel is a scenario that contains various obstacles like walls, almirahs, racks, tables, chairs, etc. Therefore, transceivers are surrounded by lots of multipaths and clutter signals. So, along with line of sight (LOS) signals, strong multipath and clutter signals make the system complex. Also, there is a high possibility of the absence of LOS signal, rather communication to be established with non-line of sight signal components [8]. The obstacles present contribute differently to the EM propagation and thus to be accounted for better channel sounding. So, type of indoor can be broadly categorized into home, office, airports, stadiums, classrooms, railway stations, etc., and as can be observed that each environment has its typical arrangement and behavior toward EM propagation [9, 10]. Therefore, various parameters like coverage area, leakage, channel capacity, interference, etc., need to be addressed.

As established earlier that the solution for high QOS is the use of the mmW. But mmW has various limitations as well. It is already known that the path loss, penetration loss for high-frequency signal is higher compared to the low-frequency applications [11]. Another factor to be taken into consideration is the oxygen absorption of high-frequency signals like at 60 GHz. Even foliage loss in case of outdoor-to-indoor communication at 28 GHz is nearly 16–28 dB depending on the condition of foliage [12, 13]. At 73 GHz band, the foliage loss of 0.4 dB/m is obtained by [14].

Even the scenario changes from city-to-city, room-to-room. Thus, the accuracy of the same model becomes questionable from one scenario to another. So, researchers are considering separately study as variable condition as possible to come up with more accurate channel models.

3 Indoor Technologies

Indoor environments mainly use the following wireless technologies:

(a) **Wi-Fi**: Wi-Fi is a well-known technology used worldwide by many users especially in indoors and the need of high-speed Internet via Wi-Fi is highly increasing. Earlier standards 802.11 a/b/g/n supports maximum data rate of 54–600 Mbps which again is limited and the intercell interference is also high. 802.11ac now promises an increased data rate of 1Gbps at 5 GHz band with the addition of multiuser MIMO (MU-MIMO). 802.11ad further improves the situation exploiting a wider bandwidth (up to 2 GHz) at 60 GHz promising typically 5 Gbps of data rate by exploiting 32 antennas used for narrow beamformation [9–11, 15–18].

(b) **Cellular**: Cellular technology has brought about a drastic change in communication system bringing people around the world closer. The advent of cellular technology starting from GSM to LTE is gigantic [15]. The increase in data rate is nearly about 1 Gbps in LTE. But the limitation is the use of macro-cell structure which makes it incompatible for indoor wireless communication. Therefore, the world is looking forward to address this problem in 5G, where techniques like MIMO, beamforming, and mmW have huge contribution [12, 13]. As already discussed, the use of mmW will lead to the usage of multiple numbers of antennas leading to high gain and less prone to interference thus reducing the shortcomings of mmW. 5G is the union of all the known technologies [19]. Even services like Internet-of-things (IoT) and Internet-of-vehicles (IoV) are also becoming part of 5G service termed as narrowband IoT (NB-IoT) [20]. Though 5G is not yet standardized, Verizon has already launched pre-5G in 28 GHz band in the US. Even Japan is also planning to launch 5G in the year 2020 during the Olympics. Therefore, it is clearly understandable that the need for larger bandwidth as in future humans will not only be the sole user of wireless network but the demand of machines will be addressed. Therefore, frequencies like 60 GHz and 73 GHz will be more likely to be used as the available bandwidth is 7 GHz and 2 GHz, respectively [21–24].

4 Radio Propagation

EM propagation undergoes various types of losses while traveling via the channel which ultimately impacts the health of signal received by the receiver. The different propagation mechanisms are reflection, refraction, diffraction, and scattering [25–27]. So the propagation is affected by the type of material of the obstacles, their penetration losses etc. Therefore, the path losses will vary accordingly depending on the type of objects present in the room. Even the penetration loss is to be considered for outdoor-to-indoor communication which leads to shadowing effect [10, 11, 15]. As discussed earlier, the impact is more severe in case of closed environments due to the presence of different obstacles like wall, floor, furniture etc. thus fast fading is a common problem faced. The papers show the effect on EM wave with the presence of [23]. These behaviors are again frequency dependent. The phenomenon like scattering and diffraction starts changing when one shifts from low frequency to high frequency, i.e., to the mmW band. At high frequency, absorption is another issue to be related. Frequencies at 60 GHz have got high oxygen absorption, which restricts its performance to a shorter distance. Therefore, while channel characterization, the above-discussed losses are to be taken into consideration.

Researchers have also taken into consideration, the effect in penetration loss depending on the antenna polarization and found an increment attenuation by 18 dB at 73 GHz [10].

5 Channel Modeling

Channel modeling is the main aim of the channel characterization. It is the way of predicting the effect of the channel on the EM wave. It tells about the coverage, interference, etc., present in the channel. The accuracy of the model is of utmost importance to ensure that the real values are closely related to the predicted values.

Therefore, proper model tuning is very important [22, 25, 26, 28, 29].

5.1 *Empirical Models*

Empirical models are used in order to predict the channel model from the experimentally taken path loss measurements. There are various empirical models as follows:

(a) **Power Law Model**
It is considered to be the simplest form of an empirical path loss model, where n and K are the only two model parameters to be tuned with measurements.

$$L = 10n \log r + K \tag{1}$$

where K = path loss intercept, which represents the path loss taken at a reference distance, often taken as 1 m; n = path loss exponent.

This model does not support the accurate predictions.

(b) **Keenan–Motley Model**

It is the path loss model mainly for the prediction of macrocell.

$$L = L1 + 20n \log r + n_f \times a_f + n_w \times a_w \tag{2}$$

where n_f and n_w = loss factor corresponding to the number of floors and walls, respectively, intersected by the straight-line distance r between the terminals. a_f and a_w = attenuation factor per floor and wall, respectively, in decibels.

This model fails in sharp corners which makes it unreliable for indoor applications.

(c) **ITU-R Indoor Model**

This model accounts the floor loss, and the loss between points on the same floor is included by varying the path loss exponent.

$$L = 20 \log f_c + 10n \log r + L_f \times n_f - 28 \tag{3}$$

where n_f = path loss exponent Ω, and L_f = floor penetration loss, which varies with the number of penetrated floors n_f.

(d) **Indoor Dominant Path Model (DPM)**

It considers the signal of the dominant path rather than the direct path.

$$L = \frac{\lambda}{4\pi} + 10n \log r + L_e + \sum_{i=1}^{m} f(\phi, i) + \sum_{i=1}^{k} L_i - \Omega \tag{4}$$

(e) **COST-231 Multiwall Model**

This model accounts for the presence of wall and floor loses for individual partitions.

$$L = L_{FSL} + L_c + \sum_{i=1}^{W} L_{wi} n_{wi} + L_{fnf}^{((n_f+2)/(n_f+1))-b)} \tag{5}$$

5.2 Physical Models

A physical model tells about the propagation mechanisms with the actual geometry and EM nature of the scenario under consideration as propagation mechanisms such as reflection, refraction, diffraction, and others are taken into account when computing the signal strength at any point. Intelligent ray-tracing (IRT) model is widely used for channel characterization [2, 27, 30]. There are other various mod-

els like measurement-based prediction (MBP) which depends on the signal strength measurement, antenna coordinates, and radiation pattern.

On completion of the designing the models, proper validation is a requirement and for that repetitive measurement is must [26].

6 Types of Measurement Systems

One of the ways to characterize mobile radio propagation channel is continuous wave (CW) measurements which are often performed by transmitting an unmodulated single tone carrier. VNA is used widely for this type of measurements but narrowband is easily affected by path loss [31]. As CW measurements only use unmodulated single tone carriers, they are not suitable for measuring network performance, and therefore, other narrowband measurement equipment should be used [22, 26].

Another way adapted is the measurement using wideband which can easily characterize the parameters such as delay spread, delay profile, average delay, and coherence bandwidth [2, 3, 21, 26, 30, 31, 32]. Various techniques may be employed, some of which use the principle of transmitting many narrowband signals, either sequentially or simultaneously. However, due to the limitations of the abovementioned methods, genuine wideband sounding techniques are required. Examples of such techniques are periodic pulse sounding and pulse compression. The measurement results using wideband signals are reported in using spread spectrum technology with 11 order PN code [4, 8–10, 14, 22, 26, 30, 33].

7 Conclusion

This paper describes the need for channel characterization and increasing demand in the recent communication era. The requirement of multi-gigabit-speed everywhere is the urge of the users. Along with the humans, now machines are also becoming the customers. So, serving multi-gigabit-speed for billions of consumers not only at outdoors but also in indoors has become a huge task. When one comes from outdoor to indoor, the radio propagation starts changing, as fast fading becomes more severe in indoors. So, the properties such as reflectivity, refractivity, scattering, etc., are to be kept in mind. Therefore, the solution to the problem is the use of mmW band which provides huge bandwidth at a stretch. So, the typical frequency bands considered are 28, 39, 60, and 73 GHz. The use of mmW also leads to use of multiple antennas, thus forming sharper antenna beam which ultimately enhances the signal-to-noise ratio by reducing the interference, multipath effect. The concept of microcell is also coming into the picture as the mmW suffers huge path loss.

Another challenge is the outdoor-to-indoor communication. Researchers are also considering the model for such situation as the penetration loss becomes an issue.

The application of mmW also leads to the wideband measurement which helps in estimating the channel parameters like delay spread, average delay, coherence bandwidth, etc., thus enhancing the accuracy of the predicted channel model. Therefore, with all the predicted model, better receiver designing is possible that will lead to improved quality of service to the users, thus fulfilling the concept "anywhere, anytime".

References

1. Zavala AA. Indoor wireless communications
2. Ryu J, Partyka A, Subramanian S, Sampath A (2015) Study of the indoor millimeter wavelength channel. In: 2015 IEEE global communication conference GLOBECOM 2015, pp 1–5
3. Merwaday A, Rupasinghe N, Güvenç I, Saad W, Yuksel M (2014) USRP-based indoor channel sounding for D2D and multi-hop communications. In: 2014 IEEE 15th annual IEEE wireless and microwave technology conference WAMICON 2014
4. Anderson CR, Rappaport TS, Bae K, Verstak A, Ramakrishnan N, Tranter WH, Shaffer CA, Watson LT (2002) In-building wideband multipath characteristics at 2.5 and 60 GHz. In: Proceedings of the IEEE 56th vehicle and technology conference, vol 1, pp 97–101
5. Peng B, Rey S, Kürner T (2016) Channel characteristics study for future indoor millimeter and submillimeter wireless communications. In: 10th European conference on antennas and propagation, vol 110, pp 1–5 (2016)
6. Ma J, Shrestha R, Moeller L, Mittleman DM (2018) Invited article: channel performance for indoor and outdoor terahertz wireless links. APL Photonics 3:051601
7. Masa-Campos JL, Lalueza-Mayordomo JM, Taha-Ahmed B (2010) RF propagation in indoor environment at WiMAX band of 3.5 GHz. J Electromagn Waves Appl 24:2495–2508
8. Maccartney GR, Yan H, Sun S, Rappaport TS (2017) A flexible wideband millimeter-wave channel sounder with local area and NLOS to LOS transition measurements. IEEE Int Conf Commun 1–7 (2017)
9. Maccartney GR, Deng S, Sun S (2015) NYU WIRELESS TR 2015-002 technical report indoor office wideband millimeter-wave propagation measurements and channel models at 28 GHz and 73 GHz for ultra-dense 5G wireless networks, vol 3
10. Ryan J, Maccartney GR, Rappaport TS (2017) Indoor office wideband penetration loss measurements at 73 GHz. In: 2017 ieee international conference on communication workshop, ICC workshop, pp 228–233
11. Koymen OH, Partyka A, Subramanian S, Li J (2015) Indoor mm-wave channel measurements: comparative study of 2.9 GHz and 29 GHz. In: 2015 IEEE global communication conference GLOBECOM 2015, pp 7–12
12. Papazian PB, Lo Y (1999) Seasonal variability of a local multi-point distribution service radio channel. In: 1999 IEEE radio wireless conference, RAWCON 99 (Cat. No. 99EX292), pp 211–214
13. Chavero M, Polo V, Ramos F, Marti J (1992) Impact of vegetation on the performance of 28 GHz LMDS transmission, pp 35–40
14. Rappaport TS, Deng S (2015) 73 GHz wideband millimeter-wave foliage and ground reflection measurements and models. In: 2015 IEEE international conference on communication workshop, ICCW 2015, pp 1238–1243
15. Akdeniz MR, Liu Y, Samimi MK, Sun S, Rangan S, Rappaport TS, Erkip E (2014) Millimeter wave channel modeling and cellular capacity evaluation. IEEE J Sel Areas Commun 32:1164–1179
16. Sun S, Rappaport TS, Heath RW, Nix A, Rangan S (2014) MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both? IEEE Commun Mag 52:110–121

17. Kristem V, Sangodoyin S, Bas CU, Kaske M, Lee J, Schneider C, Sommerkorn G, Zhang CJ, Thoma RS, Molisch AF (2017) 3D MIMO outdoor-to-indoor propagation channel measurement. IEEE Trans Wirel Commun 16:4600–4613

18. Sun S, Rappaport TS (2017) Millimeter wave MIMO channel estimation based on adaptive compressed sensing

19. Sun S, Maccartney GR, Rappaport TS (2017) A novel millimeter-wave channel simulator and applications for 5G wireless communications. IEEE Int Conf Commun 10

20. Wang Y, Lu W, Zhu H (2010) Experimental study on indoor channel model for wireless sensor networks and internet of things r l. Antenna 624–627

21. Ko J, Lee SU, Kim YS, Park DJ (2016) Measurements and analyses of 28 GHz indoor channel propagation based on a synchronized channel sounder using directional antennas. J Electromagn Waves Appl 30:2039–2054

22. Al-Samman AM, Abd Rahman T, Azmi MH (2018) Indoor corridor wideband radio propagation measurements and channel models for 5G millimeter wave wireless communications at 19 GHz, 28 GHz, and 38 GHz Bands. Wirel Commun Mob Comput 2018

23. Aihua H, Zhongliang D, Yao Z (2017) Study on the method of indoor wireless channel characteristic measurement and analysis based on vector network analyzer

24. Sulyman A, Nassar A, Samimi M, Maccartney G, Rappaport T, Alsanie A (2014) Radio propagation path loss models for 5G cellular networks in the 28 GHZ and 38 GHZ millimeter-wave bands. IEEE Commun Mag 52:78–86

25. Sulyman AI, Alwarafy A, MacCartney GR, Rappaport TS, Alsanie A (2016) Directional radio propagation path loss models for millimeter-wave wireless networks in the 28-, 60-, and 73-GHz bands. IEEE Trans Wirel Commun 15:6939–6947

26. Rappaport TS, MacCartney G, Samimi M, Sun S (2015) Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. IEEE Trans Commun 1–1

27. Bou-El-Harmel A, Benbassou A, Belkadid J, Mechatte N (2017) Effect of quasi-isotropic antenna orientation on indoor multipath propagation characteristics in RSN applications. Int J Antennas Propag 2017

28. Zhang M, Polese M, Mezzavilla M, Rangan S, Zorzi M (2017) ns-3 implementation of the 3GPP MIMO channel model for frequency spectrum above 6 GHz

29. MacCartney GR, Rappaport TS (2017) Rural macrocell path loss models for millimeter wave wireless communications. IEEE J Sel Areas Commun 35:1663–1677

30. MacCartney GR, Rappaport TS (2017) A flexible millimeter-wave channel sounder with absolute timing. IEEE J Sel Areas Commun 35:1402–1418

31. Martinez-Ingles MT, Gaillot DP, Pascual-Garcia J, Molina-Garcia-Pardo JM, Rodríguez JV, Rubio L, Juan-Llácer L (2016) Channel sounding and indoor radio channel characteristics in the W-band. Eurasip J Wirel Commun Netw 2016:1–8

32. Jagannatham AK, Erceg VO (2004) MIMO indoor WLAN channel measurements and parameter modeling at 5.25 GHz. In: IEEE 60th vehicle technology conference 2004, VTC2004-Fall, vol 1, pp 106–110

33. Sun S, Yan H, Maccartney GR, Rappaport TS (2017) Millimeter wave small-scale spatial statistics in an urban microcell scenario. IEEE Int Conf, Commun

Phased-MIMO Radar in Low SNR Regime

Samarendra Nath Sur, Rabindranath Bera and Bansibadan Maji

Abstract Exploitation of coherency gain and diversity gain to improve the multiantenna system performance is a hot research topic. This paper deals with the performance analysis of phased-MIMO radar by utilizing the above-said gains. The main focus of this paper is to demonstrate the superiority of phased-MIMO radar over MIMO radar in low SNR regime. Detailed mathematical analysis along with simulation has been presented in this paper

1 Introduction

Researchers are continuously exploring and engaging them self to improve the radar performance [1, 2] through advanced signal processing techniques. In this regards, multiple antenna signal processing algorithms provide a tremendous boost in radar performance [2]. The demand for sophisticated radar is increasing day by day with target localization and characterization being the most important aspect. But it becomes more complex in rich scattering environment, particularly in low SNR condition. In low SNR condition, radar performance is greatly influenced by target scintillations [3, 4]. On the contrary, radar should have a waveform which can not be detected by

---

S. N. Sur (✉) · R. Bera · B. Maji
Electronics and Communication Engineering Department, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: Samar.sur@gmail.com; samarendra.s@smit.smu.edu.in

R. Bera
e-mail: rbera@smit.smu.edu.in; rbera50@gmail.com

B. Maji
e-mail: bansibadan.maji@ece.nitdgp.ac.in

S. N. Sur · R. Bera · B. Maji
Electronics and Communication Engineering Department, National Institute of Technology, Durgapur, Durgapur, West Bengal, India

the enemy receiver. In present-day electronic warfare, the receiver system should be equipped in such a way that it can easily recognize the waveform with classification, and identification capabilities. Therefore, from both the way it is necessary to analyze the system performance in low SNR condition. Particularly for designing low probability of intercept radar (LPI), through the wall (TTW), ground penetrating radar (GPR), and also for the stealth target detection, low SNR of the received signal severely effects the radar performance. Now as in [5], MIMO, by exploiting the diversity radar performance can be enhanced [6–8]. Detection and localization of target with TTW imaging radar and GPR radar is a really tough task because of its effects on electromagnetic propagation due to phase distortion and delay [6]. Here in this paper, phased-MIMO radar [1, 5] performance has been evaluated with low SNR situation. As it utilizes both the coherency gain and the diversity gain, it improves the system performance tremendously under such circumstances. In other words, phased-MIMO [1, 9] system enhanced the robustness of the target detection.

In this paper, the output SNR corresponding to MIMO and phased-MIMO radar have been compared. The mathematical expression has been derived based on the low SNR approximation and validated with the simulated results.

2 Mathematical Model

Let us consider the radar detection problem at delay τ as follows [3, 4]:

H_0: Fall Detection
H_1: Target Detected

Based on the Neyman–Pearson sense, the optimal detector likelihood ratio test (LRT) can be given as [10]

$$T = \log \frac{f(r(t)|H_1)}{f(r(t)|H_0)} \underset{>H_0}{\overset{<H_1}{}} \delta_{th} \tag{1}$$

where $f(r(t)|H_0)$ and $f(r(t)|H_1)$ are the probability density function of the observation corresponding to the radar detection and δ_{th} is the threshold, which is related to the probability of false alarm (Pfa).

In this paper, we have considered M_t, transmitter antenna, and N_r, receiver antennas.

Let x be the output of the matched filter banks and x is a complex random variable with zero mean. And the elements of the correlation matrix are $\sigma_n^2 I_{M_t N_r}$ and $\left(\left(\frac{E}{M_t}\right) + \sigma_n^2\right) I_{M_t N_r}$ under the radar detection hypothesis, where E be the total transmitted energy, and $\sigma_n^2 is$ the noise level per receive element.

2.1 Discussion Output SNR

As in [4], the detector's output SNR, β is defined as given below:

$$\beta = \frac{|E(T|H_0) - E(T|H_1)|^2}{\frac{1}{2}[Var(T|H_0) + Var(T|H_1)]} \tag{2}$$

2.2 MIMO Radar

For MIMO radar, $(T/H_0) = NM(\sigma_n)^2$ and $E(T/H_1) = MN\left(\sigma_n^2 + \frac{E}{M}\right) = MN\sigma_n^2 + EN$. And also $Var(T/H_0) = MN\left(\sigma_n^2\right)^2 = NM\sigma_n^4$ and $Var(T/H_1) = MN\left\{\left(\sigma_n^2 + \frac{E}{M}\right)^2\right\} = MN\left(\sigma_n^4 + \frac{E^2}{M^2} + 2.\sigma_n^2\frac{E}{M}\right)$.

Therefore, using Eq. (2), the output SNR level can be calculated as mentioned below:

$$\beta_{MIMO} = \frac{|E(T/H_0) - E(T/H_1)|^2}{\frac{1}{2}[Var(T/H_0) + Var(T/H_1)]} = \frac{(EN)^2}{\frac{1}{2}\left[2MN\sigma_n^4 + \frac{E^2}{M^2} + 2\frac{\sigma_n^2 E}{M}\right]}$$

$$= \frac{\rho^2 N}{M\left(1 + \frac{\rho^2}{2M^2} + \frac{\rho}{M}\right)} \tag{3}$$

where the SNR is denoted by ρ, as the ratio between the total transmitted energy and the noise level per receive element and can be defined as $\rho = \frac{E}{\sigma_n^2}$.

2.3 Phased-MIMO Radar

To analyze the phased-MIMO radar following parameters are considered, M_{coh} represents the number of elements per sub-array at the transmitter side. N_{coh} is the number of elements per sub-array at the receiver side. M_{div} and N_{div} represent the number of Sub-Array at the transmitter and receiver side, respectively.

In case of phased-MIMO radar, $E(T/H_0) = \sigma_n^2 N_{coh}M_{div}N_{div}$ and $E(T/H_1) = \left(\sigma_n^2 N_{coh} + \frac{EN_{coh}^2 M_{coh}}{M_{div}}\right)M_{coh}$. And $Var(T/H_0) = M_{div}N_{div}\left(\sigma_n^2 N_{coh}\right)^2$ and $Var(T/H_1) = M_{div}N_{div}\left(\sigma_n^2 N_{coh} + \frac{EN_{coh}^2 M_{coh}}{M_{div}}\right)^2$.

Therefore, the output SNR level can be calculated as

$$\beta_{phased-MIMO} = \frac{\left(\frac{EN_{coh}^2 M_{coh}}{M_{div}}.M_{div}N_{div}\right)^2}{\frac{1}{2}M_{div}N_{div}\left[\sigma_n^4 N_{coh}^2 + \sigma_n^4 N_{coh}^2 + \frac{EN_{coh}^2 M_{coh}^2}{M_{div}^2} + 2\frac{\sigma_n^2 N_{coh}^3 E}{M_{div}}\right]}$$

$$= \frac{E^2 N_{coh}^2 M_{coh}^2 N_{div}/\sigma_n^4}{M_{div}\left[1 + \frac{N_{coh}^2 M_{coh}^2}{2M_{div}^2}\frac{E^2}{\sigma_n^4} + \frac{E}{\sigma_n^2}\frac{N_{coh}M_{coh}}{M_{div}}\right]} \tag{4}$$

$$\beta_{phased-MIMO} = \frac{\rho^2 N_{coh}^2 M_{coh}^2 N_{div}}{M_{div}\left[1 + \frac{N_{coh}^2 M_{coh}^2}{2M_{div}^2}\rho^2 + \frac{N_{coh}M_{coh}}{M_{div}}\rho\right]} \tag{5}$$

2.4 Low SNR Approximation

In high SNR regime, Eqs. (3) and (5) can be expressed as

$$\beta_{MIMO}|_{low} = \frac{N/M}{\left(\frac{1}{2\rho^2} + \frac{1}{\rho M}\right)} \tag{6}$$

$$\beta_{phased-MIMO}\big|_{low} = \frac{N_{coh}^2 M_{coh}^2 N_{div}/M_{div}}{\left[\frac{1}{\rho^2} + \frac{N_{coh}M_{coh}}{\rho M_{div}}\right]} \tag{7}$$

$$\frac{\beta_{MIMO}|_{low}}{\beta_{phased-MIMO}\big|_{low}} = \frac{N/M}{\left(\frac{1}{2\rho^2} + \frac{1}{\rho M}\right)} \times \frac{\left[\frac{1}{\rho^2} + \frac{N_{coh}M_{coh}}{\rho M_{div}}\right]}{N_{coh}^2 M_{coh}^2 N_{div}/M_{div}} \tag{8}$$

Therefore, Eq. (8) can be written as

$$\frac{\beta_{MIMO}|_{low}}{\beta_{phased-MIMO}\big|_{low}} = \frac{\rho N_{coh}M_{coh} + M_{div}}{(\rho + M_{coh}M_{div})N_{coh}M_{coh}^2} \tag{9}$$

Again for a symmetrical configuration, the above equation can be written as $N_{coh} = N_{div} = M_{coh} = M_{div} = h$.

Therefore,

$$\frac{\beta_{MIMO}|_{low}}{\beta_{phased-MIMO}\big|_{low}} = \frac{\rho h + 1}{(\rho + h^2)h^2} < 1[k(say)] \tag{10}$$

Equation 10 indicates that phased-MIMO radar is more advantageous over the MIMO radar in low SNR condition.
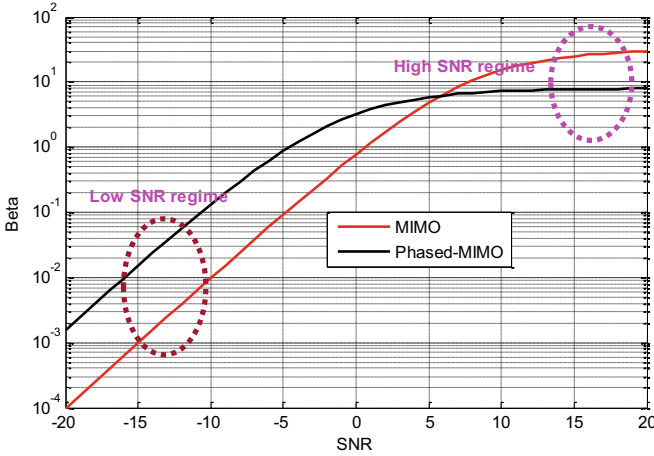
Fig. 1 Output SNR comparison for different radar systems

3 Simulation Results

Figure 1 represents the output SNR comparison for the MIMO and phased-MIMO radar system. Here, MIMO radar is of 4×4 antenna arrangement, whereas phased-MIMO radar has two sub-arrays with two antennas in each array. It is clear from the figure that at low SNR regime phased-MIMO performs significantly well with respect to MIMO system. As presented, at an SNR of -15 dB, the output SNR of phased-MIMO system is 10 times more than the MIMO radar system.

Figure 2 depicts the performance of a phased-MIMO ($M_{coh} = 2, N_{coh} = 2, M_{div} = 2, N_{div} = 2$) radar system. Here, Pd is observed with the variation in SNR (-22 dB to 0 dB) and Pfa (10-5 to 1). For a fixed SNR level, the curve between Rd and Pfa represents the receiver operating curves for the phased-MIMO system under the known noise condition.

4 Conclusion

Through the simulated results, it is clear that the choice of coherent signal processing and space diversity play an important role to enhance the radar performance. As presented, in low SNR condition, phased-MIMO radar is the best solution in low SNR environment. Therefore, in severe channel condition, like in presence of a highly dynamic condition or in sea environment, or in case of through wall or underground situation, exploitation of both signal processing techniques will provide the best solution.
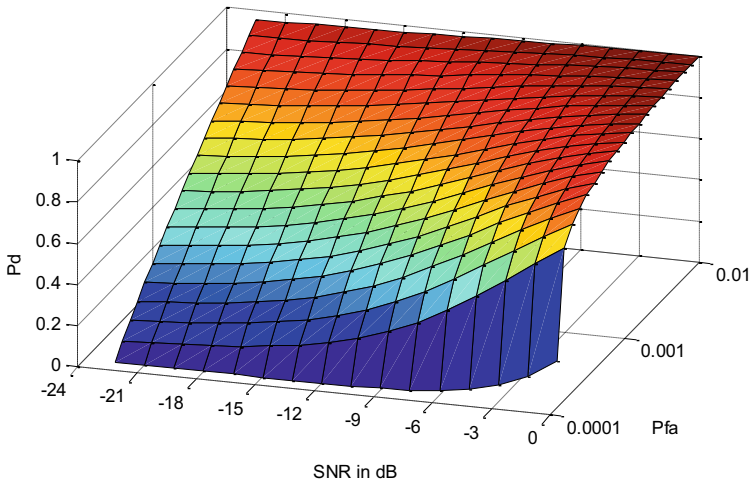
Fig. 2 Variation in Pd with the variation in SNR and Pfa

References

1. Hassanien Aboulnasr, Vorobyov Sergiy A (2010) Phased-MIMO radar: a tradeoff between phased-array and MIMO radars. IEEE Trans Signal Process 58(6):3137–3151
2. Haykin S, Litva J, Shepherd TJ (1993) Radar array processing. Springer, New York
3. Levanon N (1988) Radar principles, 1st edn. Wiley, New York
4. Fishler E, Haimovich A, Blum RS, Cimini LJ, Chizhik D, Valenzuela RA (2006) Spatial diversity in radars—models and detection performance. IEEE Trans Signal Process 54(3):823–838
5. Skolnik M (2002) Introduction to radar systems, 3rd edn. McGraw-Hill, New York
6. Lehmann N, Fishler E, Haimovich A, Blum R, Chizhik D, Cimini L, Valenzuela R (2007) Evaluation of transmit diversity in MIMO radar direction finding. IEEE Trans Signal Process 55(5):2215–2225
7. Li J, Stoica P (2007) MIMO radar with colocated antennas. IEEE Signal Process Mag 24:106–114
8. Bekkerman, Tabrikian J (2006) Target detection and localization using MIMO radars and sonars. IEEE Trans Signal Process 54(10):3873–3883
9. Wang W-Q, Shao H (2012) A flexible phased-MIMO array antenna with transmit beamforming. Int J Antennas Propag 2012, Article ID 609598
10. Trees HLV (1968) Detection and estimation, and modulation theory, Vol I. Wiley, New York

Implementation and Analysis of Spectrum Sensing Using Energy Detection Method

Dhrubajyoti Bera, Aditya Chandra, Samarendra Nath Sur and Arun Kumar Singh

Abstract The continuously increasing demand for wireless communication has highlighted the problem of underutilization of available spectrum. Cognitive radio technology provides a solution to this problem by utilizing the white spaces in the spectrum and allowing the unlicensed user to share licensed spectrum. There are various techniques to sense the licensed spectrum; one such technique is energy detection method. The performance parameters in spectrum sensing are probability of detection and probability of false alarm. This paper explores the performance of energy detection method in case of LOS communication and NLOS communication.

Keywords Cognitive radio · Spectrum sensing · Energy detection · Line of sight (LOS) · Non-line of sight (NLOS) · Primary user · Secondary user

1 Introduction

Electromagnetic spectrum (or just "spectrum") is used to send electromagnetic signals through space. Spectrum is a valuable, scarce, and finite natural resource that is needed for many different applications [1]. It carries signals used for radio, television, mobile phones, mobile broadband, scientific research, defense activities, public safety, and other personal communications systems. Conventionally, a portion of the spectrum is licensed to a user called primary user (PU) who can use the spectrum whenever they wish. However, it leads to underutilization of the spectrum. So, when the licensed user is not utilizing the spectrum allotted to it, unlicensed users or sec-

D. Bera (✉) · A. Chandra · S. N. Sur · A. K. Singh
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: dhrubajyoti.bera@gmail.com

S. N. Sur
e-mail: samar.sur@gmail.com; samarendra.s@smit.smu.edu.in

A. K. Singh
e-mail: arunsingh.smit@gmail.com

ondary users (SU) use the spectrum. But for this to take place without unintended interference and chaos, the secondary user must know when the spectrum is free for use and is not being used by a PU. The method used by these secondary users to sense whether a spectrum is free or not is called "Spectrum Sensing". When the primary user returns, the secondary user has to abandon that frequency band immediately. This is called "Forced Termination" [2]. Spectrum sensing can be performed as either distributed or cooperative. Cooperative spectrum sensing requires extra bandwidth for negotiation. Noncooperative spectrum sensing techniques like energy detection technique are based on the detection of the weak signal from a primary transmitter [3–5]. Hence, the complexity is less but the trade-off is quality of sensing [6].

2 Mathematical Modeling

In Fig. 1, a model of communication system is given. We can see a transmitter antenna that generates a signal and passes through a wireless channel. This signal is received at the receiver antenna, where the receiver signal can be processed in various spectrums sensing techniques, e.g., cyclostationary feature detection method, energy detection method, matched filter feature detection method [7–12]. The transmitted signal is also processed through various real-life channels like the AWGN channel and Rician channel, which are discussed in detail in the later part of the report.

A binary sequence is generated at first, which is modulated and transmitted afterward. The data stream is encoded using nonreturn to zero (NRZ) encoding. The first thing here is to keep the binary signal's amplitude of 1, where it has value on and replace the value of "0" with "−1".

Second, the numbers of elements of binary signals are upsampled by repeating the number of 1's and number of −1's to make the size of matrix equal to the size of carrier wave over the specific time. Afterward, the NRZ encoded data is multiplied
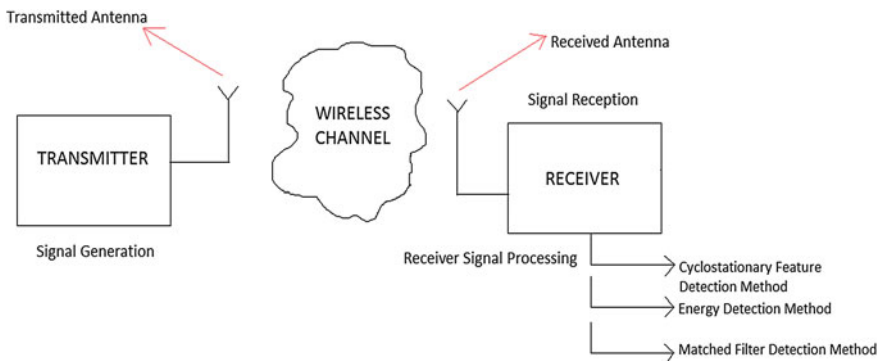
Fig. 1 Pictorial representation of the spectrum sensing techniques

by a carrier wave, according to the given equation. This will act to make a 180° phase shift of the carrier wave where -1 is present and a zero degree phase shift where 1 is present.

$$Sb(t) = \sqrt{2} \text{ ts/T} * \cos(2\pi \text{ fct} + \pi(1 - n)) \tag{1}$$

$$n = 0 \text{ for '0' and } n = 1 \text{ for '1'} \tag{2}$$

In order to make the decision whether the PU user is using the spectrum or not, a statistical model is needed for the PU signal for detection and then consider the situation without PU. Let us assume a simple received signal is modeled as

$$y(n) = s(n) + w(n) \tag{3}$$

where s(n) is the signal under detection, w(n) is the additive white Gaussian noise (AWGN) sample, and n is the sample index. When there is no PU signal present, s(n) = 0. Detection algorithms calculate a detection statistic to be compared with a detection threshold:

$$\rho x > \lambda X \tag{4}$$

in which ρx and λX are the detection statistic and detection threshold, respectively. Now we can make a decision of the spectrum usage by comparing ρx to a fixed threshold value λX depending on the detection scheme in use. This can be expressed as a hypothesis comparison

$$H_0: y(n) = w(n) \tag{5}$$

$$H_1: y(n) = s(n) + w(n) \tag{6}$$

where hypothesis H_0 denotes that no PU is present and hypothesis H_1, PU is present.
The detection performance can be expressed with two probabilities:

Probability of detection P_d.
False alarm rate P_{fa}.

P_d means the probability of detecting a signal on a spectrum band when it really exists. Therefore, a higher P_d equals better performance. P_d can be obtained as

$$P_d = P_r(\rho x > \lambda X | H_1) \tag{7}$$

P_{fa} is the probability for the test to falsely indicate that the spectrum is in use when it really is not. P_{fa} can be expressed as

$$P_{fa} = P_r(\rho x > \lambda X | H_0) \tag{8}$$

P_{fa} should be kept as low as possible to prevent underutilization of spectrum. A decision metric for the energy detector can be expressed as

$$M = \sum_{n=0}^{N} |y(n)|^2 \tag{9}$$

The AWGN can be modeled as a zero-mean Gaussian random variable with variance, σ_w^2, i.e., $w(n) = N(0, \sigma_w^2)$, and the signal, as a zero-mean Gaussian variable $s(n) = N(0, \sigma_w^2)$. Under these assumptions, the decision metric M follows chi-square X^2 distribution with 2N degrees freedom X_{2N}^2 and therefore, it can be modeled as

$$H_0 : M = \frac{\sigma_w^2}{2} X_{2N}^2 \tag{10}$$

$$H_1 : M = \frac{\sigma_w^2 + \sigma_s^2}{2} X_{2N}^2 \tag{11}$$

Now the detection probabilities P_{fa} and P_d for energy detector can be calculated as

$$P_{fa} = 1 - \Gamma\left(L_f L_t, \frac{\lambda_E}{\sigma_w^2}\right) \tag{12}$$

$$P_d = 1 - \Gamma(L_f L_t, \frac{\lambda_E}{\sigma_w^2 + \sigma_s^2}) \tag{13}$$

where λ_E is the decision threshold, and $\Gamma(a, x)$ is the incomplete gamma function. The threshold in energy detector-based sensing algorithms depends on the noise variance, and even small noise estimation errors can cause significant performance loss (Fig. 2).

3 Results

Figure 3 shows the experimental setup including two antennas: a reflector plate and WARP V2 MANGO board (FPGA). For line of sight (LOS) communication, both antennas were placed in front of each other so that signal can pass directly from transmitter to receiver. For non-line of sight (NLOS) communication, both antennas were placed in such a way that signals from transmitter reach the receiver after being reflected by the reflector plate.

Figure 4 shows the spectrum of the signal along with the noise. The peak (in the main lobe) in the middle is the band under use (higher SNR) and rest of the band (side lobes) contains noise.

On adding threshold it is decided that the primary user is present or not based upon the detections of signal whether it is above the threshold level.
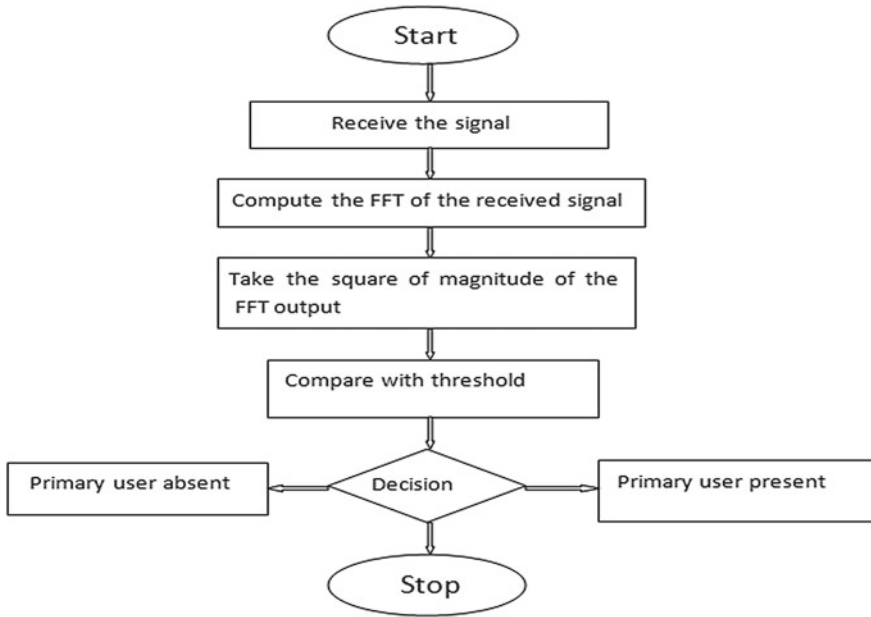
Fig. 2 Flowchart for energy detection method

Fig. 3 Image of the experimental setup

It is clear from Fig. 5 that at the higher threshold value, the desired signal is detected along with minimal noise signal. So the probability of detection is high and the probability of false alarm is low. However, at a low threshold value, noise will be detected along with the desired signal. It would adversely affect the detection to a great extent which makes the threshold a very important factor for the energy detection method.
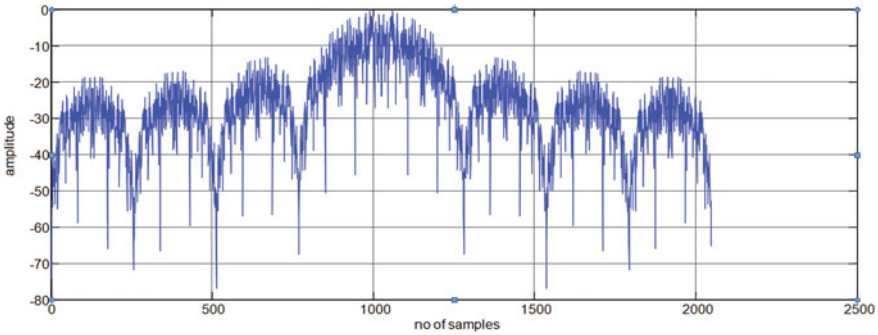
Fig. 4 Signal spectrum
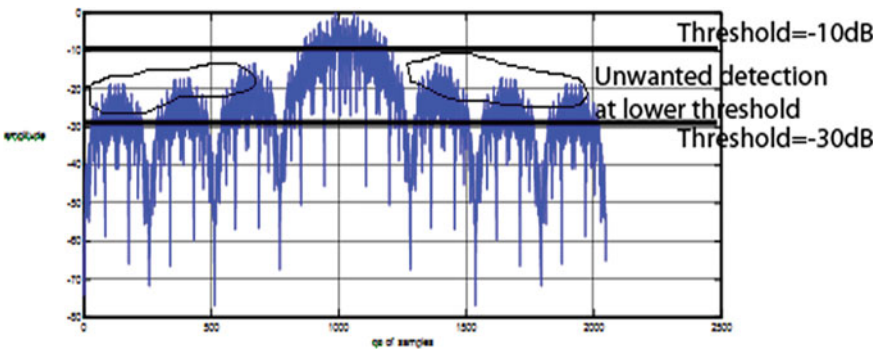
Fig. 5 Effect of threshold variation on detection ability in frequency domain

The effectiveness of a constant threshold value also changes with a change in SNR. Therefore, system performance is measured at different thresholds to understand how threshold works with varying SNR. This study pointed out the requirement of the adaptive thresholding.

Figures 6, 7 and 8 depict the nature of detection probability and false detection probability curves with increasing threshold level at different SNR levels, i.e., 20, 30, and 45 dB for line of sight (LOS) communication system.

Figures 9, 10 and 11 depict the nature of detection probability and false detection probability curves with increasing threshold level at different SNR levels, i.e., 20, 30, and 45 dB for non-line of sight (NLOS) communication system.

It is also observed that the probability of detection and false detection curves meet at a point. This crossover point implies the minimum threshold value required by the system for successful detection. Comparing figures, it is also noticed that the crossover point is shifting toward the right.
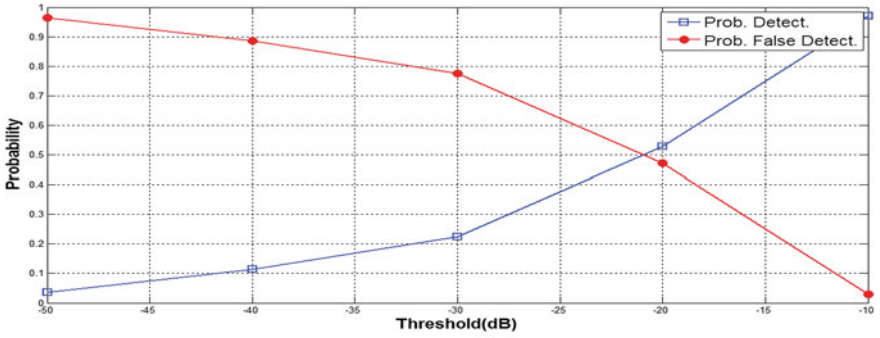
Fig. 6 Threshold versus detection prob. and false detection prob. curve at SNR = 20 dB
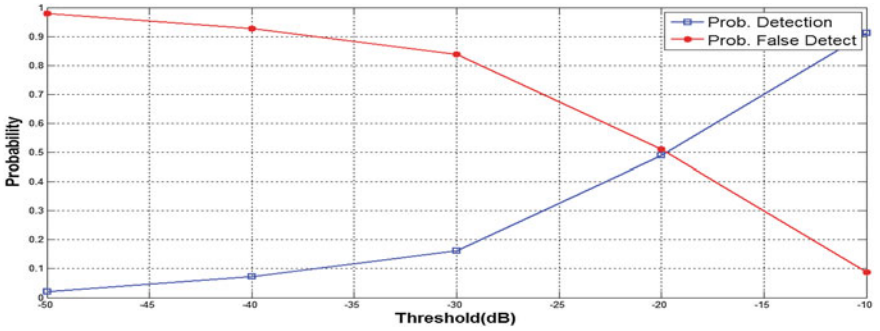
Fig. 7 Threshold versus detection prob. and false detection prob. curve at SNR = 30 dB
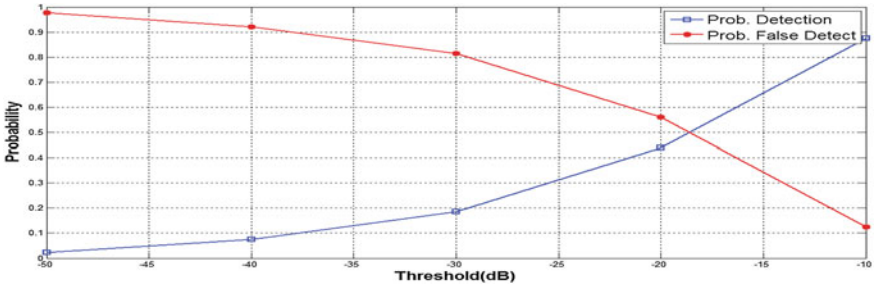
Fig. 8 Threshold versus detection prob. and false detection prob. curve at SNR = 45 dB
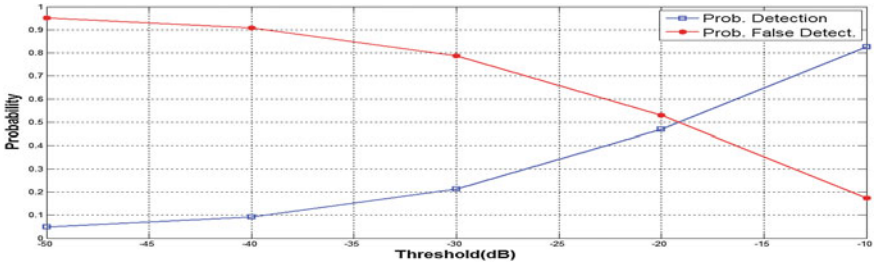
Fig. 9 Threshold versus detection prob. and false detection prob. curve at SNR = 20 dB
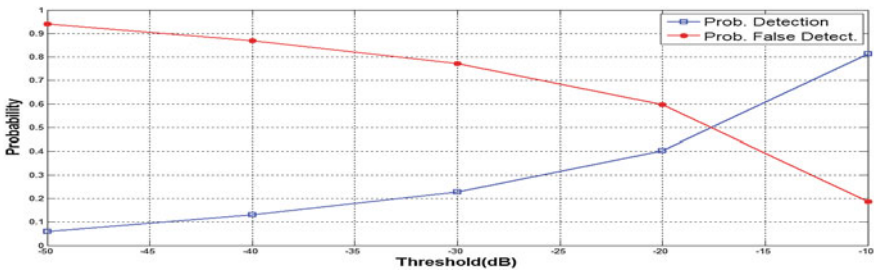
Fig. 10 Threshold versus detection prob. and false detection prob. curve at SNR = 30 dB
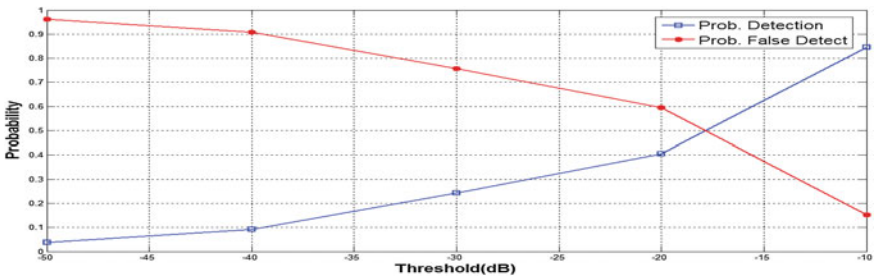
Fig. 11 Threshold versus detection prob. and false detection prob. curve at SNR = 45 dB

The other factors that have an impact on the result are the distance between transmitter and receiver antennas and presence of multiple reflectors.

The presence of multiple reflectors degrades the SNR at receiver side due to multipath effect and this leads to false detection.

4 Conclusion

This paper deals with the hardware implementation of the energy detection-based spectrum sensing. Here, the system performance has been analyzed for both LOS and NLOS condition. The probability of false detection is low with the increase in SNR values. The maximum probability of detection for a fixed transmitter power is more in case of LOS communication than NLOS communication. In other words, the probability of false detection for a fixed transmitter power is less in case of LOS communication than NLOS communication.

References

1. Federal Communications Commission (2005) Notice of proposed rulemaking and order: facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies. ET Docket No. 03-108
2. Zhu X, Shen L, Yum TP (2007) Analysis of cognitive radio spectrum access with optimal channel reservation. IEEE Commun Lett 11:304–306
3. Yucek T, Arslan H (2009) A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun Surv Tutor 11(1):116–130
4. Cabric D, Tkachenko A, Brodersen R (2004) Spectrum sensing measurements of pilot, energy, and collaborative detection. IEEE Mily Commun Conf 1–7
5. Engelman R (2002) Report of the spectrum efficiency working group. Technical report, FCC
6. Garhwal A, Bhattacharya PP (2011) A survey on spectrum sensing techniques in cognitive radio. Int J Next Gener Netw 3(4):196–205
7. Sreenivasan R, Sasirekha GVK, Bapat J (2011) Adaptive cooperative spectrum sensing using group intelligence. Int J Comput Netw Commun 3(3):1–7
8. Akyildiz IF, Lo BF, Balakrishnan R (2005) Cooperative spectrum sensing in cognitive radio networks: a survey. Phys Commun 4(1):137–143
9. Mirza A, Arshad FB (2011) Performance analysis of cyclostationary sensing in cognitive radio networks. Technical report, Halmstad University
10. Kokinnen K et al (2010) On the implementation of autocorrelation-based feature detector. In: 4th international symposium on communications, control and signal processing
11. Dandawate AV, Giannakis GB (1994) Statistical tests for presence of cyclostationarity. IEEE Trans Signal Process 42(9):2355–2369
12. Kim K, Akbar IA, Bae KK, Um J-S, Spooner CM, Reed JH (2007) Cyclostationary approaches to signal detection and classification in cognitive radio. In: IEEE international symposium on New Frontiers in dynamic spectrum access networks 212–215

A Comparative Study on Different Genres of Image Restoration Techniques

Monica Singh, Sujala Pradhan, Md Ruhul Islam and N. Chitrapriya

Abstract Numerous types of images are used as sources of information for investigation and elucidation. While a picture is going through conversion from one form to another such as scanning, transmitting, digitizing, and storing, it may get corrupted with unwanted signal called noise. Before applying any image processing tools to the image, it must go through a preprocessing phase to de-noise it. This paper highlights the diverse categories of noise that can possibly affect an image. The technique used to restore the image is largely influenced by the sort of noise affecting the image. All the available algorithms to do so have their pros and cons. A juxtaposition of these proposals has been done here.

1 Introduction

The most significant of all our senses is vision, and hence, images play an important part in human cognizance. The principal query that arises over here is "what is noise". In a layman's language, anything that is visible as spots in an image thus lowering the quality of the image is called noise. Technically speaking, noise is a random variation of image intensity arising due to an unwanted signal corrupting the image

M. Singh · S. Pradhan · M. R. Islam · N. Chitrapriya (✉)
Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology,
Majhitar 737136, Sikkim, India
e-mail: cpriya.work@gmail.com

M. Singh
e-mail: monica.cse.smit@gmail.com

S. Pradhan
e-mail: sujala_1207@yahoo.co.in

M. R. Islam
e-mail: ruhulislam786@gmail.com

373

quality. An alternate way of explaining noise is in terms of pixels. The pixels in the image show an intensity value different than its true value. The sole aim of any good image restoration method should be to remove the noise while preserving the details. To sustain the low contrast, fine details are not an easy task.

1.1 Varieties of Noise

The commencement of image noise can be at the time of transmission or acquisition. A myriad of facets can be accountable for the initiation of image noise, the magnitude of which can vary from imperceptible tiny specks to large grainy like appearances.

While explaining the types of noises the utmost significant aspect is the selection of the apt "frame of reference".

The first frame of reference selected is that of image dependency, corresponding to which the following two types exist:

1. Image-independent noise
2. Image-dependent noise

The image-independent noise is modeled upon the additive noise system, in which the image data (d) achieved is the "addition" of the noise (n) and true image (t). Such kinds of noise are additive in nature. A multiplicative or nonlinear base is the most apt model upon which an image-determined noise can be based, making it mathematically complex. Thus for maintaining the simplicity, an assumption is to be made that any corrupting noise in an image as an independent noise.

Second, we can classify noise on the premises of their source and inherent characteristics.

a. Gaussian Noise

In earlier times, Gaussian noise appeared during image acquisition, i.e., during taking the image like sensor noise (noise in a digital camera sensor) caused due to poor lighting and/or high temperature and/or transmission. Such a kind of noise is individualistic at each pixel, thus modifying each pixel differently.

b. Salt-and-Pepper Noise

Fat-tail distributed or "impulsive" noise can additionally be accosted as salt-and-pepper noise or spike noise. The utmost trait of a picture corrupted by such noise is that it consists of dark pixels in brighter regions and bright pixels in darker regions. The basic source of this noise is the error triggered due to analog to digital converters, bit errors in transmission, etc. Due to the similarity in the aspect of a salt-and-pepper, the same name is given as the name of the noise.

c. Shot Noise

The term exposure can be defined as the amount of light per unit area w.r.t photography. When an image is being acquired, the quantity of photons perceived at a

Fig. 1 Image affected by Gaussian noise [1]

stated exposure level can vary, termed as "statistical quantum fluctuations". Because of such fluctuations, a noise is usually generated in the darker area of an image. Usually, a shot noise follows a Poisson distribution.

d. Quantization Noise (Uniform Noise)

The noise induced by compressing a range of pixels values to a solitary quantum value (discrete levels) in a perceived image is called quantization noise. It forms a uniform distribution approximately. Though it can be contingent on signal, it will be signal individualistic on the condition that other origins of noise are in abundance so as to cause fluctuations or if the fluctuation is directly applied.

e. Film Grain

The photographic film grain is a signal-determined noise, with indistinguishable statistical dissemination as shot noise. If the film grains are equally dispersed and each grain has a uniform and individualistic probability of evolving into a dark silver grain after photon absorption. After the absorption, the size of such non-illuminated grains in an area will become inconsistent followed by a binomial distribution.

f. Anisotropic Noise

Some sources of noise might cause a substantial amount of inclination in outputs. For instance, image sensors are infected by row/column noise but rarely (Figs. 1, 2, 3, 4, 5 and 6).
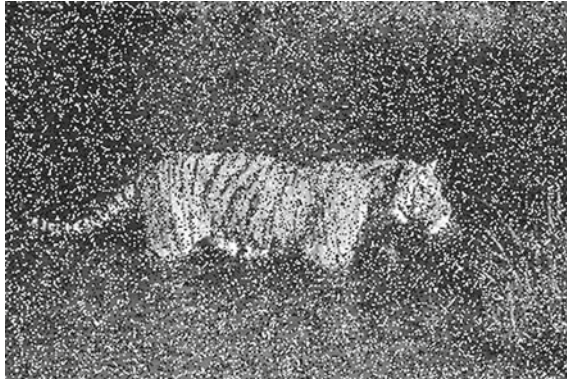
Fig. 2 Image affected by salt-and-pepper noise [2]

Fig. 3 Image affected by shot noise [3]

Fig. 4 Image affected by quantization noise [4]

Fig. 5 Image affected by film grain [5]
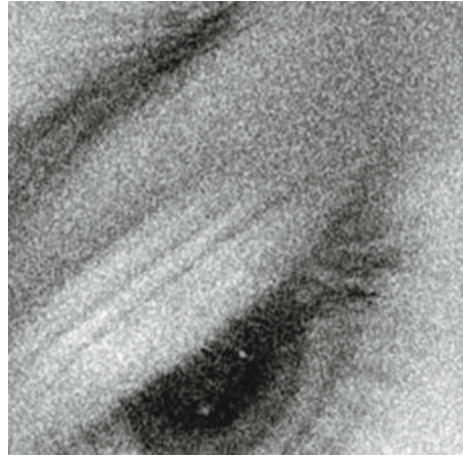
Fig. 6 Image before and after anisotropic noise [6]

(A) Image before Anisotropic Diffusion algorithm applied

(B) Image after Anisotropic Diffusion algorithm applied
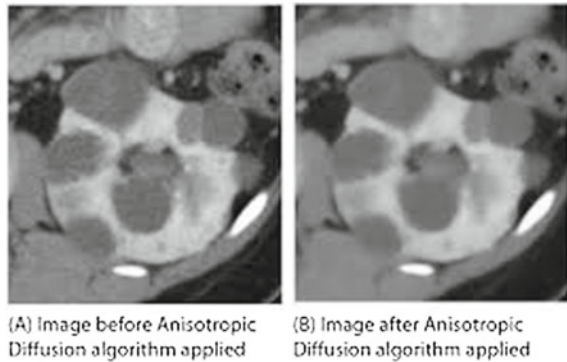
2 Noise Removal Techniques

While determining a technique for removing noise, one must consider several factors:

- The available time and computer power: The application of noise reduction within a fragment of a second using a tiny onboard CPU should be the minimal capability of a digital camera, while a desktop CPU has much more time and power.
- Is forfeiting some real detail sufficient, if that allows more noise to be eliminated?
- The attributes of the noise and of the image features for better decision-making.
- The fore coming image processing approaches.

As mentioned earlier, the noise reduction method is application-dependent, i.e., it is reliant on the genre of noise affecting the image. Hence for most categories of noise, an algorithm for its removal or reduction has been devised. The existing de-noising tactics or proposals are filtering methodologies, multi-fractal approach, and wavelet-based approach

2.1 Filtering Techniques

There are two extensively varying methods used for noise removal: (a) linear filter and (b) nonlinear filters. Thus, we are compelled to analogize these two filters so as to select the better existing one.

Comparison between Linear and Nonlinear Filters:

To compare linear and nonlinear filters, we need to describe the attributes that comprise both of them.

2.1.1 Linear Filters

A linear system is delineated by several propositions. The following two are the rudimentary elucidations of linearity.

If a system is described to have an input as $x[n] = ax[n1] + bx[n2]$, then the linear system reverberation should be $y[n] = ay[n1] + by[n2]$. This is known to be the superposition principle, considered as most crucial to a linear system.

The second property is shift invariance. If $y[n]$ is retaliation to a linear, shift invariance system with an input $x[n]$, then $y[n - n0]$ should be the retaliation to a system with input $x[n - n0]$.

Two additional conditions are seminal and stability, the seminal condition is needed when considering systems in which future values are not unknown. To keep a filter's output within a finite limit, a stability should be established for any given input which does not exceed the limit.

2.1.2 Nonlinear Filters

Nonlinear filters exhibit a different behavior than linear filter, that is, the filter response does not adhere to the principles listed above, especially linearity and shift variance. Also, a nonlinear filter produces results that can vary in an unpredictable manner. A simple example is considered for a nonlinear filter that used a median filter.

The following example also compares the two types of filters on grounds of working of the two types of filters.

Consider a filter based on five values. In the region of interest, $x0 \ldots x4$, the values are in an ascending order. The value at position 2 is chosen as output.

- In low frequency, almost all the values are either same or somewhat close to it. Thus, the value that is selected will be either actual value plus small error or actual value minus small error.
- In case of high frequency, like an edge, all the values on one side of the edge will be small while all the values on the other side will be large. When the orientation is done, the low values will be in low position and the high values will be in high

Fig. 7 Noisy image going through Gaussian and median filter [7]

position. The middle value selected will be either in the high group or the low group, as would be the case while using a low-pass linear filter.
- For this reason, this category of filter is also called the edge-preserving filter, and hence, it is highly useful in removing the impulse noise (Fig. 7).

3 Selection of the Complaisant Filter

Both the filters have a crucial place in the timeline of image processing and both are a preprocessing step. There can be dozens of both types included to create, shape, encounter, and chisel data related to the image. They can be customized to perform in one way under certain conditions and to perform differently when exposed to another set of conditions adhering to the adaptive filter rule generation.

An adaptive filter is a linear filtered system that has a transfer function measured by mutable specifications and a mode to calibrate those parameters as directed by an optimization algorithm. This filter can also work with the nonlinear filter.

The goals of filtering image data fluctuate from noise removal to feature abstraction. Filter construction has mainly two forms: (a) linear and (b) nonlinear. The selection of a relevant filter depends solely on the targets and disposition of the image data that has to be restored.

- Cases in which the input image data comprises a huge amount of noise but the degree of measure is low, a linear low-pass filter may be adequate.
- Contrariwise, if an input has a low amount of noise but with a comparably high magnitude, then a nonlinear filter may be more relevant.
- While linear methods are fast enough, they do not conserve the details of the image, whereas nonlinear filters retain the details well enough.

4 Various Types of Filter

4.1 Mean Filter

It is defined as averaging a linear filter. In this form, filter calculates the mean or average value of all the distorted images in an area which is already pre-decided. Then, the center pixel intensity value is substituted using that mean value. This procedure is performed iteratively on all pixels of the image.

4.2 Median Filter

The median filter is best order statistic and nonlinear filter, whose reciprocation is constructed on the grading of pixel values present in the filter region. In this, the nucleus pixel value is substituted by the intermediate pixel values in the filter vicinity. The median filter performs well for salt-and-pepper noise. The central asset of a median filter is that it can remove input noise with large magnitudes (Fig. 8).

Fig. 8 Effect of nonlinear median filter [8]

4.3 Order Statistics Filter

It is a type of nonlinear filter whose riposte is determined by the sequence of pixels covering filtering area. When the nucleus value of the pixel in a picture region is substituted by the 100th percentile, that methodology is termed as max-filter. While on the other hand, if the same pixel value is substituted by the zero percentile, then the filter is considered as the min-filter.

4.4 Adaptive Filter

The behavior of these filters is transmuted based on the demographic traits of the image region, beset by the filter region.

For example, BM3D is sort of adaptive filter. The de-noising algorithm of the adaptive filter can be divided into three steps:

1. Analysis,
2. Processing, and
3. Synthesis.

4.5 Wiener Filter

The Wiener filter intends to eradicate noise which has altered a signal. It follows a statistical approach.

5 Implementation of Median Filter

See Fig. 9.

6 Conclusion

1. The achievement of the Wiener filter after performing noise removal for all speckle, Poisson, and Gaussian is of higher quality than mean filter and median filter.
2. The median filter works better for salt-and-pepper noise than mean and Wiener filter.
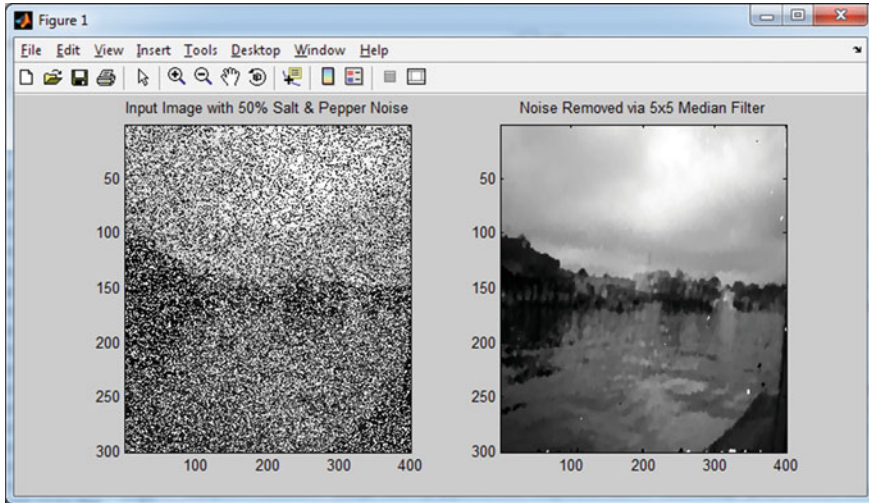
Fig. 9 Effect of nonlinear median filter [9]

All the available filters have their specific benefits and disadvantages. One type of filter may work very well with a certain genre of noise but not so good with another kind in the same image. There is a lot of scope for research in this field. Here, we have discussed the various grades of noise that can affect an image during processing and acquisition. Light is thrown on the various ways available to de-noise these images. It can be concluded by saying that using a method that can well preserve the edges of an image is more important than using a method which is fast though the speed factor is equally important. Another technique being used these days is the application of fuzzy systems in filters [10]. Fuzzy logic is related to vagueness and in case of image noise, certain data can be vague [11], [12]; thus, fuzzy systems can work well in this field.

References

1. https://www.google.co.in/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwjqjfCa-DdAhUJaI8KHY4DB6UQjRx6BAgBEAU&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FImage-03-with-Gaussian-noise-s-5_fig3_291043468&psig=AOvVaw2JU2KcbwyxXoczjILQ4LPG&ust=1536137213527988
2. http://www.fit.vutbr.cz/~vasicek/imagedb/img_corrupted/impnoise_005/108073.png
3. https://homepages.inf.ed.ac.uk/rbf/HIPR2/images/fce5noi4.gif
4. https://www.google.co.in/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uac=8&ved=2ahUKEwiX1LvFgKHdAhUQTo8KHWW1DMkQjRx6BAgBEAU&url=https%3A%2F%2Fwww.researchgate.net%2Ffigure%2FVisual-results-for-the-Barbara-image-with-JPEG-quantization-noise-Q-7_fig3_220320603&psig=AOvVaw2auKTt5fZpJC20RiwX7082&ust=1536138644050333

5. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTjA8FQQ36oYjEHr4TbwnhLo4rp R0Kf-qwP0sk0lmpEi1NNZ3bI

6. Verma R, Ali J (2013) A Comparative study of various types of image noise and efficient noise removal techniques. IJARCSSE, (2277-128X) 3(10)

7. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcR84YI5yGqy8yZkPcE-KQHxDN vhkTq9R5JVvCldlIX5CB6mXhcOHA

8. https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRlz4KKImcLE5QKNJTAsB0Umg 1LURICLBVIBg-g8kFaMWFVMBax

9. https://www.google.co.in/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact =8&ved=2ahUKEwiQlN3h-DdAhWMq48KHUcCAnsQjRx6BAgBEAU&url=https%3A %2F%2Fwww.giassa.net%2F%3Fpage_id%3D635&psig=AOvVaw2956XMBL6dsaOcTCZD 0AEL&ust=1536137338290174

10. Mehan S, Singla N (2012) Introduction of image restoration using fuzzy filtering. IJARCSS (2277-128X) 2(3)

11. Mittal A, Goel V (2012) Removal of impulse noise using fuzzy techniques. IJAER (0973-4562) 7(11)

12. Ville DVD, Nachtegael M, Weken DV, Kerre EE, Philips W, Lamahieu I (2003) Noise reduction by using fuzzy image filtering. IEEE Trans Fuzzy Syst 11(4)

Realization of Communication System with Remote Access Using LabView

Harsh Anand, Shravan Cunchala, Debdatta Biswas and Soumyasree Bera

Abstract The aim of the paper is to develop a communication system using LabView (Laboratory Virtual Instrument Engineering Workbench). With the help of remote sensing, any system can be controlled by PC wherever placed without operating the required instruments. The goal is to implement the system in a way which will have minimal cost, and the system is accessible anywhere.

Keywords Transmitter · Receiver · Modulation · Demodulation · BER · Random fading · Multipath delay spread

1 Introduction

When information is exchanged between two points, it is known as communication. A communication system consists of transmitter, transducer, channel, and receiver. The message is transmitted from the transmitter, and it is received at the receiver. A message is transmitted through the channel. This will allow communication from any place across the world using small devices.

The message is originated from the source which can be voice, picture or data. The input transducer helps to convert the message signal to electrical signal which is known as baseband signal. The transmitter is used to convert the baseband signal to channel friendly signal. The baseband signal is transmitted through a medium known as channel. The channel could be a wire, a coaxial cable, an optical fiber, wireless, etc.

H. Anand · S. Cunchala · D. Biswas · S. Bera (✉)
Department of Electronics & Communication, Sikkim Manipal Institute of Technology, Majitar, Rangpo 737136, Sikkim, India
e-mail: soumyasree.bera@gmail.com

Based on the channel type, communication is of two types: wired communication and wireless communication. The signal is received at the receiver and further processed. The message signal is extracted from the noisy signal. The receiver output is given to the output transducer which converts the electrical signal back to its original form [1, 2].

2 Extraction of Message Bit and BER Calculation

Triggering is another important aspect of communication. Triggering is used to identify the start bit (start of the message signal) and the stop bit (end of the message signal) in the message signal. In the process of triggering a particular trigger level is chosen. It is also used to extract the signal at the receiver side. The triggering level shows the start of the signal. One frame of a signal consists of sync bit (start of signal), training bits, message signal, and sync bit (end of the signal after that there is another frame of message signal). Initially, we know the length and the time duration of the sync and the training bits. So for extraction of the training bit, we add up the time duration of the sync bit and specify the length of the training bits we are able to extract the training bits, and the same process is followed for extraction of message signal [3–5].

3 Modulation

Modulation is an important key in the transmission of message from transmitter to receiver. Modulation is of two types: analog modulation and digital modulation. Modulation techniques should be chosen in such a way that it gives better spectral efficiency also less bit error rate. Higher the number of bit gives good spectral efficiency (transmission of more number of bits within the given bandwidth) but has higher bit error rates Modulation is the process in which the property (amplitude, frequency, and phase) of a signal is varied. Modulation is done to achieve better spectral efficiency, reduction in the height of antenna, avoiding mixing of signals [3, 6].

Some of the digital modulation techniques used are ASK, PSK, FSK, etc. Out of ASK, PSK, and FSK, PSK is the most advantageous to use. ASK is more sensitive to noise, as noise affects the modulation so in ASK the amplitude gets distorted. PSK requires higher bandwidth to transmit the data. PSK overcomes the disadvantages of ASK and FSK. PSK has constant amplitude and frequency. For modulation through PSK, 8 PSK is most widely used. For calculation of bandwidth of message signal, the circuit was designed [7].

4 For Calculation of Bandwidth of Message Signal, Circuit Was Designed

To design the circuit for calculation of the bandwidth of the message signal, the P4 code signal is taken as input which is converted from time domain to frequency domain by taking its fast Fourier transform. After passing through the FFT block, the absolute value of the signal is taken using the abs block and then a unit converter is used for conversion to dB scale, which is attached to a graph indicator which displays the output waveform. Here, the sampling frequency considered is 100 Hz and the code length of P4 is taken as 100.

5 Methodology

5.1 Block Diagram

- Message signal (P4 code) is generated in LabView.
- The message signal is modulated using 8 PSK modulation technique; this step is carried out in LabView.
- The modulated message signal is dumped in hardware for generation. Hardware used for generation of the signal is function generator.
- The generated signal is viewed through digital signal oscilloscope (DSO).
- Signal from DSO is taken back (data acquisition) to LabView for further processing.
- The received signal is demodulated and output is received in LabView (Fig. 1).
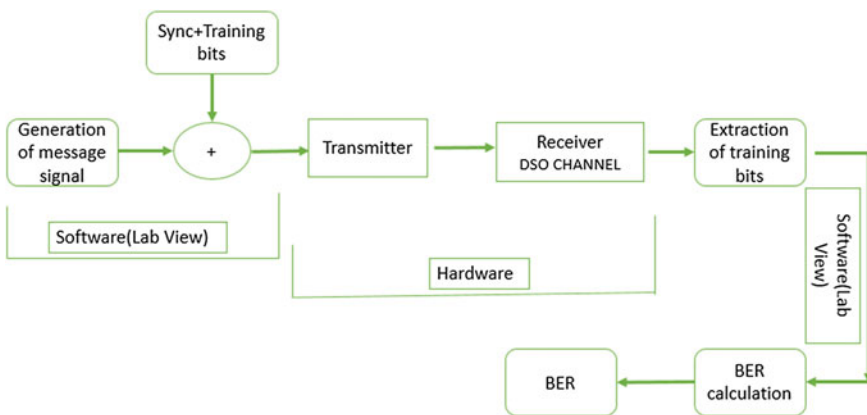
Fig. 1 Block diagram of the proposed system

6 Result Analysis

6.1 Generation and Transmission of Message Signal

While designing a communication system our first task is to establish communication between transmitter and receiver. So for this purpose before transmitting any message signal (Fig. 4), sync bits are appended with the message signal. Doing so the receiver comes to know that data is coming toward it and same sync bit is also present at the receiver so that the receiver should only accept the desired signal and not any other signal. In the system, the sync bit is developed by the developer which is fixed and the message signal is generated by the user which can be varied. If there is any change in the sync bit at the transmitter the same should be informed at the receiver for proper synchronization of the message signal. A message signal directly cannot be used for synchronization because the message signal is not fixed. The sequence of the sync bit is "0101" (Figs. 2, 3, 4 and 5).

6.2 Extraction of Message Signal and BER Calculation

To find the efficiency of the system BER is calculated. Higher the BER, lesser the efficient our system. So for calculating BER along with message and sync bit training bits are appended. At the receiver side, the training bits are extracted, and it is compared with the reference bits which were appended at the receiver side. The extracted training bits and the reference bits are compared and BER of the system is calculated. The sequence of the training bits is "1011" (Fig. 6).
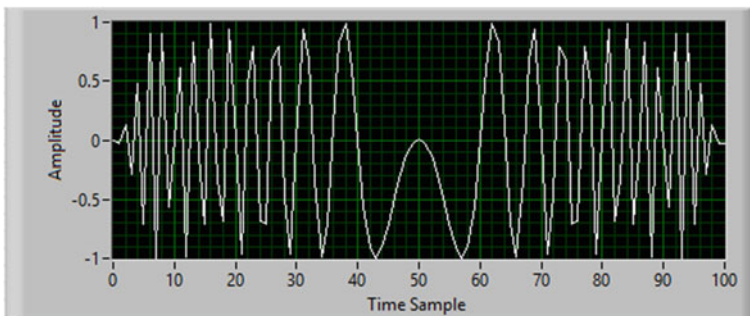
Fig. 2 Transmitted message signal (P4 code)
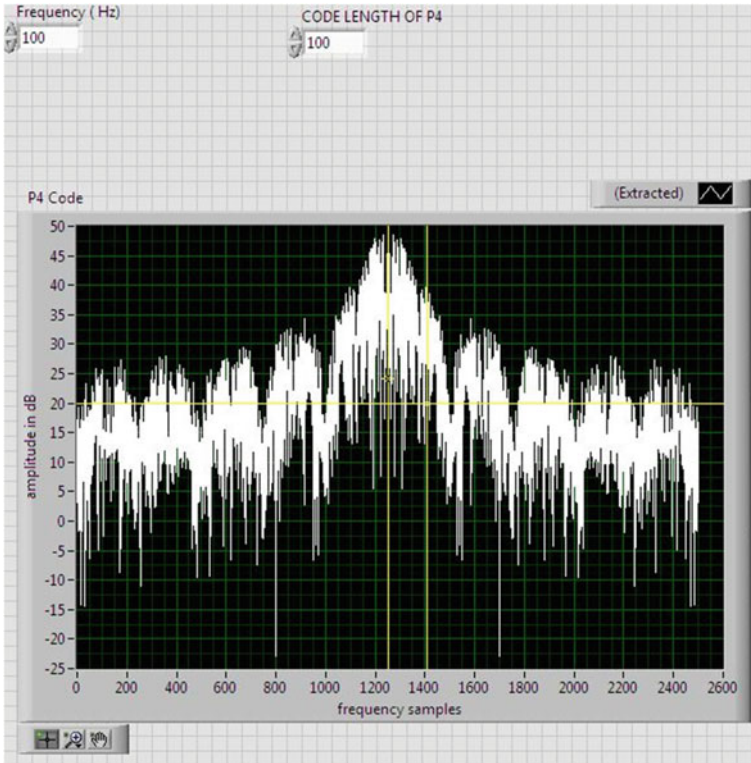
Fig. 3 Spectrum of the message signal

Fig. 4 Transmitted message frame
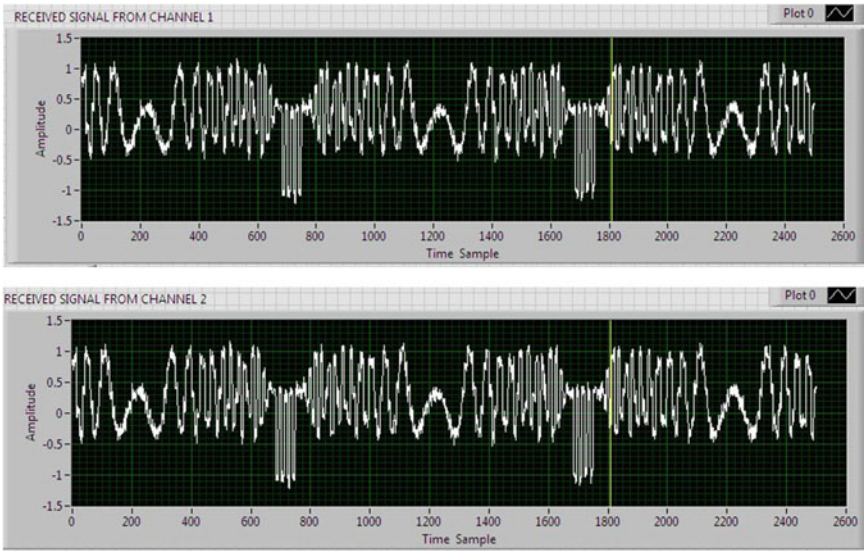
Fig. 5 Received message signal
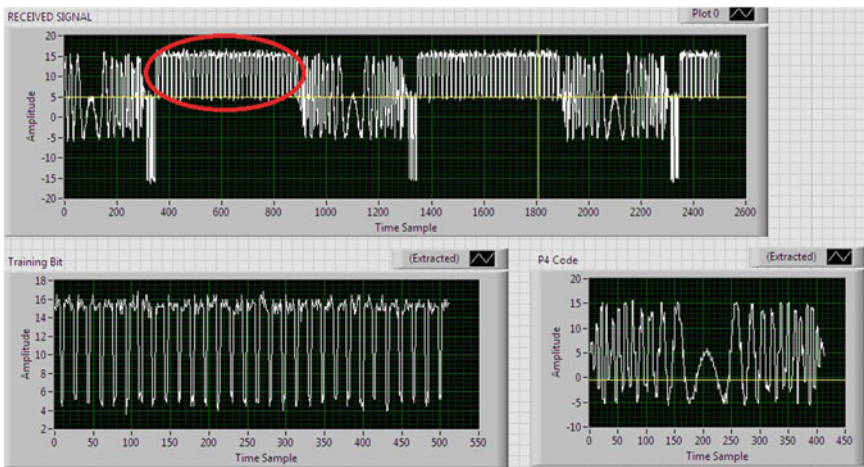
Fig. 6 Extraction of message and training bits from received signal and BER calculation (noise amplitude—0.1 V, BER—0.0110156)

7 Modulation

Figure 7 shows the user input panel. The user giver input to the modulator and gets output from the demodulator.

Figure 8 represents the constellation diagram for 8 PSK modulation. 8 PSK has less bit error rate and also it gives good spectral efficiency. In 8 PSK modulation, 3 bits are transmitted in a symbol. The symbols are mapped to complex sequence. In the demodulator side, the complex sequence is taken as input and the bits are demodulated. PSK overcomes the disadvantages of ASK and FSK. PSK has constant amplitude and frequency.
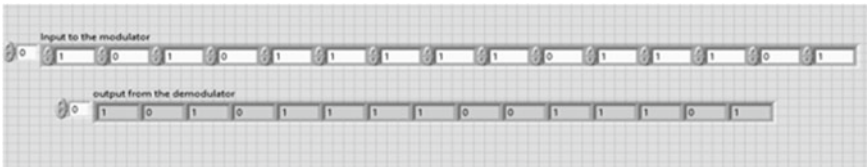
Fig. 7 Input and output of modulator and demodulator
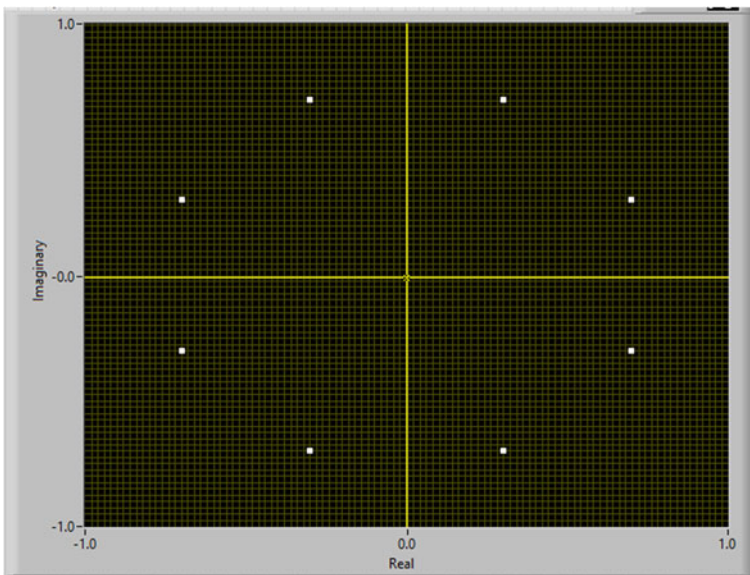
Fig. 8 Constellation diagram for 8 PSK
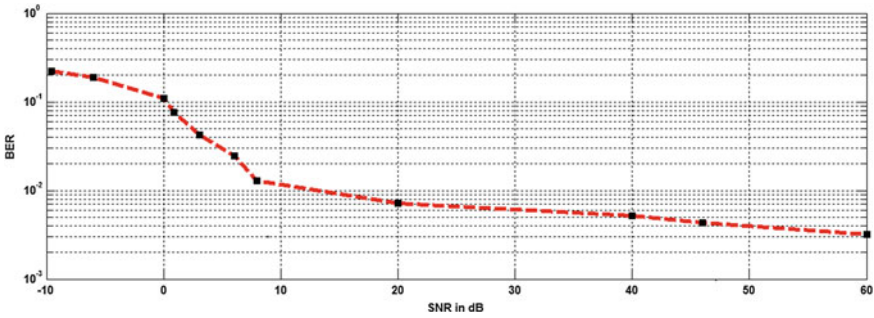
Fig. 9 SNR versus BER plot

7.1   SNR Versus BER Plot

SNR known as signal-to-noise ratio is the ratio of signal power to noise power. It is an important factor for comparing the BER (Bit error rate) of any system using any modulation technique. For the relation between SNR and BER, SNR is inversely proportional to BER, i.e., as the SNR increases, BER decreases for an efficient system. SNR of any system is represented by (Fig. 9)

$$\text{SNR (dB)} = 10 \log_{10} \frac{S}{N} \tag{1}$$

where S is signal power and N is noise power.

8   Conclusion

From this paper, it is seen that communication is established between transmitter and receiver. The message signal is generated in LabView. Proper synchronization was observed between transmitter and receiver by the help of sync bit appended with the message signal. The efficiency of the system was observed by calculating the BER, and it observed that the BER was found to be 0.0032 at SNR 60 dB, which shows that the system has good efficiency. From the SNR vs. BER plot, it was observed that SNR is inversely proportional to BER, last but not the least remote access of the communication system was established through PC in LabView.

References

1. Amraoui A, Benmammar B, Fethi FK, Bendimerad T (2012) Intelligent wireless communication system. Int J Distrib Parallel Syst 3(2):181–187
2. Shah SK, Rane SJ, Vishwakarma DD (2012) Performance evaluation of wired and wireless local area networks. Int J Eng Res Dev 1(11):43–48. ISSN: 2278-067X
3. Singh H, Singh M (2012) Optimization method for analysis of bit error rate with BPSK modulation technique. Int J Sci Eng Res 3(6):1–4
4. Bohra DD, Bora A (2014) Bit error rate analysis in simulation of digital communication systems with different modulation schemes. Int J Sci Eng Res 1(3):406–413
5. Pandey R, Awasthi A, Srivastava V (2013) Comparison between bit error rate and signal to noise ratio in OFDM using LSE algorithm. In: Proceedings of the conference on advances in communication and control systems 2013 (CAC2S 2013), pp 463–466
6. Tewari J, Singh HM (2016) Performance comparison of digital modulation techniques used in wireless communication system. Int J Innov Res Comput Commun Eng 4(7):12425–12431
7. Singh H, Arora M (2015) Comparison of bit error rate performance of different modulation techniques over turbulent FSO link. Int J Comput Appl (0975-8887) 109(12):20–24

Rational Approximation Methods for a Class of Fractional-Order SISO System in Delta Domain

Jaydeep Swarnakar, Prasanta Sarkar and Lairenlakpam Joyprakash Singh

Abstract Rational approximation is fundamental for practical implementation of infinite dimensional fractional-order system in the area of system and control. In this paper, a class of fractional-order single-input-single-output (SISO) system has been realized in delta-domain using two rational approximation techniques, namely continued fraction expansion (CFE) approximation and Oustaloup approximation, respectively. Both the techniques use indirect discretization method for obtaining the integer-order approximants of the ideal fractional-order SISO system in delta-domain. The frequency responses of two discretized integer-order approximants of the original fractional-order system are compared subsequently using MATLAB.

Keywords Fractional-order system · Oustaloup approximation · CFE approximation · Discretization · Delta operator · Delta-domain

1 Introduction

Most of the natural systems around us are typically of fractional-order. Although the idea of fractional calculus was initiated over 300 years back [1], the topic has earned significant relevance even today among the researchers in many fields of science and engineering. In the field of system and control, the fractional-order controllers have been designed with greater flexibility and robustness due to their higher

J. Swarnakar (✉) · L. J. Singh
Department of Electronics & Communication Engineering, School of Technology, North-Eastern Hill University, Shillong 793022, Meghalaya, India
e-mail: jaydeepswarnakar@gmail.com

L. J. Singh
e-mail: jplairen@gmail.com

P. Sarkar
Department of Electrical Engineering, National Institute of Technical Teachers' Training and Research, Kolkata 700106, West Bengal, India
e-mail: psarkar@nitttrkol.ac.in

degree of freedom [2]. The basic component of any fractional-order system is either fractional differentiator or fractional integrator. For implementation of fractional differentiator/integrator, one needs infinite memory requirements which is not feasible at all in a practical scenario. Therefore, an efficient rational approximation method is desirable to realize the fractional-order system within a limited band. Rational approximation methods are of two types, namely continuous-time approximation and discrete-time approximation. For continuous-time approximation, we cite [3, 4]. For digital realization of fractional-order system, discrete-time approximation or discretization is essential. Discretization comprises of indirect method and direct method, respectively. Indirect approach replaces the fractional-order model by its equivalent integer-order model in S-domain using continuous-time rational approximation method, and subsequently the fit continuous-time system is discretized [5, 6]. For direct approach, we refer [7, 8].

Discrete-time systems are modelled using shift operator in general. At fast sampling rate, shift operator based modelling fails to reproduce the meaningful description of the original continuous-time system. Delta operator was introduced to alleviate this problem which is an alternate modelling formulated on the principle of signal differencing as oppose to signal shifting [9]. Delta operator based modelling shows resemblance of its continuous-time counterparts at the lower sampling time limit. Therefore, both the analog and digital system can be studied under a unified framework capitalizing the feature of delta operator.

In this work, an indirect discretization method is presented to realize a class of fractional-order SISO system in delta-domain using two rational approximation methods. Section 1 gives the introduction. Section 2 focusses on the two rational approximation techniques of fractional-order system under a unified framework. The results and simulation are deliberated in Sect. 3. The entire work is concluded with some remarks in Sect. 4.

2 Rational Approximation of Fractional-Order SISO System

Let a SISO system is represented by the following fractional-order transfer function:

$$G(s) = \frac{1}{h_p s^{\rho_k} + h_{p-1} s^{\rho_{k-1}} + h_{p-2} s^{\rho_{k-2}} + \cdots + h_1 s^{\rho_1} + h_0} \tag{1}$$

where h_i ($i = 0, 1, 2 \ldots p$) are the real numbers and s^{ρ_j} ($j = 0, 1, 2 \ldots k$) are the various fractional power terms. It is necessary to find the integer-order approximation of each fractional power term separately existing in Eq. (1) to obtain the rational approximation of $G(s)$ as a whole. In this work, $G(s)$ has been approximated by using CFE method [10] and generalized Oustaloup approximation method [1]. Both the methods handle each fractional power term (s^{ρ_j} ($j = 0, 1, 2 \ldots k$) separately

to obtain the integer-order approximation of $G(s)$. The mathematical expression of CFE approximation method is specified in Eq. (2).

$$(1 + h)^\rho = 1 + \cfrac{\rho h}{1 + \cfrac{(1-\rho)h}{2 + \cfrac{(1+\rho)h}{3 + \cfrac{(2-\rho)h}{2 + \cfrac{(2+\rho)h}{5 + \cfrac{(3-\rho)h}{2 + \cdots}}}}}} \tag{2}$$

Substituting $h = s - 1$ in Eq. (2), the third-order approximation of each fractional power term is obtained as follows:

$$s^\rho = \frac{\alpha_3 s^3 + \alpha_2 s^2 + \alpha_1 s + \alpha_0}{\beta_3 s^3 + \beta_2 s^2 + \beta_1 s + \beta_0} \tag{3}$$

where,

$$\alpha_3 = \beta_0 = \rho^3 + 6\rho^2 + 11\rho + 6$$
$$\alpha_2 = \beta_1 = -3\rho^3 - 6\rho^2 + 27\rho + 54$$
$$\alpha_1 = \beta_2 = 3\rho^3 - 6\rho^2 - 27\rho + 54$$
$$\alpha_0 = \beta_3 = -\rho^3 + 6\rho^2 - 11\rho + 6$$

Using generalized Oustaloup approximation technique, the Jth order approximant of each fractional power term is derived within a frequency range (ω_e, ω_f) as specified in Eq. (4):

$$s^\rho = M \frac{\left(s + \omega_1'\right)\left(s + \omega_2'\right)\cdots\left(s + \omega_J'\right)}{(s + \omega_1)(s + \omega_2)\cdots(s + \omega_J)} \tag{4}$$

where ω_i $(i = 0, 1, 2 \ldots J)$ and ω_i' $(i = 0, 1, 2 \ldots J)$ are the recursively distributed poles and zeros, $M = \omega_f^\rho$, $\omega_i' = \omega_e \omega_q^{(2i-\rho-1)/J}$, $\omega_i = \omega_e \omega_q^{(2i+\rho-1)/J}$ and $\omega_q = \sqrt{\omega_f / \omega_e}$.

Using Eqs. (3) and (4), $G(s)$ is approximated to $G_{c1}(s)$ and $G_{c2}(s)$ using CFE approximation and Oustaloup approximation respectively. Discretization is accomplished subsequently to realize the fractional-order system in delta-domain. In this work, the indirect discretization method is applied in delta-domain on two integer-order approximants obtained from two continuous-time rational approximation methods of s-domain. The δ operator is represented by the following mathematical expression:

$$\delta = \frac{q - 1}{\Delta} \tag{5}$$

where Δ represents the sampling period and q is the forward shift operator. Applying δ-operator on $x(t)$ and considering the lower sampling time limit $\Delta \to 0$, it converges to the corresponding continuous-time operator as exhibited in Eq. (6).

$$\underset{\Delta \to 0}{Lt}\, \delta x(t) = \underset{\Delta \to 0}{Lt}\, \frac{x(t + \Delta) - x(t)}{\Delta} = \frac{d}{dt}x(t) \tag{6}$$

The frequency variable of delta-domain (γ) and the frequency variable of conventional discrete-time domain (z) uphold the similar relationship indicated in Eq. (7):

$$\gamma = \frac{z - 1}{\Delta} \tag{7}$$

The discretization formula in δ-domain is developed by rearranging Eq. (7) after replacing $z = e^{s\Delta}$ as specified in Eq. (8):

$$s = \Delta^{-1} \ln(\Delta\gamma + 1) \tag{8}$$

Using Eq. (8), $G_{c1}(s)$ and $G_{c2}(s)$ are discretized in delta-domain to attain the corresponding discretized transfer functions $G_{\delta1}(\gamma)$ and $G_{\delta2}(\gamma)$, respectively. $G_{\delta1}(\gamma)$ originates from the indirect discretization of ideal fractional-order system using CFE approximation (CFE + Delta) method, whereas $G_{\delta2}(\gamma)$ originates from the indirect discretization of fractional-order system using Oustaloup approximation (Oustaloup + Delta) method. The frequency responses of the discretized systems $G_{\delta1}(\gamma)$ and $G_{\delta2}(\gamma)$ are compared at the lower sampling time limit with respect to the frequency response characteristics acquired from their respective continuous-time counterparts.

3 Simulation and Results

The following fractional-order SISO system [1] is taken as an example:

$$G(s) = \frac{1}{h_2 s^{\rho_2} + h_1 s^{\rho_1} + h_0} \tag{9}$$

where $h_2 = 1$, $h_1 = 0.75$, $h_0 = 0.9$, $\rho_2 = 0.8$ and $\rho_1 = 0.4$. So, $G(s)$ can be written as follows:

$$G(s) = \frac{1}{s^{0.8} + 0.75s^{0.4} + 0.9} \tag{10}$$

$G(s)$ contains two fractional power terms $s^{0.8}$ and $s^{0.4}$, respectively. Using Eqs. (3) and (4), the integer-order (third-order) approximate of fractional power terms are calculated using CFE method and Oustaloup method within a frequency range 0.001–1000 rad/s as presented in Table 1.

Table 1 Rational approximation of the fractional power terms in s-domain

Fractional power term	Oustaloup approximation	CFE approximation
$s^{0.8}$	$\dfrac{251.2s^3+4021s^2+637.3s+1}{s^3+637.3s^2+4021s+251.2}$	$\dfrac{19.15s^3+70.22s^2+30.1s+0.528}{0.528s^3+30.1s^2+70.22s+19.15}$
$s^{0.4}$	$\dfrac{15.85s^3+637.3s^2+253.7s+1}{s^3+253.7s^2+637.3s+15.85}$	$\dfrac{11.42s^3+63.65s^2+42.43s+2.496}{2.496s^3+42.43s^2+63.65s+11.42}$

Table 2 Rational approximation of the fractional-order SISO system in s-domain

Method	Integer-order approximants of $G(s)$
CFE	$G_{c1}(s) = \dfrac{1.318s^6+97.52s^5+1486s^4+4949s^3+5626s^2+2021s+218.8}{53.51s^6+1359s^5+7666s^4+14900s^3+11010s^2+2937s+238.8}$
Oustaloup	$G_{c2}(s) = \dfrac{s^6+891.1s^5+166400s^4+1427000s^3+2637000s^2+223800s+3981}{264\,s^6+76610\,s^5+1683000\,s^4+6059000\,s^3+3729000\,s^2+263000\,s+3787}$

Table 3 Rational approximation of the fractional-order SISO system in delta-domain for $\Delta = 0.0001$ s

Method	Discretization of $G(s)$ in delta-domain
CFE + Delta	$G_{\delta1}(\gamma) = \dfrac{0.02463\,\gamma^6+1.822\,\gamma^5+27.74\,\gamma^4+92.39\,\gamma^3+105\,\gamma^2+37.72\,\gamma+4.084}{\gamma^6+25.38\gamma^5+143.1\gamma^4+278.1\gamma^3+205.5\gamma^2+54.82\gamma+4.457}$
Oustaloup + Delta	$G_{\delta2}(\gamma) = \dfrac{0.003788\gamma^6+3.359\gamma^5+621.8\gamma^4+5329\gamma^3+9845\gamma^2+835.5\gamma+14.86}{\gamma^6+286.6\gamma^5+6287\gamma^4+22620\gamma^3+13920\gamma^2+981.9\gamma+14.14}$

Now, by substituting the corresponding integer-order approximants of the fractional power terms in the Eq. (10), $G_{c1}(s)$ and $G_{c2}(s)$ are obtained which represent the CFE approximation and the generalized Oustaloup approximation of $G(s)$ respectively as shown in Table 2. $G_{c1}(s)$ and $G_{c2}(s)$ are discretized to obtain $G_{\delta1}(\gamma)$ and $G_{\delta2}(\gamma)$, respectively, in delta-domain by taking $\Delta = 0.0001$ s as exhibited in Table 3.

Figure 1 shows the frequency responses of the ideal fractional-order system and its discrete-time approximations obtained using "Oustaloup + Delta" method and "CFE + Delta" method respectively for $\Delta = 0.0001$ s. Both the approximation techniques provide satisfactory result especially in terms of magnitude approximation. "Oustaloup + Delta" method provides better phase approximation in the boundary range and produces some ripple in the mid-frequency range. "CFE + Delta" method produces better magnitude fit and phase fit in the mid-frequency range while showing some inconsistency in terms of phase approximation at the boundary range of frequencies. Figure 2 shows the frequency response of the original fractional-order system and its continuous-time approximations obtained from the CFE method and the Oustaloup method respectively.

From the continuous-time approximations, it is apparent that the Oustaloup approximation shows a moderate magnitude approximation and phase approximation throughout the entire range of frequency while the CFE approximation works better particularly in the mid-frequency range and exhibits a larger amount of phase error near the boundary of the desired frequency range. From Tables 4 and 5, it is also apparent that the errors obtained from the discrete-time approximations and the
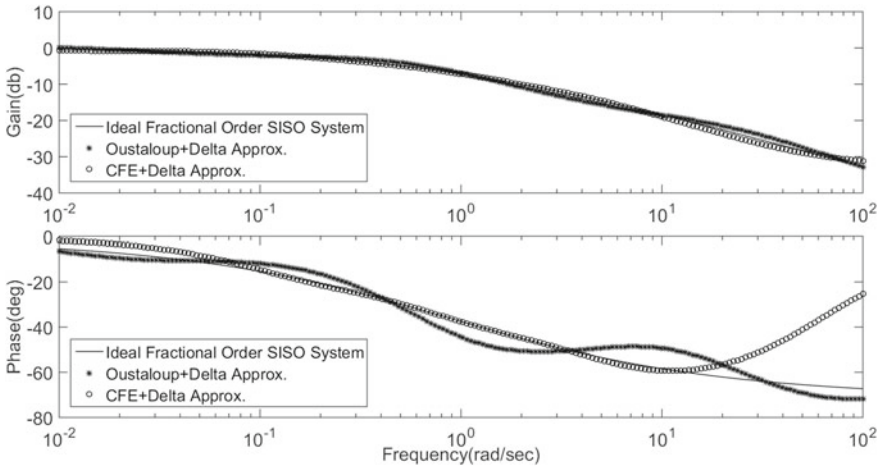
Fig. 1 Frequency response comparison in delta-domain for $\Delta = 0.0001$ s
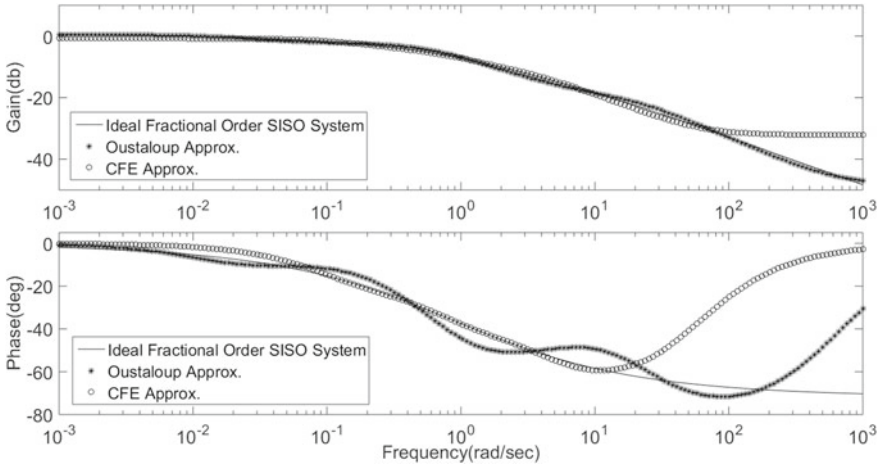
Fig. 2 Frequency response comparison in s-domain

continuous-time approximations of the original fractional-order system are almost

Table 4 Magnitude approximation errors and phase approximation errors obtained in delta-domain for $\Delta = 0.0001$ s

Method	Maximum absolute errors in delta-domain	
	Magnitude error (dB)	Phase error (°)
Oustaloup + Delta	1.26	38.9
CFE + Delta	16.16	67.5

Table 5 Magnitude approximation errors and phase approximation errors obtained in s-domain

Method	Maximum absolute errors in delta-domain	
	Magnitude error (dB)	Phase error (°)
Oustaloup	1.3	40
CFE	16.2	67.5

analogous in nature. This clearly signifies that the discrete-time systems converge to their continuous-time counterparts at the lower sampling time limit $\Delta = 0.0001$ s, which emphasizes the effectiveness prevailing in the delta-domain modelling of discrete-time systems.

4 Conclusion

In this work, two rational approximation methods are presented for realizing a class of fractional-order SISO system in delta-domain which incorporate CFE approximation and Oustaloup approximation to acquire the approximate integer-order models of the ideal fractional-order system separately. Both the integer-order approximants are subsequently discretized in delta-domain using "Oustaloup + Delta" approximation and "CFE + Delta" approximation respectively. "Oustaloup + Delta" approximation maintains a moderate magnitude approximation and phase approximation throughout the entire frequency range and indicates better magnitude and phase fit near the boundary frequency range compared to 'CFE + Delta' approximation. In contrary, 'CFE + Delta' approximation exhibits consistent magnitude approximation and phase approximation in the mid-frequency range while showing larger errors nearby the boundary region. At fast sampling rate or lower sampling time limit, both the rational approximation methods of delta-domain produce similar results as that of obtained in s-domain. Therefore, the proposed approach has implication in some sense to unify the study of both continuous-time system and discrete-time system together at the high sampling frequency limit.

References

1. Monje CA, Chen YQ, Vinagre BM, Xue D, Feliu V (2010) Fractional-order systems and controls. Springer, London
2. Chen YQ, Petras I, Xue D (2009) Fractional-order control—a tutorial. In: Proceeding of the American control conference (ACC), St. Louis, USA, pp 1397–1411
3. Vinagre BM, Podlubny I, Hernande A, Feliu V (2000) Some approximations of fractional-order operators used in control theory and applications. J Fract Calc Appl Anal 3:231–248
4. Xue D, Zhao C, Chen YQ (2006) A modified approximation method of fractional-order system. In: Proceedings of the IEEE international conference on mechatronics and automation (ICMA), Luoyang, China, pp 1043–1048
5. Krishna BT, Reddy KVVS (2008) Design of fractional-order digital differentiators and integrators using indirect discretization. Int J Theory Appl 25(33):143–151
6. Maione G (2011) High-speed digital realizations of fractional operators in the delta-domain. IEEE Trans Autom Control 56:697–702
7. Chen YQ, Moore KL (2002) Discretization schemes for fractional-order differentiators and integrators. IEEE Trans Circuits Syst I Fundam Theory Appl 49:363–367
8. Vinagre BM, Chen YQ, Petras I (2003) Two direct Tustin discretization methods for fractional-order differentiator/integrator. J Frankl Inst 340(5):349–362
9. Middleton RH, Goodwin GC (1990) Digital control and estimation: a unified approach. Prentice Hall, Englewood Cliffs
10. Khovanskii AN (1963) The application of continued functions and their generalizations to problems in approximation theory (Trans Wynn P). P. Noordhoff Ltd

Utilization of Electric Vehicle for LFC of Multi-area System Using Wind Driven Optimized Classical Controllers

Pushpa Gaur, Nirmala Soren and Debashish Bhowmik

Abstract This paper focuses on load frequency control (LFC) of an integrated unequal three area thermal system with the incorporation of electric vehicle (EV) in all three areas. An appropriate generation rate constraint of 3% per minute is considered in each of the three areas. The performance of conventional controllers like Integral (I), Proportional-Integral (PI) and Proportional-Integral-Derivative (PID) controllers are evaluated on the system. A new nature-inspired optimization technique named as Wind Driven Optimization (WDO) technique is utilized for optimizing parameters of the controllers. Comparison between the responses obtained by using I, PI and PID discloses the superior performance of PID controller from the point of view of time of dying out of response, highest deviation and degree of oscillations. The study has been stretched to the application of random loading in Area-1 in place of step load. Also, an analysis is carried out to evaluate the impact of incorporation of EV in the system, which reveals the improvement of system dynamics with the incorporation of EV.

Keywords Load frequency control · Electric vehicle · Wind driven optimization algorithm · Generation rate constraint · Classical controllers

1 Introduction

The uninterrupted operation of an integrated power system demands the balance between power-generated consumer demand along with the associated losses occurring in the system. A minor alteration in load demand may result in deviation in

P. Gaur (✉) · N. Soren
Electrical Engineering Department, National Institute of Technology Silchar, Silchar, Assam, India
e-mail: pushpa_gaur@yahoo.com

D. Bhowmik
Electrical Engineering Department, Tripura Institute of Technology Narsingarh, Agartala, Tripura, India

nominal frequency and power exchange amongst connected areas, leading to adverse effects. Load frequency control (LFC) comes into play when there is any mismatch in this power balance. The enormous amount of work has been found on LFC of multi-area systems. The concept of LFC was given by Elgerd [1, 2], which is carried forward by many authors. Nanda et al. [3] have taken conventional hydrothermal plants for study. Pham et al. [4] presented the LFC of multi-area system with the effect of electric vehicle. Barisal et al. [5] presented a two area multi-source system incorporating thermal, hydro and gas systems in a single area. It is evident from the literature that very limited work has been done on LFC of thermal system with EV. Hence, this provides further scope for investigation.

In an automatic generation control (AGC) study, both primary and secondary controls are applied at the same time. AGC mainly focuses on the design and implementation of secondary controller. Most of the reported studies are based on utilization of classical controllers, namely, Integral (I), Proportional-Integral (PI) and Proportional- Integral-Derivative (PID). The working of I and PI controller has been tested on a hydrothermal system in the presence and absence of electric governor [3]. Some researchers have employed artificial intelligence based controllers like fuzzy logic (FL) and artificial neural network (ANN) controllers as a supplementary controller in their studies [6, 7]. But, the disadvantage of FL and ANN controllers lies in the fact that both are time-consuming. Many authors have used fractional order (FO) controllers also and have successfully utilized in AGC [8].

A very important task in LFC is to set the controller gains at optimum values so that the objective function is minimized. For this, a number of optimization techniques have been employed by researchers all over the world. Various heuristic and metaheuristic optimization techniques like genetic algorithm [9], particle swarm optimization [10], bacteria foraging [11], artificial bee colony algorithm [12], firefly algorithm [13], bat algorithm [14], etc. have been utilized for obtaining optimum values of controller gains. An algorithm has been used by Singh et al. [15] called Wind Driven Optimization (WDO). The developed technique is very simple and robust. However, the utility of this algorithm is yet to be evaluated in LFC. Hence, further studies using WDO is required to be explored in LFC.

From the extensive survey carried out, the main objectives of the presented work are listed below:

(1) To develop a multi-area thermal system with the implementation of electric vehicle (EV) considering generation rate constraint (GRC).
(2) To apply of Wind Driven Optimization (WDO) technique for simultaneously optimizing controller gains of classical controllers.
(3) To compare the dynamic responses obtained by using I, PI and PID controller in the system developed in (1) for nominal system condition for estimating the best for the system.
(4) Test the behaviour of classical controllers under random load perturbation (RLP).
(5) To evaluate the impact of including EV in the thermal system.

2　System Modelling

An unequal multi-area thermal system with the involvement of electric vehicle (EV) is taken for study. The ratio of the generation capacity of the areas is Area-1:Area-2:Area-3 = 1:3:5.

All the three areas are provided with conventional thermal plants considering generation rate constraint (GRC) of 3% per minute. Single reheat turbine is considered in each of the three areas. The study is carried out under nominal system conditions the parameters of which are taken from [3, 4] and are given in Appendix. The dynamic responses are attained by giving 1% step load as disturbance in Area-1. The dynamics are obtained individually for the secondary controllers such as Integral (I) controller, Proportional-Integral (PI) controller and Proportional-Integral-Derivative (PID) controller. A new nature-inspired optimization technique named as Wind driven Optimization (WDO) technique is utilized for simultaneously optimizing the gains of the secondary controllers. The objective function considered for this study is Integral of squared error (ISE) and is given by Eq. (1). The transfer function model for the system considered is depicted in Fig. 1.

$$J = \int_0^T \left\{ (\Delta f_i)^2 + (\Delta P_{tiej-k})^2 \right\} dt \tag{1}$$

The block diagram for EV is given in Fig. 2. EV is a very good example of battery energy storage system, which is nowadays highly employed by researchers all over the world. The energy stored in the batteries of the EVs will help in providing ancillary support to the power system in the form of frequency regulation by restoring
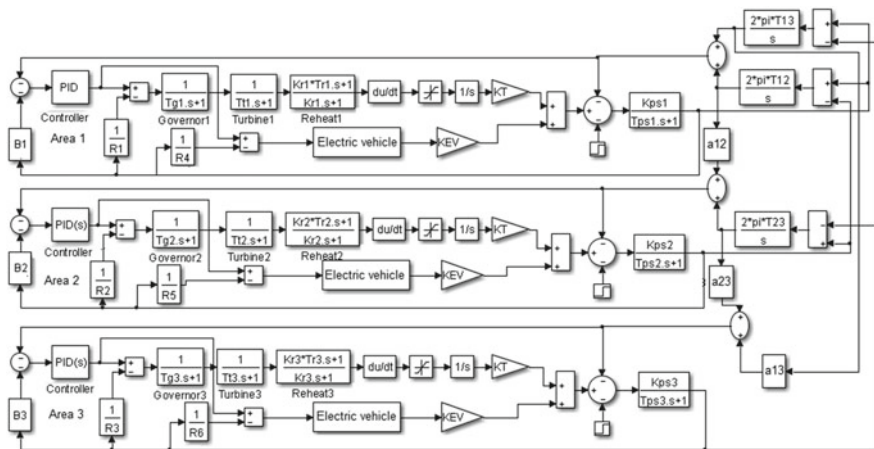
Fig. 1 Transfer function model for thermal system incorporating electric vehicle
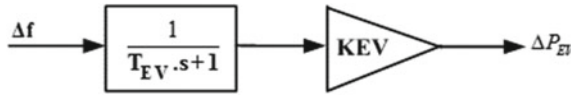
Fig. 2 Transfer function model of electric vehicle

system frequency back to its nominal value on the occasion of load fluctuations in the power system by serving the fluctuations in load demand.

The transfer function of an aggregate EV fleet is given by (2)

$$TF_{EV} = \frac{K_{EV}}{1 + sT_{EV}} \tag{2}$$

where K_{EV} is the gain of EV which is taken to be 1 for the EV fleet to participate in LFC, and $T_{EV} = 1$ is the battery time constant for EV.

3 Wind Driven Optimization Technique

A novel metaheuristic algorithm inspired by nature called Wind Driven Optimization (WDO) was applied by the authors in [15]. The direction of wind is from high- to low-pressure region for balancing the air pressure in our atmosphere with a speed relative to the pressure gradient. Taking assumptions of hydrostatic air balance and vertical movement is weaker than the horizontal movement, the change in pressure as well as the motion of wind can be taken as a horizontal movement. Although our world is three-dimensional, the motion of wind takes into account addresses multidimensional problems. Furthermore, to derive the operators employed in WDO algorithm, specific assumptions and simplifications are required. The algorithm initializes with Newton's second law of motion that offers very specific results when utilized to analyze the motion of wind:

$$\rho a = \sum F_i \tag{3}$$

where, the air density is denoted by ρ, a denotes the acceleration vector, and the forces acting on the mass are denoted by F_i. The relation between air pressure, density and temperature is given by:

$$P = \rho RT \tag{4}$$

where pressure is denoted by P, universal gas constants denoted by R and temperature is denoted by T.

The motion of wind in any specific path is controlled by four main factors, as specified by Eq. (3), out of which the most dominant force that helps in the movement of

air is the pressure gradient force (F_{PG}) which is given by Eq. (8), whereas the friction force (F_F) opposes such movement, and is given by Eq. (5). Gravitational force (F_G) given by Eq. (6) pulls the particles in the direction of the origin of the coordinate system. Another force called Coriolis force (F_C) as given in Eq. (7) happens due to earth's rotation and causes the movement of wind from one to other dimension. The implementation of WDO as the movement in one dimension influences the speed in another.

The equations defining the forces are given below, where δV denotes a minute air volume. Rotation of the earth is represented as Ω, gravitational acceleration is denoted by g, and the velocity vector of the wind is denoted by u.

$$F_F = \rho \alpha u \tag{5}$$

$$F_G = \rho \delta V g \tag{6}$$

$$F_C = -2\Omega * u \tag{7}$$

$$F_{PG} = -\nabla P \delta V \tag{8}$$

All the above equations defining forces may be added up and put in the right-hand side of Newton's second law of motion. The equation formed is given below:

$$\rho \Delta u = (\rho \delta V g) + (-\nabla P \delta V) + (-\rho \alpha u) + (-2\Omega * u) \tag{9}$$

The velocity update equation can be derived from the above equation by considering an infinitesimal air particle that moves with the wind. The pressure in Eq. (3) can be substituted in place of and an assumption of time step to be unity ($\Delta t = 1$) is taken which gives the velocity update equation:

$$u_{new} = (1 - \alpha)u_{cur} - gx_{cur} + \left(\frac{RT}{P_{cur}} |P_{opt} - P_{cur}|(x_{opt} - x_{cur}) \right)$$
$$+ \left(\frac{cu_{cur}^{other\ dim}}{P_{cur}} \right) \tag{10}$$

The updating of the location of air parcels and their velocity is done in each iteration, which is done by using Eq. (10) whereas updating their position is done by employing the equation given below:

$$x_{new} = x_{cur} + (u_{new}\Delta t) \tag{11}$$

Wind Driven Optimization to Generate OGRs Begin /*
WDO parameter initialization */

Channels (marks) 'n' initialization and set upper bound on the ruler length

While not Popsize /* User defined size of population*/
Define Pressure function, coefficients and boundaries;
Generate initial random population and velocity of air parcels;
  Check Golombness of each air parcel;
  If Golombness is satisfied
   Keep that individual;
  Else
   Delete that specific air parcel from the generated population;
 End if
End while /* End of WDO parameter initialization */
Evaluate initial population;
Find best air parcel in the initial population;
While not S /* S is a stopping criterion */
Get an air parcel velocity randomly;
Evaluate its pressure (fitness);
Update velocity and check velocity limits;
Update position and check position limits;
A: Recheck Golombness of new air parcels;
 If Golombness is satisfied Keep that air parcel and then go to B;
 End if
Delete that particular air parcel from the population;
Generate a new air parcel and then go to A;
B: Recompute the pressure (fitness) value of the modified air parcels;
Sort the population from best to worst based on pressure values (fitness);
Replace the Global pressure values with the best;
Keep the best solutions;
Solutions are ranked and best solution currently is found;
End while End

The steps initiate from initialization of parameters and end at ranking the population from best to worst and finally obtaining the optimum value at the end of iterations.

4 Result and Analysis

4.1 Response Comparison for of I, PI and PID Controllers

The multi-area system under investigation is tested with classical controllers like I, PI and PID as secondary controllers individually. The gains of the secondary controllers considered one at a time is optimized by using WDO technique. The optimized values of the gains of controllers are tabulated and compared in Table 1. The dynamic responses obtained using the three classical controllers at the optimum gains are obtained, compared, and shown in Fig. 3. The time of settling and maximum devia-

Table 1 Optimum values of secondary controller gains

Controller	Controller gains	Optimized values of controller gains		
		Area-1	Area-2	Area-3
I	K_{Ii}^*	0.4221580312	0.50452474778	0.43795012961
PI	K_{Pi}^*	0.6390804372	0.03484840724	0.15645388866
	K_{Ii}^*	0.2882525993	0.69650594833	0.59324807083
PID	K_{Pi}^*	0.67164469847	0.02411225635	0.15666835764
	K_{Ii}^*	0.3313021942	0.23598585107	0.30987478884
	K_{Di}^*	0.10146835875	0.077841928705	0.31564659450
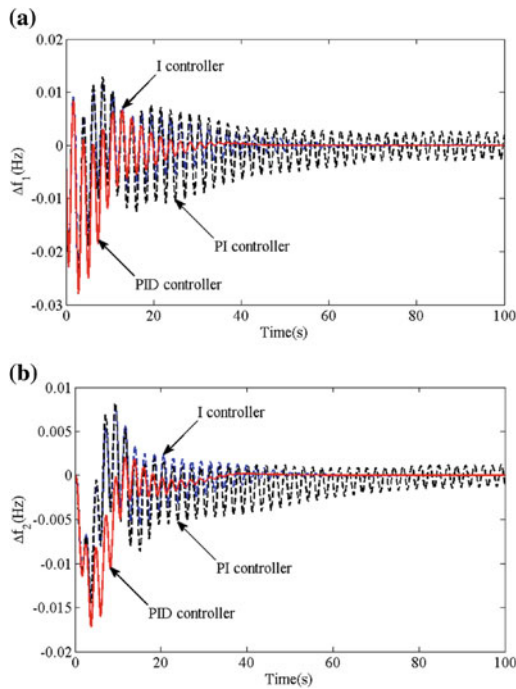
Fig. 3 System dynamics comparison for I, PI and PID controllers under nominal condition **a** Δf_1-t, **b** Δf_2-t

tions are tabulated in Table 2. From observing Fig. 3 and Table 2, it is clearly found out that the responses of PI controller have not died out and has been abbreviated as NS for not settled (NS) till time T. Again, a few of the values of peak deviations cannot be obtained, so, blank space is left in Table 2. From the investigation of the responses, it is found that PID controller outperforms other conventional controllers.
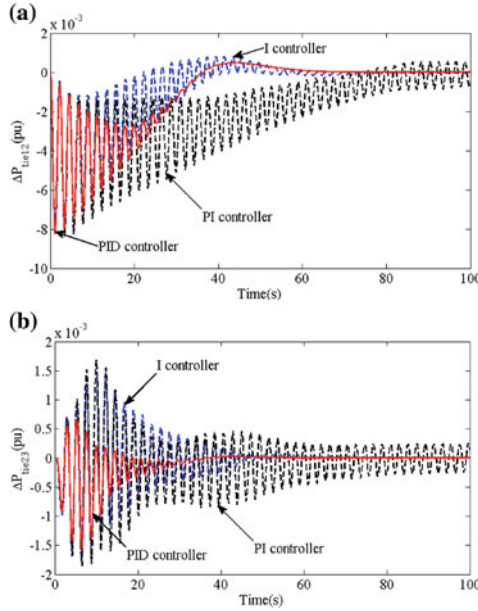
Fig. 4 System dynamics comparison for I, PI and PID controllers under nominal system condition **a** ΔP_{tie12}-t, **b** ΔP_{tie23}-t

Table 2 Values of settling time, peak overshoot and undershoot for Figs. 3 and 4

	Controllers	Figure 3a	Figure 3b	Figure 4a	Figure 4b
Settling time (s)	I	73.73	65.29	75.39	65.46
	PI	NS	NS	NS	NS
	PID	43.07	38.06	62.5	39.86
Peak overshoot	I	0.01285	0.008237	0.0008187	0.001561
	PI	0.01285	0.008247	0.009879	0.001681
	PID	0.00828	0.00204	0.0004804	0.000671629
Peak undershoot	I	0.0711	0.01361	0.007971	–
	PI	0.02726	0.0144	0.008353	0.002899
	PID	0.02773	0.01707	0.008222	0.001581
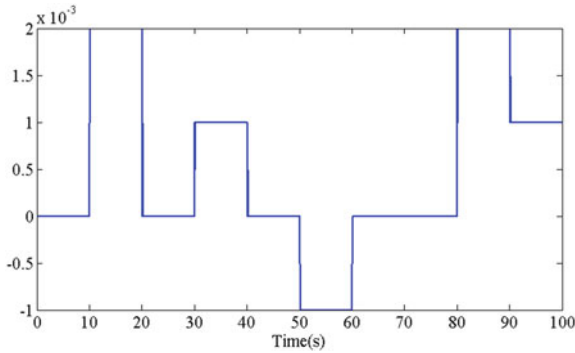
Fig. 5 Random load pattern (RLP)

4.2 Performance Comparison Under Random Loading

In Sect. 4.1, the analysis was carried out under nominal system condition, that is, 1% step loading, 50% nominal loading and inertia constant of 5 s. To study the behaviour and performance of the secondary controllers under random loading condition instead of step loading, a case study is done by replacing step deviation by random loading of pattern shown in Fig. 5 in Area-1. The responses obtained by using the classical controllers are compared and provided in Fig. 6.

From studying the comparison of the dynamic responses it is concluded that the responses of PID controller are better than the other classical controllers both in terms of magnitude of deviations and number of oscillations.

4.3 Comparison of Dynamic Responses in Presence and Absence of Electric Vehicle

From the analysis carried out in Sects. 4.1 and 4.2, it can be determined that the functioning of PID controller is better than that of I and PI controllers under both nominal and random loading conditions. Hence, PID controller is used for further investigation. In this section, an analysis is carried out to evaluate the impact of incorporating electric vehicles (EVs) in the traditional thermal power system. With the optimized values of PID controllers, the dynamic responses obtained in the presence of EVs are compared with those obtained in the absence of EVs and shown in Fig. 7. It can be clearly observed from the figures that the responses in the presence of EVs are far better than those obtained in the absence of EVs in terms of magnitude of oscillations. Hence, it can be concluded that the incorporation of EV in the system helps in overcoming the changes in load demand of the system.
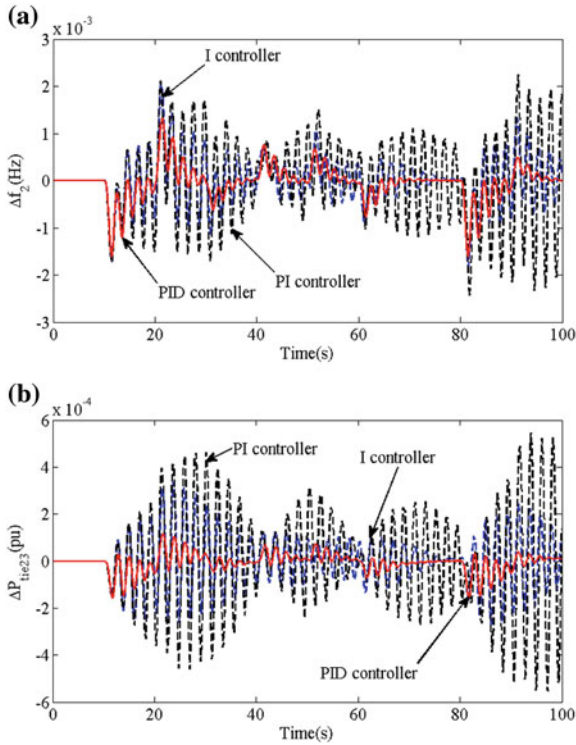
Fig. 6 **a** Comparison of Δf_2 versus t for I, PI and PID controller. **b** Comparison of ΔP_{tie23} versus t for I, PI and PID controller
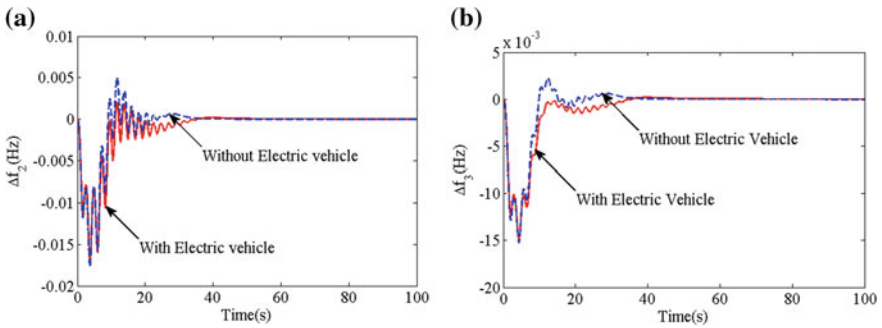
Fig. 7 **a** Comparison of Δf_2 versus t in presence and absence EV, **b** comparison of Δf_3 versus t in presence and absence EV

5 Conclusion

An unequal multi-area thermal system with the integration of electric vehicle (EV) in all the areas is considered for investigation. To give a more realistic view to the system, the test system is provided with generation rate constraint of 3% per minute. The applications of Wind Driven Optimized classical controllers have been attempted in this work. Investigations expose that the responses attained for PID controller are superior to those attained for I and PI controller when the system is under nominal conditions. Also, PID controller behaves better than I and PI controllers when random load pattern is applied in Area-1. EVs can be used for a number of ancillary services, out of which frequency regulation is one. From analysis, it is revealed that the system dynamics improve a lot in the presence of EVs in the system.

Appendix

System parameters are $f = 60$ Hz; $T_{sg} = 0.08$ s; $T_t = 0.3$ s; $T_r = 10$ s; $K_r = 0.5$; $K_p = 120$ Hz/pu MW; $T_p = 0.08$ s; $T_{12} = 0.086$ pu MW/rad H = 5 s; $D = 8.33 \times 10^{-3}$ pu MW/Hz; $B = -\beta = 0.425$ pu MW/Hz; $R = 2.4$ pu Hz/MW; loading = 50%.

References

1. Elgard OI (1982) Electric energy systems theory. McGraw Hill, New York
2. Elgerd OI, Fosha C (1970) Optimum megawatt frequency control of multi-area electric energy systems. IEEE Trans Power App Syst 89(4):556–563
3. Nanda J, Mangla A, Suri S (2006) Some findings on automatic generation control of an interconnected hydrothermal system with conventional controllers. IEEE Trans Energy Convers 21(1):187–193
4. Pham TN, Trinh H, Hien LV (2016) Load frequency control of power systems with electric vehicles and diverse transmission links using distributed functional observers. IEEE Trans Smart Grid 7(1):238–252
5. Barisal AK (2015) Comparative performance analysis of teaching learning based optimization for automatic load frequency control of multi-source power systems. Int J Electr Power Energy Syst 66:67–77
6. Chown GA, Hartman RC (1998) Design and experience with a fuzzy logic controller for automatic generation control (AGC). IEEE Trans Power Syst 13(3):965–970
7. Khuntia SR, Panda S (2012) Simulation study for automatic generation control of a multi-area power system by ANFIS approach. Appl Soft Comput 12(1):333–341
8. Sondhi S, Hote YV (2014) Fractional order PID controller for load frequency control. Energy Convers Manag 85:343–353
9. Ghoshal SP, Goswami SK (2003) Application of GA based optimal integral gains in fuzzy based active power frequency control of non-reheat and reheat thermal generating systems. Electr Power Syst Res 67:79–88
10. Ghosal SP (2004) Optimization of PID gains by particle swarm optimization in fuzzy based automatic generation control. Electr Power Syst Res 72(3):203–212

11. Ali ES, Abd-Elazim SM (2013) BFOA based design of PID controller for two area load frequency control with nonlinearities. Int J Electr Power Energy Syst 51:224–231

12. Gozde H, Taplamacioglu MC, Kocaarslan I (2012) Comparative performance analysis of artificial bee colony algorithm in automatic generation control for interconnected reheat thermal power system. Int J Electr Power Energy Syst 42:167–178

13. Naidu K, Mokhlis H, Bakar AHA (2013) Application of firefly algorithm (FA) based optimization in load frequency control for interconnected reheat thermal power system. In: IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), pp 1, 5, 3–5

14. Naidu K, Mokhlis H, Bakar AHA, Terzija V, Illias HA (2014) Application of firefly algorithm with online wavelet filter in automatic generation control of an interconnected reheat thermal power system. Int J Electr Power Energy Syst 63:401–413

15. Singh PK, Gupta N (2016) A wind driven optimization based WDM channel allocation algorithm. Int J Innov Res Sci Technol 5(7):12073–12080

Lift Automation and Material Sorting Using PLC

Gyanshree and Shabbiruddin

Abstract In this modern world, automation plays an important role in development. Industries are nowadays incomplete without automation. This paper explains about the sorting of materials of different sizes. Based on the requirements, any industry develops materials of different sizes and it will not be economically feasible if the separation is done manually. Labour cost, human effort and chances of human error will increase in this process of sorting. To overcome these disadvantages, automation plays a very important role which makes it economically feasible and time-saving. Chances of human error will also minimize. This paper is basically about the sorting of medium-sized objects to one level and large-sized objects to another level through vertical upward movement of lift by using Programmable Logic Controller (PLC) and Supervisory control and data acquisition (SCADA).

Keywords Automation · Programmable logic controller · Micrologix 500 english · RS view 32 works · Sorting · SCADA

1 Introduction

Improvement in manufacture sector depends upon various explorations in manufacturing process and innovation of new technology in new products. Those countries who are having high manufacturing and production rates are called to be developed whereas those with little manufacturing and production rates are considered to be underdeveloped [1].

Gyanshree · Shabbiruddin (✉)
Department of Electrical & Electronics Engineering,
Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Majitar, Sikkim, India
e-mail: shabbiruddin85@yahoo.com

Gyanshree
e-mail: gyanshree49@gmail.com

In many industrial applications, separation of objects is needed. Sorting can be done in many ways like sorting of object according to their dimensions (height, length, thickness, etc.), according to their colours, according to their weight, using machine vision (image processing), according to the material of an object, etc. For example, in thermal power station, electromagnetic sorting procedure is used to sort ferromagnetic materials from coal.

For growth of any industry, automation is an important aspect. Quality and flexibility of the product are the main criteria. Robotics are basically used to perform automated tasks like sorting, painting, transferring. Earlier, robots does not require high accuracy and precision [2].

An advantage of using PLC in controlling a mechanically designed system is much easier as compared to relay control. PLC allows easy accessing of variety of sensors, inputs, and outputs. Henceforth, PLC becomes the required choice for industrial purposes. The first and necessary step in any industrial application is identifying number of inputs, outputs, appropriate sensors and actuators which has to be designed and executed [3].

Earlier, when demand in the market was less, it was possible to use manual labour for sorting similar object, but now with increase in demand and production, it is not suitable for industries to have any human errors in sorting these products. These phenomena forced the industries to adopt automation in sorting process. Proper and accurate storing of finished materials is an important aspect for the industry. In developing any industry economy, it is very much necessary to make cost-effective automation system which is affordable. Also, system should be able to elevate productivity and discard defective objects accurately and precisely. The stacking should be done properly without any breakdown [4].

To reduce human errors and their contributions, automation plays a very important role in control systems. Any system which is automated uses complex architecture which ultimately increases the overall cost of design layout and the power involved. With evolving these technologies, human efforts reduces which ultimately provides more time to work on different methods and aspects. With the implementation of automation, life risks to human reduce when they are made to work in hazardous conditions [5].

Control system in the field of automation is the combination of various components such as boilers, heat ovens and many more. It is being used in switching telephone networks and for stability of ships and aircrafts in which human interference is reduced to minimum value. There are some processes in the field of automation which are completely automatic and self-improvising [6].

Belonging to any sector in the field of automation, it has an important role and contribution towards efficiency, effectiveness and reducing the human error to minimum extent. PLC is an industrial computer whose state of action depends on the ladder program, and hence continuously monitors the input states. It comprises of power supply, CPU and I/O devices [7].

2 Problem Definition

In many industrial applications, it has been observed that once the machine is switched on, conveyor belt continuously moves. During the upward movement of lift, it is not acceptable to place any object over the belt, but still it continues to move. It is very obvious that it will lead to wastage of electrical energy, and hence money. When object reaches to any particular level then pushing mechanism is done manually. In general, industries have automatic pushing mechanism, but it might happen that the paths through which objects have to move further (after getting out of the lift) is not vacant and this will struck the objects in a queue. Also it might happen that the robot which is used to pick and place the object from the position, just after material sorting to another position, is busy at that particular moment, and henceforth it will hamper the entire process. So, there is a need to develop technique for tackling such kind of problems.

3 Solution Strategy

To tackle the issues, ladder diagram is designed in such a way that when the lift starts to move upward, conveyor belt stops and it will save electrical energy to some extent. When lift comes vertically downward at the base position then only belt starts moving again. When the object reaches any particular level, manual pushing mechanism is used with the help of Micro limit switch.

This is all about representation of lift automation and sorting of different size objects using PLC and SCADA. It is very much necessary to separate the objects as per their sizes for convenience. As soon as the Start button is pressed, conveyor belt starts moving. Two photoelectric sensors are placed in the path of conveyor belt which is used to sense medium-sized and large-sized objects. Photoelectric sensors are placed at different heights which can easily sense the objects of different sizes. Object is placed over the conveyor and depending upon the size it is sensed by the sensor. In the path of conveyor, a photosensor to detect large objects is placed first and then to detect medium objects. When medium objects pass through belt, sensors used to detect large objects will not sense it. When large objects pass through belt, sensor used to detect these objects will get activated. As the belt is moving, after a certain duration of time, material gets inside the lift. As soon as the material gets inside the lift, it is again sensed by photoelectric sensor, which will decide the level up to which lift has to move upward and the vertical upward movement of lift starts. Medium-sized and large-sized objects get separated on Level 1 and 2, respectively. If the object present inside the lift is of medium size then lift automatically gets stopped at Level 1. Inductive proximity sensor will stop the upward movement of lift at any particular level. When lift gets stopped at level 1, then with the help of Micro limit switch, manual pushing mechanism starts and it will take the object out of the lift. After the material gets out of the lift, it starts moving downward. It is very

important to note that during upward movement of lift, conveyor belt automatically gets stopped. Same scenario takes place for the large size objects, but it will take the object out of the lift at Level 2.

4   Results Obtained

As soon as the Start button is pressed, conveyor belt gets started. Since it is a bistable switch, latch is connected in parallel to Start button, with the address of belt. Bistable means switch gets back to its same state after getting pressed at once. The purpose is to make Start button turn ON even after switch gets back to its original state. When medium-sized object starts moving over the belt, sensor used for that purpose gets active and for a period of 5 s, object will move forward. When one object is counted in the path then up counter becomes high, which can be observed from Fig. 1.

Upward movement of lift (O:0.0/1) starts as soon as object inside lift is sensed by sensor present inside the lift. Upward movement of lift is done using a timer for a duration of 10 s (T4:0). At Level 1, Proximity sensor (I:0.0/4) and Micro limit switch(I:0.0/8) becomes high and hence, Pushing mechanism starts for a duration of 5 s, as shown in Figs. 2 and 3.

After pushing mechanism is over, down counter (C5:0) is connected across it which will compare that one object is out of the lift, and hence downward movement of lift starts for the same duration of 10 s, which can be executed by connecting a timer across it (T4:4), as shown in Fig. 4.
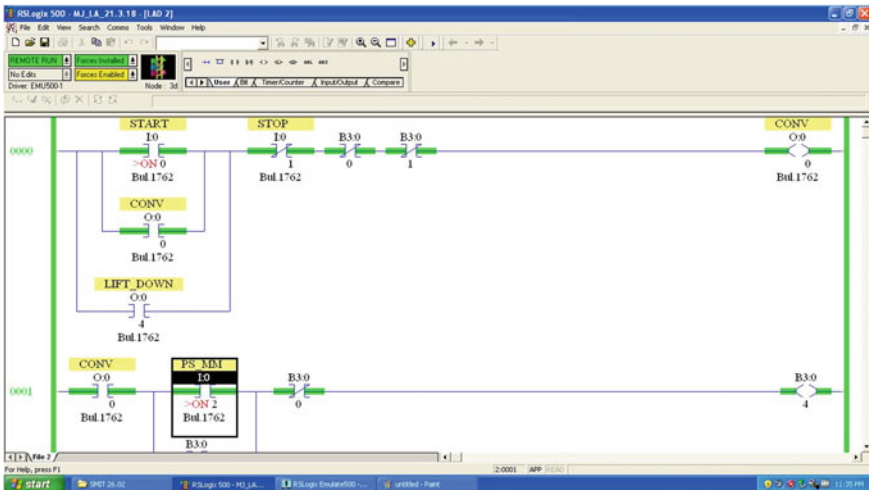
Fig. 1 Ladder diagram representing conveyor belt part

Fig. 2 Upward movement of lift

Fig. 3 Representation of pushing mechanism

The entire process can be done in same way for large size objects. Level 2 is assigned for separation of large size objects. Sensor is present inside lift to sense large-sized object and for the vertical upward movement of lift. Just like medium object case, proximity sensor and Micro limit switch are connected at level 2 for pushing mechanism, and to take the object out of the lift, which can be seen from Figs. 5 and 6.

Fig. 4 Downward movement of lift

Fig. 5 Sensing of large-sized objects

This process can also be shown in working mode with the help of SCADA via simulation as shown in Fig. 7.
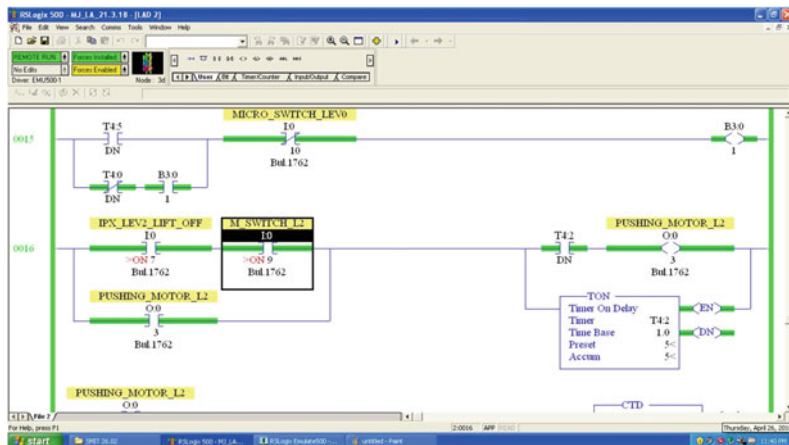
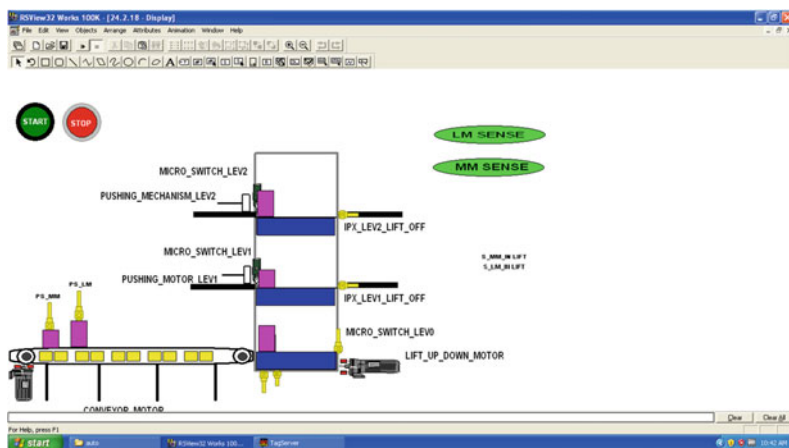**Fig. 6** Pushing mechanism at another level

Fig. 7 Simulation on SCADA

Experimental data are mentioned below in table:

Input high state	Output high state
Start (I:0.0/0)	Conveyor (O:0.0/0)
Memory bit (B3:0/0 and B3:0/5)	Lift up (O:0.0/1)
Inductive sensor at Level 1 (I:0.0/4) Micro switch at Level 1 (I:0.0/8)	Pushing motor at Level 1(O:0.0/2)
Memory bit (B3:0/3)	Lift down (O:0.0/4)
Photo sensor for large objects(I:0.0/5)	Memory bit (B3:0/6)
Large material sensor inside lift (I:0.0/6)	Counter (C5:1)
Inductive sensor at Level 2 (I:0.0/7) Micro switch at Level 2 (I:0.0/9)	Pushing motor at Level 2 (O:0.0/3)

5 Conclusion

Lift automation and object sorting is done through Programmable Logic Controller and working mode is executed through SCADA, where it can be observed that the entire process is in animation mode. Software used for simulation in PLC is RS Logix 500 English and for SCADA is RS View 32 Works. These softwares are compatible with 32-bit operating systems. Ladder diagram is made using Rockwell Automation. The problems that are being faced are resolved. Ladder diagram to simulate only in PLC need some changes to make it compatible with SCADA. Photoelectric sensors are used to sense any object in their paths and inductive proximity sensors are used to detect the metallic objects. Micro limit switches work in NO and NC modes. Motors are required for lift movement and belt rotation. Inductive proximity sensors are used to stop the lift movement at any particular level, depending upon the size of object. Proximity sensors should be of large range so that it can access long area. In any practical situation, we can use operate motors using 24 V DC relay. Same motor can be used for forward and reverse rotation by the use of two relays. Ladder diagram is made in such a way that when object of any particular size is sensed then it lift will get automatically stop at the respective level. If any other sensor gets active at another level, it will not stop lift at the undesired level. The model developed will be helpful for design engineers in upgrading automation in industries.

References

1. Bargal N (2016) PLC based object sorting automation. Int Res J Eng Technol (IRJET) 03(07)
2. Aruna YV, Beena S (2015) Automatic convey or system with in–process sorting mechanism using PLC and HMI system. Int J Eng Res Appl 5(11):37–42 (Part-3). ISSN: 2248-9622

3. Mangalore ZB (2015) PLC based elevator system with colour sensing capabilities in industrial applications. NCRIET-2015 Indian J Sci Res 12(1):186–191. ISSN: 2250-0138 (Online)
4. Tailor D Object sorting and stacking automation with PLC. Int J Eng Technol (IJET). ISSN (Print): 2319-8613, ISSN (Online): 0975-4024
5. Rautu SV, Shinde AP, Darda NR, Vaghule AV, Meshram CB, Sarawade SS (2017) Sorting of objects based on colour, weight and type on a conveyor line using PLC. IOSR J Mech Civil Eng (IOSR-JMCE) 04–07. e-ISSN: 2278-1684, p-ISSN: 2320-334X
6. Nagapure MB, Deshmukh RM, Nagapure MB (2017) PLC based segregation of scrap material. Int J Eng Res Appl 7(4):06–10 (Part-3). ISSN: 2248-9622
7. Aruna YV, Beena S (2015) Automatic convey or system with in–process sorting mechanism using PLC and HMI system. Int J Eng Res Appl 5(11):37–42 (Part-3). ISSN: 2248-9622

Automated Computer-Aided Lung Cancer Detection System

Ananya Bhattacharjee and Swanirbhar Majumder

Abstract Cancer is one of the most deadly diseases in the world. Cancer of the lung or lung cancer is one of the most commonly occurring variants of cancer. This is due to an increase in pollution, presence of harmful gases in the environment and work area, smoking etc. Lung cancer can be diagnosed in a better way only if pulmonary lung nodules are detected at an early stage of growth. Popularly used techniques used for assisting the computer-aided systems in various stages of image processing include image acquisition, pre-processing, candidate nodule detection, segmentation and post-processing. Several pre-processing techniques are employed like image enhancement, smoothing, edge detection and region of interest selection. The main goal of this paper is to establish such an efficient automated computer-aided detection system which is able to detect malicious nodules in the same manner as it is detected manually by a doctor. Soft computing will be employed for the decision-making process.

Keywords Lung cancer · Computer-aided detection · Nodule detection · LIDC-IDRI · CT images

1 Introduction

Lungs are the most important organs of our respiratory system. Lung cancer is the most dangerous and brutal disease. It is a disease whereof the many cells present in the lungs some abnormal cells rapidly multiply and grow into a tumour. Most of the lung cancer cases are diagnosed late because most of the people are not aware of its early symptoms. As they are camouflaged under common symptoms like

A. Bhattacharjee
Department of ECE, NERIST, Nirjuli, Auranachal Pradesh, India
e-mail: ananya0409@gmail.com

S. Majumder (✉)
Department of IT, Tripura University, Suryamaninagar, Tripura, India
e-mail: swanirbhar@ieee.org

Coughing/Shortness of breath, etc. Lung cancer is frequently diagnosed in advanced stage where probability of survival is too low. As a result, effective treatments are not possible on time. The major causes of lung cancer are related to the usage of tobacco and cigarette and among them, 80% of the total number of cases is due to tobacco usage. Cancer of the lung represents both the highest mortality rate as well as the smallest survival rates after diagnosis.

1.1 CAD System

CAD system helps doctors to correctly interpret medical images more accurately. This system is of two types: computer-aided detection (CADe) system and computer-aided diagnosis (CADX) system. The former is responsible for detecting lesions with the help of medical images whereas the later is responsible for measuring various lesion characteristics like determination of malignancy and stages of cancer. The main goals of CADe system are diagnosis accuracy, early detection of cancer and minimal time in evaluation by the radiologists.

Figure 1 represents various stages of Computer-aided diagnosis (CAD) system.

Pre-processing is the first stage of CAD system which consists of several methods like image enhancement, smoothening, edge detection and so on. The main purpose of nodule detection stage is to properly identify the presence of suspicious nodules in the analyzed images because patient's survival rate may be increased if these lung nodules are detected in an early stage but this is quite a tedious task. The segmentation substage is applied to separate the target region from other organs. Post-processing stage includes feature selection and classification.

1.2 Soft Computing

It is a branch of artificial computational intelligence which is able to analyze varied complex medical data by employing different kinds of optimization techniques in order to improve the diagnosis and detection of cancerous nodules. Its main methodologies are Fuzzy logic (FL), evolutionary algorithms such as Genetic algorithm (GA) and artificial neural networks (ANNs).

Fig. 1 CADe system block diagram
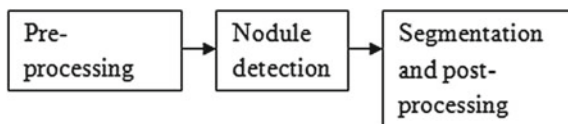
2   Review of Previous Work on Lung Nodules

Over more than a decade, various efforts have been done to develop an automated system which can detect suspicious lesions in thoracic CT and other types of imagery. In 1998, CT screening devices as well as a filtering technique known as "N-Quoit filter" are used [1]. In 2001, 2D as well as 3D feature analysis technique and linear discriminant-based classifier are used to differentiate actual nodules from false positive (FP) candidates [2, 3]. In 2002, directional gradient concentration features were developed to reduce FP generated [4]. In 2005, a surface normal overlap technique and lantern transform were developed in order to form a feature vector where rule-based classifier technique was processed to classify nodules and non-nodules [5]. In 2007, a dot enhancement filter was implemented for the selection of nodule candidate and a neural classifier was also used in order to reduce FP [6]. In 2009, temporal subtraction image was developed by implementing a technique based on artificial neural networks for detecting lung nodules [7]. In 2010, intensity thresholding as well as morphological processing were used for the detection of nodule candidates [8]. In 2011, Hopfield Neural Network (HNN) and Fuzzy C-Mean (FCM) clustering algorithm were used for segmentation in order to detect lung cancer. This will improve the survival rate of the patient. HNN provides better classification result than FCM [9]. In 2013, mean-shift methodology and geometric properties-based techniques like region of interest (ROI) were implemented [10]. In 2014, extraction and enhancement of pulmonary parenchyma were carried out and then nodule candidates are segmented. Micro-genetic algorithm was also employed in order to find out the best training model and SVM for final classification [11]. In 2015, much rarer larger nodules greater than 10 mm were mainly focused [12]. The main focus of this system was to include those large nodules which are attached to the pleural wall through the process of morphological processing. In 2016, the system presented cloud-based database was presented for detection of pulmonary nodules which is mainly characterized by 3D texture attributes [13]. Not only Structured Query Language (NoSQL) approach was developed which mainly consists of 838 nodules, 379 exams, 8237 images comprising of 4029 CT scans and 4208 manually segmented nodules. In 2017, a new segmentation algorithm named as PropSeg was proposed [14] which comprises of pre-processing, candidate detection by three levels of fuzzy C-means (FCM) clustering, segmentation and post-processing performed by morphological edge detection method. This system performed better than other techniques and so, it is suitable for lung-related disease detection. In [15], pattern recognition techniques were used. To classify malignant and benign nodules, phylogenetic diversity technique was used. To select the best model, genetic algorithm was used. It achieved accuracy of 95.52%, sensitivity of 93.1% and specificity of 92.26%. In [16], on the basis of detection accuracy, various lung cancer detection techniques were listed which were analyzed step by step and then overall limitations were pointed out.

Table 1 Database
specifications

Sl. no.	Specifications	TAG description
1.	Manufacturer	GE medical systems
2.	Patient ID	LIDC-IDRI-00061
3.	Body part examined	Chest
4.	Scan options	Helical mode
5.	Bits allocated	16
6.	Bits stored	16
7.	Image type	Original/primary/Axial
8.	Manufacturer model name	Lightspeed 16
9.	Filter type	Body filter
10.	Modality	CT
11.	Slice thickness	2,500,000
12.	Exposure time	1160

3 Database Used

LIDC-IDRI is the database used which is an association of both the Lung Image
Database Consortium (LIDC) and the Image Database Resource Initiative (IDRI). All
the CT images are in the form of Digital Imaging and Communications in Medicine
(DICOM). Each image is having a dimension of 512×512. It consists information
about the nodule markings and size of the nodules is in between 3 and 30 mm. It
contains 888 thoracic CT scans and section thickness is about lesser than or equal to
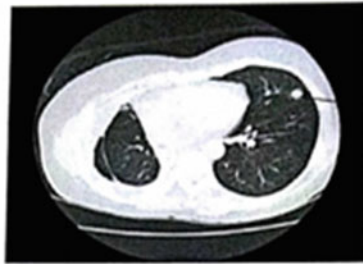2.5 mm.

Its specifications are listed in Table 1.

4 Methodology

Working methodology is based on two approaches: manual detection and automated
computer-aided detection. Manual detection is based on the marking of suspicious
nodule which is certified by a doctor. This is shown in Fig. 2. The main aim of
automated detection is to train CADe system in such a manner that malicious nodules
detected by both doctor and automated system resemble each other.

Block diagram of the working procedure is shown in the above Fig. 3. After
pre-processing, segmentation and feature extraction will be implemented. Segmen-
tation will be implemented by applying multilevel thresholding and then several
features like area, convex area, perimeter, solidity, etc., of the suspicious nod-
ules will be extracted. Among various soft computing techniques, artificial neural

Fig. 2 Certification of suspicious nodules by a doctor

network (ANN) will be implemented for decision-making process in future [9]. ANN is useful for the development of algorithms for complex pattern recognition.

5   Result

Pre-processing is the initial stage of CAD system which includes image enhancement, smoothing, edge detection and Region of Interest (ROI) selection. Image enhancement is based on two approaches:

1. Contrast Enhancement with Morphological opening.
2. Histogram equalized Enhanced image.
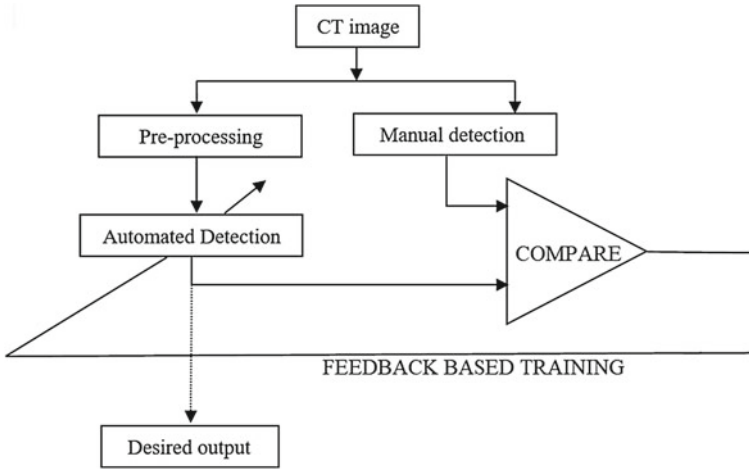
Fig. 3 Soft computing-based training of automated detection via manual detection data

The above two techniques are compared. Histogram equalization technique shows better enhancement result. Both of these techniques are clearly represented in Fig. 4a, b, respectively. Smoothing is obtained with the help of median filter. It is a non-linear filter that reduces noise in an image. Figure 4c represents smoothened image. Edge detection is performed on four operators: Sobel, Prewitt, Roberts and Canny. Among these, Canny edge detector provides the best edge details. Although Prewitt is simpler than Sobel, it produces noisier results which are not desirable. Roberts faces the problem of symmetry and detection of those edges which are multiples of 45°. Figure 5 represents all edge detection techniques.

ROI extraction algorithm is based on morphological filters which comprise operators of four types: erosion, dilation, opening and closing. All input CT image is fed into the morphological reconstruction block comprising of a marker and a mask as shown in Fig. 6a. This technique resulted in removal of those textures which are devoid of strong edges. So, these edges are extracted by applying canny edge detec-
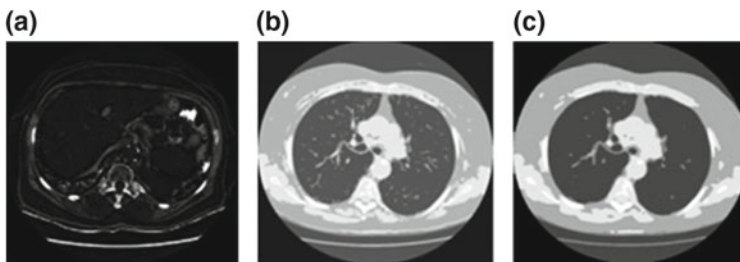
Fig. 4 Pre-processing stage **a** contrast-enhanced with morphological opening **b** histogram equalized image **c** median filter smoothened image
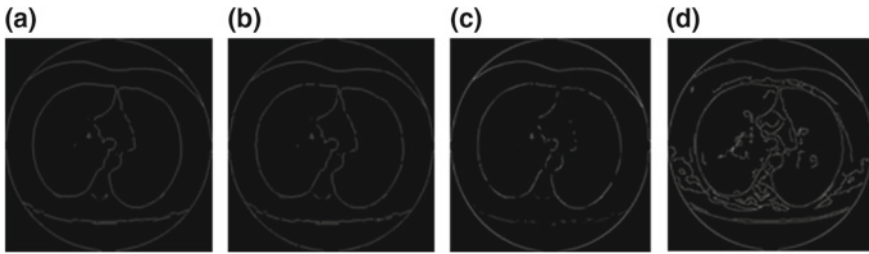
Fig. 5 Edge detection **a** Sobel **b** Prewitt **c** Roberts **d** Canny
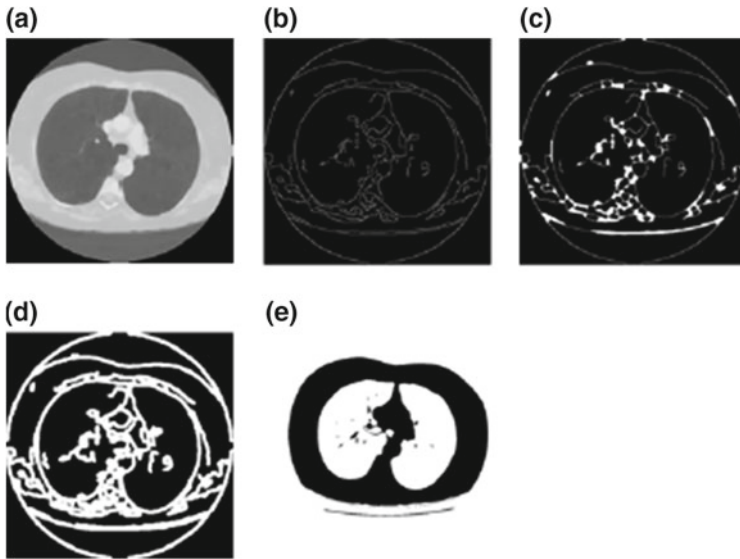
Fig. 6 ROI selection **a** reconstructed image **b** canny edge detected **c** morphologically closing operation on edge detected image **d** morphologically filtered image **e** ROI-extracted image obtained with threshold of 830 HU

tion. This is shown in Fig. 6b. However, some gaps in the edges are observed which is filled by morphological closing operation as shown in Fig. 6c Also, morphological filtering is applied to smooth the resultant image. This is shown in Fig. 6d. Finally, extraction of ROI is obtained by applying a threshold of 830 HU as shown in Fig. 6e.

6 Conclusion

Lung cancer presents the disease having one of the smallest survival rates after diagnosis. So, early diagnosis is much needed in this field to improve the condition of the patients suffering from this disease. Therefore, detection in initial stage has a

probability of being cured whereas at advanced stages it is life-threatening. The present work proposes a methodology for automatic detection of lung nodules using a combination of median filtering based smoothening, image enhancement, edge detection, histogram equalization in the pre-processing stage and then applying morphological operations to extract ROI. The authors are planning to perform segmentation and post-processing in the later stage with the help of artificial neural network.

References

1. Okumura T, Miwa T, Kako J, Yamamoto S, Matsumoto M, Tateno Y, Iinuma T, Matshmoto T (1998) Variable N-Quoit filter applied for automatic detection of lung cancer by X-ray CT. Comput Assist Radiol Surg (CAR98) 242–247
2. Armato III S, Giger M, Moran C, Blackburn J, Doi K, MacMahon H (1999) Computerized detection of pulmonary nodules on CT scans. RadioGraphics 19(5):1303–1311
3. Armato III S, Giger M, MacMahon H (2001) Automated detection of lung nodules in CT scans: preliminary results. Med Phys 28:1552–1561
4. Nappi J, Yoshida H (2002) Automated detection of polyps with CT colonography: evaluation of volumetric features for reduction of false-positive findings. Acad Radiol 386–397
5. Rubin G, Lyo J, Paik D, Sherbondy A, Chow L, Leung A, Mindelzun R, Schraedley-Desmond P, Zinck S, Naidich D (2005) Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. Radiology
6. Gori I, Fantacci M, Preite Martinez A, Retico A (2007) An automated system for lung nodule detection in low-dose computed tomography. In: Giger ML, Karssemeijer N (eds) Proceedings of the SPIE on medical imaging. computer-aided diagnosis, vol 6514, p 65143R
7. Miyake N, Kim H, Itai Y, Tan JK, Ishikawa S, Katsuragawa S (2009) Automatic detection of lung nodules in temporal subtraction image by use of shape and density features. In: Liao P, Bin-Yih, S (eds) Fourth international conference on innovative computing, information and control (ICICIC). Institute of Electrical and Electronics Engineers, pp 1288–1292
8. Messay T, Hardie RC, Rogers SK (2010) A new computationally efficient CAD system for pulmonary nodule detection in CT imaging. Med Image Anal 14:390–406
9. Taher F, Sammouda R (2011) Lung cancer detection by using artificial neural network and fuzzy clustering methods. In: IEEE GCC conference and exhibition (GCC), pp 295–298
10. Chama CK, Mukhopadhyay S, Biswas PK, Dhara AK, Madaiah MK, Khandelwal N (2013) Automated lung field segmentation in CT images using mean shift clustering and geometrical features
11. Antonio Oseas de Carvalho Filho, Wener Borges de sampaio, Aristofanes Correa Silva, Anselmo Cardoso de paiva, Rodolfo Acatavassu Nunes and Marcado galtass (2014) Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. Artif Intell Med 60:165–177
12. Setio AAA, Jacobs C, Gelderblom J (2015) Automatic detection of large pulmonary solid nodules in thoracic CT images. Med Phys 42(10)
13. Junior JRF, Oliveira MC, de Azevedo-Marques PM (2016) Cloud-based NoSQL open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. J Digit Imaging. Springer
14. Abbas Q (2017) Segmentation of differential structures on computed tomography images for diagnosis lung-related diseases. Biomed Signal Process Control 33:325–334

15. Antonio Oseas de Carvalho Filho, Aristófanes Corrêa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassú Nunes, Marcelo Gattass (2017) Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and SVM. J Digit Imaging Soc Imaging Inf Med
16. Makaju S, Prasad PWS, Alsadoon A, Singh AK, Elchouemi A (2018) Lung cancer detection using CT scan images. In: 6th international conference on smart computing and communications, pp 107–114

Indirect Field Oriented Control for Three-Phase Induction Motor Drive Using DSP Controller

Tista Banerjee and Jitendra Nath Bera

Abstract For efficient speed control of three phase induction motor mostly vector control (VC) or field-oriented control (FOC) technique is applied to the Voltage Source Inverter based-Induction Motor Drives (VSI-IM). In implementing FOC, the angular position of the rotor flux vector (θ) is to be evaluated. The accurate estimation of this vector θ along with the feedback parameters is the most critical aspects as the entire FOC logic is based on it. In this paper, an attempt has been made to implement the Indirect Field Oriented Control (IFOC) scheme by estimating θ from slip frequency and rotor frequency through a DSPIC30f4011 microcontroller using its capture module. The hysteresis modulation technique is used for generating the gate pulses of power circuit of VSI. The power MOSFETs are triggered using high-speed gate driver with bootstrap and motor protection circuits embedded in it. The IFOC scheme is then verified with 3KW IM using a prototype VSI model. It has been seen that this VSI-IM with IFOC scheme gave a satisfactory result with varied torque and speed.

Keywords IFOC · VSI · IM · DSPIC · Hysteresis controller

1 Introduction

In industries mostly Voltage Source Inverter (VSI) based variable speed induction motor drives are used with various control schemes. They are provided with high-speed controllers for implementation of the control strategies. The speed of these drives can be controlled satisfactorily by scalar control methods which give very good result in the steady state performance of the drive. But this control scheme produces oscillation to the torque developed. Again presence of inherent coupling

T. Banerjee (✉) · J. N. Bera
Department of Applied Physics, University of Calcutta, Kolkata, India
e-mail: tista.banerjee@gmail.com

J. N. Bera
e-mail: jitendrabera@gmail.com

effect between the torque and flux of an induction motor makes the control strategy nonlinear and motor's response becomes slow. To improve this, various vector control strategies, like field-oriented control (FOC), feedback linearised control, direct torque control, inverse gain based linearised method, etc. are used [1, 2].

The vector control or field-oriented control (FOC) logic decouples this inherent coupling effect of the IM for which the torque and the flux and the machine speed can be independently controlled like a separately excited dc motor. FOC transforms a 3-ϕ sinusoidal varying motor parameter to a synchronously rotating reference frame where sinusoidal parameters appear as dc quantity in steady state. In an IM the coefficient of coupling parameters changes with the rotor flux vector position θ and the entire mathematical calculation of FOC is based on it [3, 4]. Depending upon the estimation of this angle, FOC is classified as Direct FOC, which uses flux estimator to calculate θ and Indirect FOC which calculates θ from slip frequency and rotor speed. As slip frequency can be calculated from motor parameters only and no extra device is required for this purpose, implementation of IFOC is much simpler and cost-effective [1, 2].

To achieve this, high precision control strategy, the control unit of the VSI-IM drives contain very high-speed DSP-based microcontrollers. These controllers are programmed via code composer studio or with dedicated tool boxes with DSP or from motor development kits with DSP or from real-time simulators directly. Thus, costly sensing units are also required to maintain accuracy. This increases the overall cost of the system [5–7].

In VSI-IM Drives, presence of lower order harmonics gives rise to various problems like torque pulsation, heating, etc. To overcome these problems, various modulation schemes are applied like PWM, SPWM, THSPWM, Space Vector Modulation, voltage carrier in phase lock manner [7]. But for instantaneous response, Hysteresis current controller modulation scheme provides the best result. It is simple, with fast transient response and insensitive towards dc link voltage ripple [8, 9].

In this paper, IFOC scheme is applied to 3KW VSI-IM drive prototype using DSPIC30f4011 with hysteresis modulation scheme. The rotor flux angle θ is calculated using slip frequency and rotor speed. The slip gain is evaluated from motor constant parameters to avoid detuning of IFOC. The hall sensors sense the feedback current and proximity sensor senses the speed using. Along with the DSP-based microcontroller a high-speed gate driver IRS2336 $\times$ 64 is used. This IC has bootstrap circuit and can provide protection to the motor from over current, under voltage, etc. With these modifications imparted to the system, the PCB size is reduced, cost is reduced and overall performance of the drive is improved. Before designing the hardware, the system performance is simulated in an artificial environment. It has been seen that θ varies in a sinusoidal manner. This system gave satisfactory result both at transient and steady state, even when the speed is varied rapidly.

2 Materials and Methods

2.1 Field-Oriented Control or (FOC)

As mentioned earlier that Field-Oriented Control (FOC) improves the transient response of IM and it can be controlled like DC machine, where its torque component and field flux component are separated virtually and independent control of each component is possible. FOC is based on phase transformation, where a three-phase time and space variant variable parameter is transformed to a synchronously rotating, time invariant system, leading to a structure similar to that of a dc machine [1, 2]. If the stator current in 3-ϕ system is represented by i_q and i_d in the synchronously rotating reference frame, the torque (T_e) of the IM can be computed as shown in Eq. 1

$$\begin{cases} T_e = K' \times \psi_r \times i_q \\ T_e = K' \times i_d \times i_q \end{cases} \tag{1}$$

where ψ_r is magnetising flux vector. In FOC control, here i_d is analogous to the field current, and i_q is analogous to armature current of the dc machine. This means that when i_q is controlled it affects the torque directly, ψ_r remaining unaffected. Similarly, when i_d is controlled it affects the flux only and does not affect the i_q component of current. Thus, an induction motor can be treated as a dc machine.

2.2 Hysteresis Modulation

In this modulation, a triangular wave is allowed to oscillate within a defined window Δv above and below the reference wave. From the vertices of the triangular wave, the output pulses are generated. Keeping the slope of the triangular wave constant, if the frequency of modulating wave is changed, the number of pulses and pulse widths of modulated wave would change. This modulation can control the ratio of voltage to frequency. This type of modulation is basically used with instantaneous feedback current control method like Indirect Field oriented control schemes [8].

3 The Schematic of the Proposed Work

The basic plan of work is represented by the schematic diagram as shown in Fig. 1. This VSI-IM drive contains three parts, power unit, control unit, and sensing unit. The power unit is composed of a 1-ϕ VARIAC, ac to dc rectifier, VSI with power MOSFET as switches connected in H-bride configuration, and 3KW 3-ϕ IM subjected to load. The control unit contains high-speed DSP-based microcontroller provided with
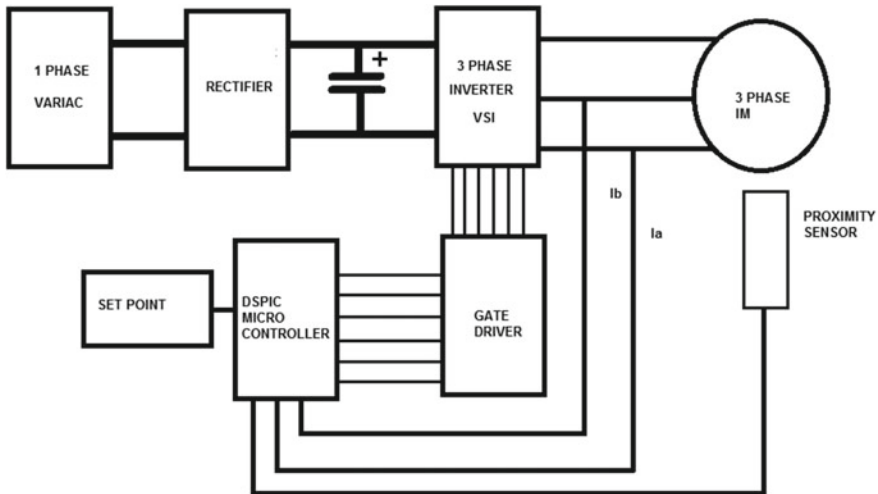
Fig. 1 Block diagram of IFOC-based VSI-IM

dedicated motor control PWM module. This module has three independent comple-
mentary PWM channels suitable for 3-φ operation. The processing speed of the IC
can be increased up to 30 MIPS by proper selection of oscillator and PLL. Thus, it
is capable enough to perform efficiently for IFOC control scheme. The sensing unit
is composed of two hall sensors to sense instantaneous current and proximity sensor
to sense speed. The power circuit is isolated from control circuit by high-speed gate
driver. This gate driver IC contains bootstrap circuit for which the isolated control
supply for the power circuit is reduced from four to one. It is also featured with
various motor protections like over current, under voltage etc. with adjustable fault
clearance. The system is operated at frequency 50 Hz and 400 V. The PWM speed
is maintained at 5 kHz thus the sampling period is 2 μs.

Before hardware implementation, the drive performance has been simulated in an
artificial environment.

3.1 Expressions for Estimation of Parameters for IFOC Implementation

For implementation of IFOC scheme some mathematical formulas are required. FOC
logic is based on projection of motor variable parameters to get the exact value of θ.
So the 3-φ instantaneous stator currents i_a, i_b and i_c, in 3-φ time-varying reference
frame (a-b-c) is transformed to a stationary orthogonal 2-φ system (α-β-0). These
processes is called Clarke's transformation. The current $i_α$ and $i_β$ in (α-β-0) frame is
given by Eq. 2

$$i_\alpha = i_a, \; i_\beta = \left(i_a + 2 \times i_b/\sqrt{3}\right) \tag{2}$$

Then from (α-β-0) reference frame, the currents are projected to a synchronously rotating orthogonal reference frame (d-q). Let the current in (d-q) axis be i_d and i_q This transformation is known as Park's transformation as shown in Eq. 3.

$$i_d = i_\alpha \cos \theta_e + i_\beta \sin \theta_e, \; i_q = -i_\alpha \cos \theta_e + i_\beta \sin \theta_e \tag{3}$$

In IFOC slip frequency ω_{sl} and rotor speed ω_r are integrated to calculate the position of synchronously rotating reference frame θ_s from synchronous speed ω_s as shown in Eqs. 4 and 5, respectively.

$$\omega_s = \omega_r + \omega_{sl} \tag{4}$$

$$\theta = \int \omega_s dt \tag{5}$$

ω_{sl} depends upon slip gain K_{sl} which depends on motor parameters like stator and rotor per phase resistances R_s, R_r, per phase inductances L_s, L_r, mutual inductance L_m, number poles p etc. as shown by Eqs. 6 and 7

$$K_{sl} = \left(3 \times p \times L_m^2\right)/(2 \times L_r) \tag{6}$$

$$\omega_{sl} = K_{sl} \times i_{qs} \tag{7}$$

If the value of demand speed and rotor speed is known the value of i_d, i_q can also be calculated from machine parameters as shown by Eq. 8.

$$i_{ds} = \frac{\psi_r}{L_m}, \; i_{qs} = \frac{\psi_r L_r}{L_m R_r}(\omega_{ref} - \omega_r) \tag{8}$$

After knowing the value of θ the demand stator current is calculated using inverse Park and Inverse Clarke transformation which is compared with the actual current to generate the gate pulses, is given by the following Eqs. 9 and 10.

$$i_\alpha = i_d \cos \theta - i_q \sin \theta, \; i_\beta = i_d \sin \theta + i_q \cos \theta \tag{9}$$

$$i_a = i_\beta, \; i_b = i_c = (i_\beta + \sqrt{3} i_\alpha)/2 \tag{10}$$

3.2 IFOC Algorithm

For the development of the IFOC algorithm, pre-defined values and mathematical calculations are involved as stated in 3.1.1. IFOC algorithm implemented as follows:

1. Get the reference speed ω_r^* and flux.
2. Calculate reference value of i_d^* i_q^*, ω_{sl} and K_{sl} above Eqs. 6–8. Also, calculate the initial value of θ_0 from $\tan^{-1}(i_q/i_d)$.
3. Read the rotor speed from proximity sensor ω_r and two-phase current i_a and i_b and i_c can be calculated using the formula ($i_a + i_b + i_c = 0$) as the system is balanced.
4. Calculate the value of synchronous speed ω_s using Eq. 4.
5. Calculate θ using Eq. 5.
6. Add the value of θ with θ_0 or with 0 if it starts from zero.
7. Calculate the increment θ_{inc} by which θ will increase per sampling instance. If n represents the sampling instance $\theta_{(n+1)} = \theta_n + \theta_{inc}$. Where $\theta_n = (\theta + \theta_0)$.
8. Check whether the value of θ has crossed 360. If yes then divide it by 360. The remainder will give the correct angular orientation.
9. Calculate the actual value of i_q and i_d using Eqs. 2 and 3.
10. Compare value of i_q and i_d with the i_d^* i_q^*
11. Tune the inner loop and outer loop PI controllers
12. Calculate the demand i_a^*, i_b^* and i_c^* using Eqs. 9 and 10 and step 2.
13. Compare the value of i_a^*, i_b^* and i_c^* with i_a, i_b and i_c to generate the gate pulses.
14. Repeat the step if demand speed is varied from step1 or else from step 3.

4 Experimentation and Results

The VSI-IM based IFOC drive has been modelled and simulated in with the available nameplate data and motor test result and then has been implemented in hardware.

4.1 IFOC Based VSI-IM Simulation Model

The IFOC is at first modelled in a MATLAB/SIMULINK shown in Fig 2. At first, the system is simulated subjected to a sudden change in torque as shown in Fig. 3. The simulation was conducted for 4 s. Then, the system has been analyzed with change in speed as shown in Fig. 4. The VSI-IM has been configured as per the name plate data and motor test results.
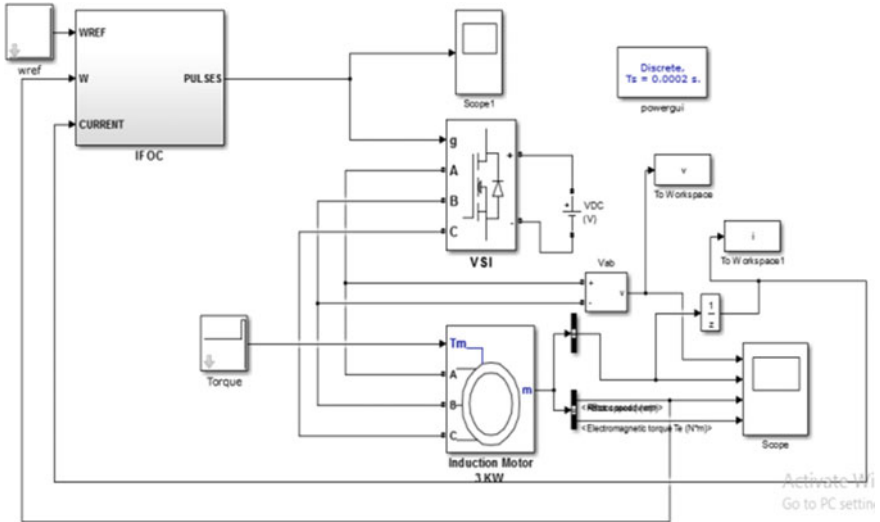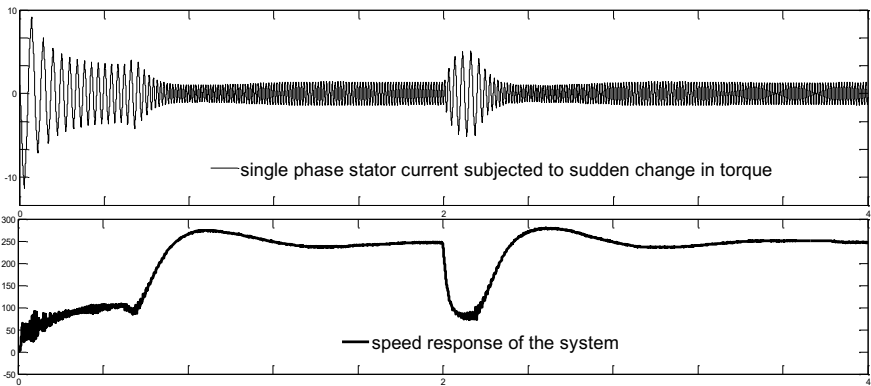
Fig. 2 Simulation model of IFOC-VSI-IM

Fig. 3 Variation of stator current and speed with sudden change in torque of the IFOC-VSI-IM

4.2 Hardware Implementation

For implementing IFOC, ADC interrupt subroutine is developed within which the instantaneous current and speed is sensed and all the mathematical operation is performed and is updated at every sampling instance. The ADC conversion rate has been kept at $12T_{ad}$. The PWM channels are provided with 0.00002 s dead time to avoid short circuit due to shoot through. The ADC buffer senses the current and capture input module in counter mode senses speed. The logic has been implemented via mplab IDE. The variation of θ with speed. The value of i_a, i_b, i_c has been tabulated in Tables 1 and 2. These values have been estimated by using counter registers.
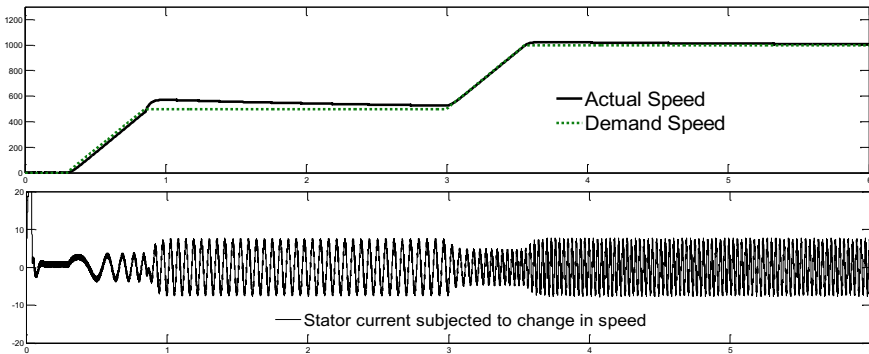
Fig. 4 Variation of stator current and speed with sudden change in speed of the IFOC-VSI-IM

Table 1 The angular position of the rotor flux vector (θ)

Speed in rmp	θmax	θmin	θincrement	No: of cycles of 360 rotation
600	169.466	−179.664	12.8811	3
1000	178.7463	−178.2448	12.8392	4
1200	178.67	−178.872	21.2335	6
1415	179.523	−178.972	25.3992	8

Table 2 Variation of stator current (ia, ib, ic) with change of speed

Speed	ia	ib	ic
600	2.9 to −3.1	2.99 to −3.1	−2.99 to 3.1
1000	3.4 to −3.2	3.45 to −3.2	−3.45 to 3.2
1200	3.26 to −3.25	3.25 to −3.25	−3.25 to 3.25
1415	3.15 to −3.2	3.15 to −3.2	−3.15 to 3.2

It has been seen that from the Figs 3 and 4 that the system is stable when subjected to change in speed and torque. The hardware system sights slight deviation of the values of current parameters as the IFOC–IM is not compensated i.e. the tuning of the PI controller has not been done. The value of θ varies almost from $-180°$ to $180°$ with sinusoidal in nature.

5 Conclusion

It has been seen that this IFOC-based VSI-IM performs satisfactorily both in transient and steady state. From the feedback information of current, speed and calculation of θ and other IFOC equations by the microcontroller is also quite accurate. The gate driver used is compatible with the microcontroller and no fault occurred. The

selection of modulation scheme is also done properly otherwise there could have been excessive waveform distortion in the stator current, harmonic generation and torque pulsation. Some minor errors are there in the hardware circuit as the PI controller parameters have not been tuned properly. This can be achieved in near future.

References

1. Bose BK Power electronics & drives. CRC press
2. Rashid MH Power electronic circuits, devices, and applications. Isbn: 978-81-317-0246-8
3. Hiware RS, Chaudhari JG (2011) Indirect field oriented control for induction motor. In: Emerging trends in engineering and technology (ICETET), 18–20 Nov 2011
4. Mon-Nzongo DL, Jin T, Ekemb G, Bitjoka L (2017) Decoupling network of field-oriented control in variable-frequency drives. IEEE Trans Ind Electron 64(7)
5. Jebali T, Jemli M, Boussak M, Gossa M, Kamoun MBT (2004) Dspace based experimental results of indirect field oriented control (IFOC) PWM VSI fed induction motor. In: International conference on industrial technology, IEEE ICIT'04
6. Senthil Kumar R, Mishra A, Asfar M, Shahzad M (2014) Speed control of field oriented induction motor using DsPIC controller. IOSR J Electr Electron Eng (IOSR-JEEE), 9(3):8, Ver. IV (May–Jun 2014)
7. Sulaimanl M, Patakorz FA, Ibrahim Z (2013) DSP based implementation of field oriented control of three-phase induction motor drives. Int J Res Eng Technol 02
8. Banerjee T, Bera JN, Choudhuri S, Sarkar G (2016) A comparative study between different modulations techniques used in field oriented control induction motor drive. In: 2nd international conference on control, instrumentation, energy & communication (CIEC) IEEE conferences, pp 358–362
9. Liu L, Du X, Shen S (2017) Indirect field-oriented torque control of induction motor considering magnetic saturation effect: error analysis. IET Electr Power Appl 11(6)

Smart Ambient Weather Monitoring System for Hikers

Mandira Biswas, Anup Dey, Sutanni Bhowmick, Bikram Biswas, Subhashis Roy and Subir Kumar Sarkar

Abstract The aim of this paper is to design and develop a user-friendly weather monitoring system that examines weather condition of the surroundings, and thus makes it available to the user from any part of the world. The proposed design is based on Internet of Things (IoT) platforms which come easily to connect the whole world on a network. With this system, it becomes simple and effortless to visualize the changing environmental conditions like relative humidity, temperature, and pressure. To aid hikers by providing information during hiking, a variety of sensors have been used to measure all necessary parameters and Intel Edison has been used to upload the sensor data wirelessly to the cloud. The recorded data can be observed on a web page in form of statistical graphs or on smartphones as required by the user. The device has numerous points of interest in terms of its smaller size, huge memory capacities, lower cost, and more prominent compactness.

Keywords Temperature sensor · Humidity sensor · Barometric pressure sensor · IoT · Wireless communication

M. Biswas (✉) · A. Dey · S. Bhowmick · B. Biswas · S. Roy · S. K. Sarkar
Department of Electronics and Telecommunication Engineering,
Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: mbiiswas@gmail.com

A. Dey
e-mail: anupetce@gmail.com

S. Bhowmick
e-mail: sutannibhow94@gmail.com

B. Biswas
e-mail: bikram94biswas@gmail.com

S. Roy
e-mail: subhashisaec@gmail.com

S. K. Sarkar
e-mail: sksarkar@etce.jdvu.ac.in

1 Introduction

A weather monitoring system is a facility, either land or sea measuring atmospheric conditions and provide us with information about weather parameters like temperature, humidity (SHT10), and a pressure sensor (BMP280). Weather monitoring system is an imperative part of human life [1]. Weather monitoring system can be either wireless or wired one, by collecting and analyzing data related to environment temperature, humidity, and air pressure over a wide area. The advantage of the wireless monitoring system over the wired system is easy to connect with other devices such as smartphones and web server for their application [2]. Human activities such as farming and transportation can be affected by weather changes. As a result of these features, an IoT device enables remote monitoring of the environment without the need to visit the site frequently. This can make even the possible observation of difficult geographical terrains. It can be able to reduce the manpower and helps to visualize the changing environmental conditions like relative humidity, temperature, and pressure. An IoT device can efficiently connect physical objects placed at a great distance from each other without the need for direct physical connection. Thus, using the Internet of Things (IoT), we can efficiently manage the sensor data, identify redundant sensor readings and visualize it with the help of web-based applications [3]. The aim of this paper is to design wireless weather monitoring system that is based on the concept for hiking purposes as the people who climb need to keep track of the atmospheric pressure using Intel Edison as an interfacing device as well as a web server. The weather condition is driven by air pressure differences between one place and another. We can automatically collect the information about humidity, temperature, and pressure through this system. For the mobile-based monitoring interface, Virtuino was used. Through this universal app, the user can access this measured data of a particular place at remote place from any display devices. The measured data is processed using a microcontroller-based system and made available wirelessly to the server for storage and access continuously [1]. The system sends the information about monitoring parameters like temperature, pressure, and humidity to the cloud. In the cloud, which shows the sensor data graphically and also allows sending of SMS to alert people when the sensor reading crosses a threshold (which can be programmed) [3–6].

2 Design of Proposed System

2.1 System Architecture

The proposed ambient weather monitoring system for hikers is as shown in Fig. 1. The proposed IOT based Weather monitoring system describes the weather parameters like temperature, humidity, and pressure levels through wireless communication. The presented design is based on the Arduino Intel Edison platform which is
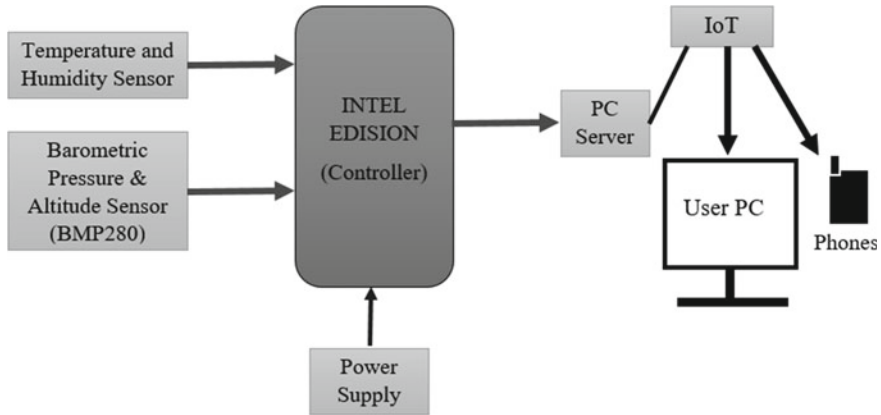
Fig. 1 Block diagram of a planned design

appropriate for the simple application. Unlike the processor-based Raspberry Pi, the Intel Edison is a microcontroller-based platform. As a result, it can be more specific in the application, optimizing memory and I/O resources and thus reducing its cost. The primary advantage of this design in terms of practicality lies in the fact that the design is made far to simpler by adding the flexibility in terms of user programmability. IoT devices based on Arduino Intel Edison have been designed in which data has to be accessed by entering the IP address or channel ID assigned to the device in the web browser [6]. It consists of two sections, one is a transmitter and another is receiving section. Transmitter section includes a controller, sensors but receiving section includes web server, and website.

The proposed system comprises of (1) the set of sensors that facilitate the monitoring of the ambient conditions, namely, pressure sensor, humidity sensor, and ambient temperature sensor, (2) a data transmission pathway through which the sensor data can be communicated wirelessly using a microcontroller and a wireless module, (3) a web-based application to visualize the sensor data.

3 Working Principle

The proposed system has been implemented by using "Intel Edison" microcontroller, which acts as a primary processing unit to make connections among the sensors for retrieving and further processing of data. Block diagram of a proposed system shown in Fig. 1 that consists of different sensor devices and other modules that are shown in Fig. 2. The proposed system uses Intel Edison wireless module that is mounted on an Arduino prototype. This sensor senses the weather data and then these data send to Intel Edison. Now the weather information is stored in Intel Edison, and then we access data, configure it for remote accessing via Wi-Fi [3]. In our system, the
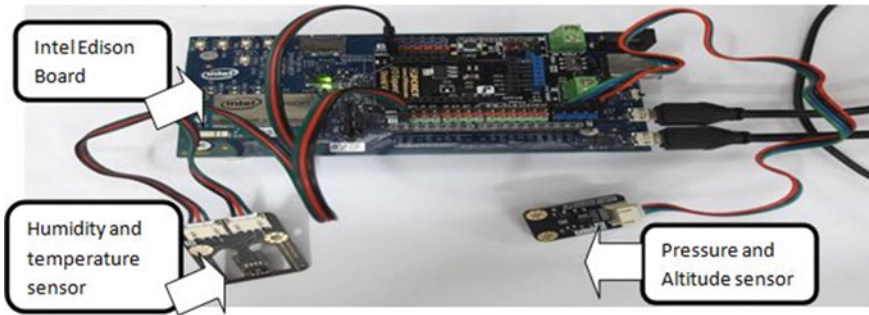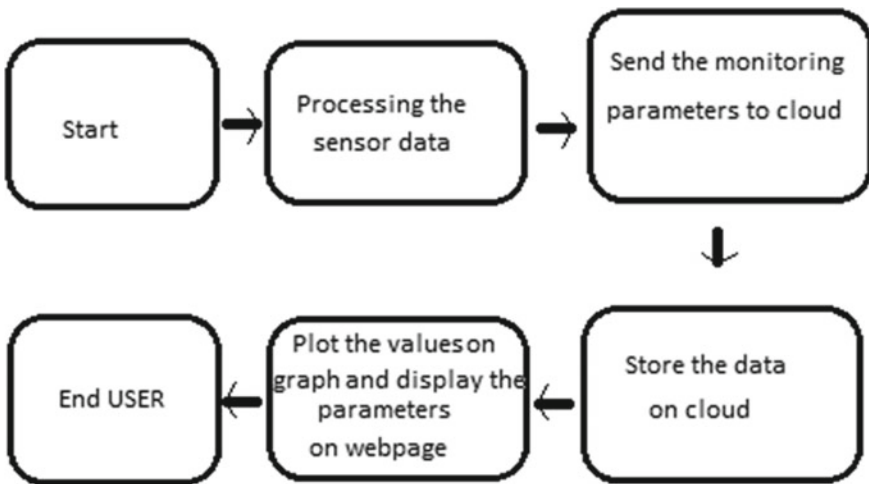
Fig. 2 Working prototype

Fig. 3 Flowchart

board runs on Linux OS with developmental support for Ardunio IDE. The board has been programmed to receive the readings from the sensors, generate an alarm when weather parameters fall below optimum threshold and upload the sensor readings onto the cloud. The program is pushed through the serial port of the Arduino prototype. Since it is Linux based, the PuTTY tool is used to configure the Edison Board with the Wireless Local Area Network (WLAN) to facilitate communication with the Internet [7]. All the sensor outputs are received by the controller and processed in necessary format. These are then transmitted using Wi-Fi to the main server. The relevant parameters get dynamically stored on the server at a faster rate which is simultaneously visualized on the web page and other interactive devices. For the mobile-based monitoring interface, Virtuino was used. Virtuino is an Android app which shows the sensor data graphically and also allows sending of SMS when the sensor reading crosses a threshold. The working model shown is in Fig. 2 (Figs. 3, 4).

Fig. 4 Intel edison computer-on-module [8]

Intel Edison—It is a master device in the proposed system; all the other devices like different sensors are connected to it. It can assist the microcontroller by making an interface between PC and microcontroller, with the help of USB cable. Moreover, to get started and maintain the power supply during operations, both of the battery and an option to connect AC to DC adapter are there.

Temperature and Humidity Sensor—SHT10 sensor is used to measure the temperature and humidity by utilizing a calibrated output signal in digital form. The signal acquisition technique can be very useful here and thus it guarantees high reliability and splendid long time period balance. This sensor has components which can measure resistive-type humidity and NTC temperature and provides a calibrated digital signal output [9].

Barometric Pressure Sensor—Gravity BMP280 is the sensor which is used to measure altitude and pressure by utilizing the piezo-resistive pressure sensors. For A/D conversions, a mixed-signal ASIC has been included. With a difference of ± 1 m in altitude, the sensor has a relative accuracy of ± 0.12 hPa [10].

ThingSpeak—ThingSpeak is an application platform for the internet of things. ThingSpeak permits to build a utility around information gathered by using sensors. It consists of real-time facts collection, facts processing, visualizations, apps, and plug-in. The open source IoT-based application ThingSpeak, which has API for both storing and retrieving data via HTTP protocol or LAN connection, can be used for various applications viz. "sensor-logging", "geographical-area tracking", and so on.

4 Result and Discussion

After sending the information from different sensors, these sensor senses the weather data and these data automatically sent to the web server via Wi-Fi. Figure 5 shows the sensor data reading of the ambient temperature sensor in degree Celsius and humidity sensor in % or relative humidity.

The proposed model has been successfully implemented to measure various parameters viz. temperature, humidity, pressure, and altitude as shown in Table 1.

A sample graph for temperature and humidity variation is shown in Fig. 6. The graphs can indicate seasonal variations. A rising trend of humidity can indicate a possibility of precipitation. The trend of temperature change can indicate seasonal variations. For a long time, temperature variations help to detect the global warming

A	B	C		A	B	C
created_at	entry_id	temperature		created_at	entry_i	Humidity
2018-02-24	285	27.339998		2018-02-24	285	50.84713
2018-02-24	286	27.419998		2018-02-24	286	52.30155
2018-02-24	287	27.429998		2018-02-24	287	52.3696
2018-02-24	288	27.359999		2018-02-24	288	50.04125
2018-02-24	289	27.349998		2018-02-24	289	50.39354
2018-02-24	290	29.129999		2018-02-24	290	51.85651
2018-02-24	291	31.339998		2018-02-24	291	59.70638
2018-02-24	292	32.269997		2018-02-24	292	63.00575
2018-02-24	293	32.789997		2018-02-24	293	66.4579
2018-02-24	294	32.989998		2018-02-24	294	67.68773
2018-02-24	295	33.219997		2018-02-24	295	68.77032
				2018-02-24	296	69.7206

Fig. 5 Cloud storage value for sensors data

Table 1 The measured parameters *from serial monitor*

Temperature (°C)	Humidity (%)	Pressure (Pa)	Altitude (m)
26.77	56.27	100724.20	50.59
26.70	56.05	100723.26	50.22
26.91	56.93	100714.63	51.07
26.95	56.96	100719.68	50.52
27.02	63.68	100711.60	51.20
28.53	66.05	100713.19	51.06
29.93	68.99	100711.41	51.21
31.19	74.31	100709.93	51.34

rate. The weather parameters uploaded by the device are available as user updates. Data can be accessed anywhere with the help of Internet-activated devices such as PCs, smartphones, tablets, laptops, etc. If the data is published, it can be accessed by anyone. The uploaded time-stamped weather data is seen in the screenshots shown in Fig. 7.

We used Virtuino which is an Android app which shows the sensor data graphically and also allows sending of SMS when the sensor the temperature and humidity values fall above or below predetermined thresholds shown in Fig. 7.
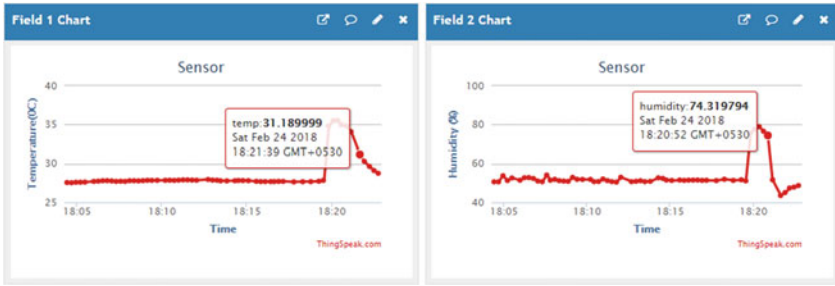
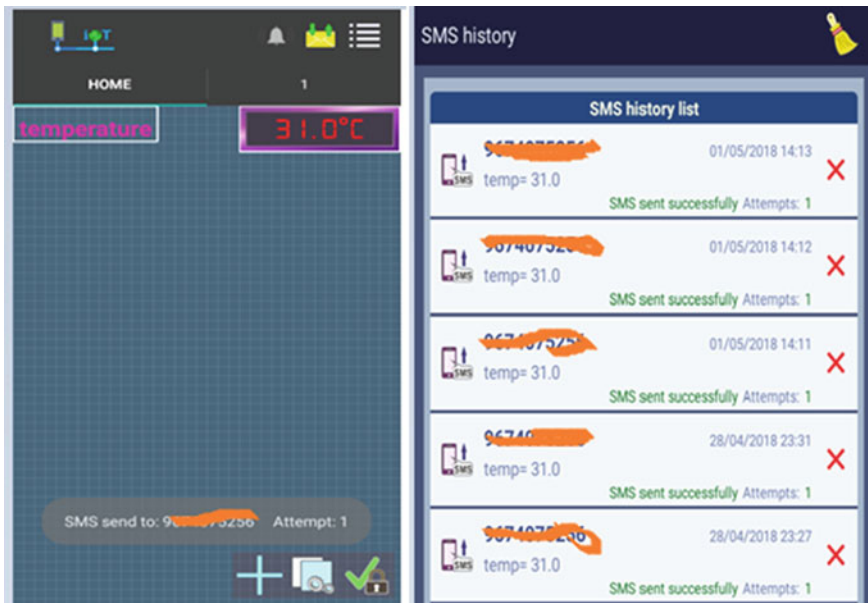**Fig. 6** Graphical representation of temperature and humidity

Fig. 7 Screenshot of monitoring parameter (Temperature) from mobile app

5 Conclusion

By deploying the weather monitoring system, the data can be stored online. Since all the parameters are displayed online, it can be made available to public for smart cities. We can inform hikers of active conditions such as various parameters like temperature, humidity, barometric pressure, and save the hiker's life through the network. This system provides a cheap, easy and real-time solution to monitor the weather parameters. The system helps in reducing human error. We conclude that the IOT-based weather monitoring system has successfully shown

real-time monitoring of weather through the mobile application and from this we can know the weather information remotely at any place. The Wi-Fi interface with the embedded application can be useful for exploiting an IoT application.

References

1. Satyanarayana KNV, Reddy SRN, SaiTeja PVYN, Habibuddin B (2016) IoT based smart weather station using Raspberry-PI3. J Chem Pharm Sci 10:1–6
2. Pawar P, Lahade S, Shinde S, Khatua P (2017) IoT based weather monitoring system. IJARIIE 3:1901–1905
3. Kulkarni P, Kute P (2016) Internet of things based system for remote monitoring of weather parameters and applications. Int J Adv Electron Comput Sci 3:68–73
4. Parashar T, Gahlot S, Godbole A, Thakare Y (2016) Weather monitoring system using Wi-Fi. Int J Sci Res (IJSR) 5:891–893
5. Patwardhan A, Sreedhar P, Henry R (2017) Portable weather monitoring station. Int J Electr Electron Data Commun 5:56–58
6. Krishnamurthi k, Thapa S, Kothari L, Prakash A (2015) Arduino based weather monitoring system. Int J Eng Res Gen Sci 3:452–458
7. Susmitha P, Sowmyabala G (2014) Design and implementation of weather monitoring and controlling system. Int J Comput Appl 97:19–22
8. https://software.intel.com/en-us/iot/hardware/discontinued. Accessed 20 Mar 2018
9. Malik A, Jalal A, Parray B, Kohli M (2017) Smart city IoT based weather monitoring system. IJESC 7:12123–12127
10. https://www.dfrobot.com/wiki/index.php/Gravity:_I2C_BMP280_Barometer_Sensor_SKU:_SEN0226. Accessed 20 Mar 2018

Implementation of a Temperature Control Process Trainer Through PID Controller Designed with Siemens S7-1200 PLC and HMI

Abhishek Kumar, Peeyush Garg, Ajay Shankar and Namrata Kar

Abstract This paper deals with the design of a temperature control process trainer interfaced with Siemens S7-1200 PLC and KTP-700 Basic HMI. The entire set-up comprises a temperature control trainer having an on–off controller as well as a PID controller mounted onto the instrumentation panel, a voltage to current converter which converts from 0–10 V to 4–20 mA, a temperature to voltage converter based on the RTD PT-100 sensor, PLC-based PID controller and HMI (Human Machine Interface). Initially, the efficiency, robustness, speed and effectiveness of the process control system are demonstrated via experimental results which are then compared with the simulation results.

1 Introduction

Process control is a part of control system which deals with the various mechanisms and algorithms required to achieve and maintain the output of any process within a specific range. Process control finds application in various industries. Temperature control is one of the most common branches of process control. A model of a temperature controller is presented in this paper. A PID controller (sometimes known as a temperature controller) is a device that controls temperature. This controller receives

A. Kumar (✉) · P. Garg · A. Shankar · N. Kar
Department of Electrical Engineering,
Manipal University Jaipur, Jaipur, India
e-mail: abhishekkumr@gmail.com

P. Garg
e-mail: peeyush01garg@gmail.com

A. Shankar
e-mail: nitrkl.ajay@gmail.com

N. Kar
e-mail: namratakar76@gmail.com

input from a temperature sensor (preferably an RTD or a thermocouple) and gives a stable output which is connected to a control element, such as a heater.

The PID controller is particularly favoured over most controllers as it more accurately and efficiently controls the process temperature. It compares the actual temperature to the required temperature (often called the set-point) and provides the output.

The paper is organized as follows. The structure of the control system is described in Sect. 2. Section 3 handles the modelling and simulation portions of the temperature control trainer. Comparison of the control system using both PLC and PID is presented in Sect. 4 and conclusion in Sect. 5.

2   Control System Description

The block diagram of the actual process control system and the proposed process control systems are shown in Fig. 1 and Fig. 2, respectively.

Initially, [1] the water tank having a capacity of 30 L is filled one-fourth with water. The check valve is closed in order to avoid loss of water through the pipes. The valve of the rotameter should be kept in a position such that the indicating pin is at 40 LPH (litres per hour).

Reference [2] The set-point of the PID controller is set at a particular temperature (say 75 °C) and the heater being the control element is switched on. The RTD PT-100 sensor connected to the heater senses the temperature of the water in the tank and sends a signal to the heater which in turn heats up the water until the set-point value is
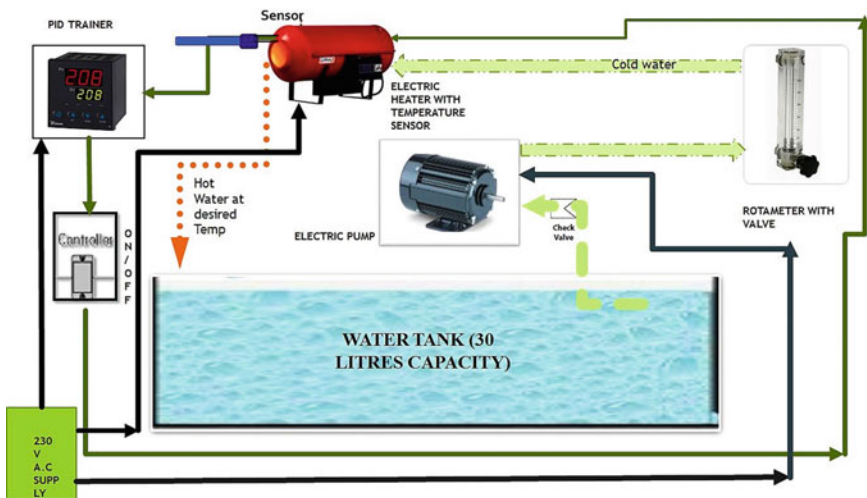

Fig. 1 Block diagram of actual Neeshionic's temperature control trainer
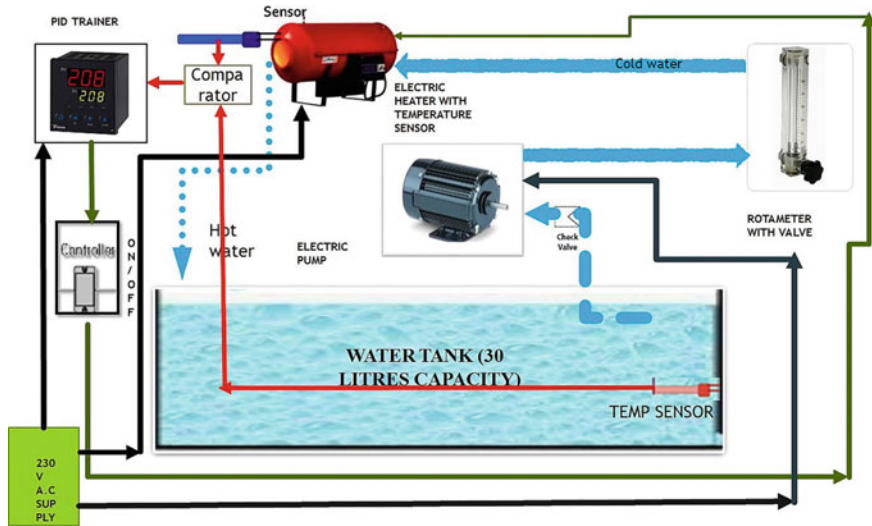
Fig. 2 Block diagram of proposed Neeshionic's temperature control trainer

reached. The heater eventually stops heating the water once the desired temperature is achieved.

The designed Neeshionic's temperature process control with PLC set-up includes the following components:

Neeshionic's Temperature control Set-up:

Reference [1] Multiprocess measurement and control system, a control for temperature loop, an SS water storage tank capacity of 30 L, ¼ HP monoblock electric pump for water circulation (½ in. × ½ in. inlet and outlet connection), ½ in. piping for water circulation, glass tube rotameter, connection ½ in. BSP socket end and microcontroller based electronic PID controller.

(a) Voltage to Current Converter (UT1101)
    This is a microcontroller-based device which converts a voltage range from 0–10 V to 4–20 mA. It also has a highly communicative serial (RS232/485) interface facility.
(b) Siemens S7-1200 PLC
    Reference [3] The Siemens S7-1200 PLC is easily adapted to suit the user's needs having CPU 1214C processor. They are scalable in terms of their performance and are equipped with integrated I/Os, integrated PROFINET interface for programming, distributed IOs and distributed drive archives.
(c) Siemens KTP-700 BASIC HMI
    Reference [3] The Siemens KTP-700 Basic HMI has touch/key functionality for intuitive operation, which provides an innovative user interface and improved usability
(d) RTD PT-100 SENSOR

Table 1 Comparison chart of the various temperature sensors [1]

Criteria	Thermocouple	RTD	Thermistor
Temp. range (°C)	−267 to 2316	−240 to 649	−100 to 500
Accuracy	Good	Best	Good
Linearity	Better	Best	Good
Sensitivity	Good	Better	Best
Cost	Best	Good	Better

The PT-100 RTD sensor has a linear increase in resistance with respect to the rise in temperature. It has a temperature measuring range of approximately −59 to 572 °F and hence has higher accuracy than other sensors-like thermocouple. Above is a chart, i.e. Table 1 showing the specifications of various temperature sensors:

(e) Temperature to Voltage Converter

A temperature to voltage converter based on the RTD PT-100 sensor which converts a temperature range of 0–100 °C to 0–10 V for the system.

(f) Neeshionics microcontroller based PID trainer

It is a microcontroller-based electronic PID controller operated at 230 V mains A.C. It has a 96 × 96 aluminium extracted DIN enclosure that includes a polycarbonate front fascia. Its input is 4–20 mA current and controlled output is 4–20 mA as well.

(g) Contactor and Relay

The contactor used in this experiment is a 3-pole device having a current of 9 A, 220 V flowing through it with input and 25 A output as 660 V. Relay is used for switching purpose of 10–220 V, 5 A.

3　Modelling and Simulation

Figure 3 shows the various experimental tools that are required for the completion of the project including the connections to each of them made by wires and also shows the functional block diagram and ladder logic of PID using PID_Compact on the TIA portal of PLC as well as the visualization of PID on HMI.

The PID temperature control module uses embedded design, module integration module control switch, blower, heat resistance, interference, RTD PT-100 measurement and temperature transmitter. In hardware configuration, the ladder programming of PID is carried out in the cyclic interrupt program block. The PID ladder logic is shown in Fig. 4.

The simulation of the PID_compact block is done by simple PID commissioning, where the various variables are set to our preferred value. After setting it enters into automatic operation mode. It records the temperature of the hot water via the sensor

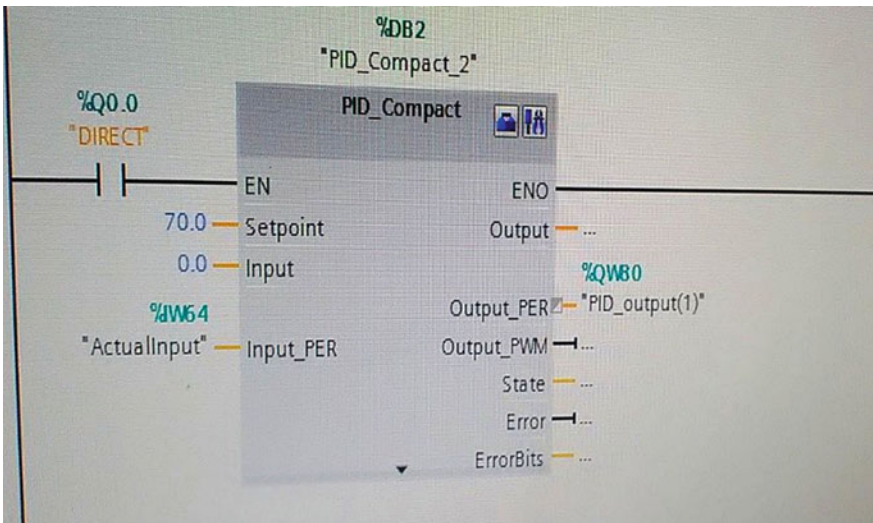**Fig. 3** The overall connection of the experimental tools

Fig. 4 PID ladder logic on the TIA portal

which is in turn connected to voltage converter that in turn sends in the analog voltage value to the PLC. During PID commissioning fine tuning is done which gives a zero percent (OFF) to hundred percent (ON) output. The various graphs that are obtained for this experiment are shown below.

In Fig. 5, the output is at zero percent, this is because the process value is near to 70.6 °C while the set value is at 70 °C.

After the PID commissioning is completed on the TIA portal, the PID controller is then designed on HMI (Human Machine Interface) [3, 4, 5, 6].
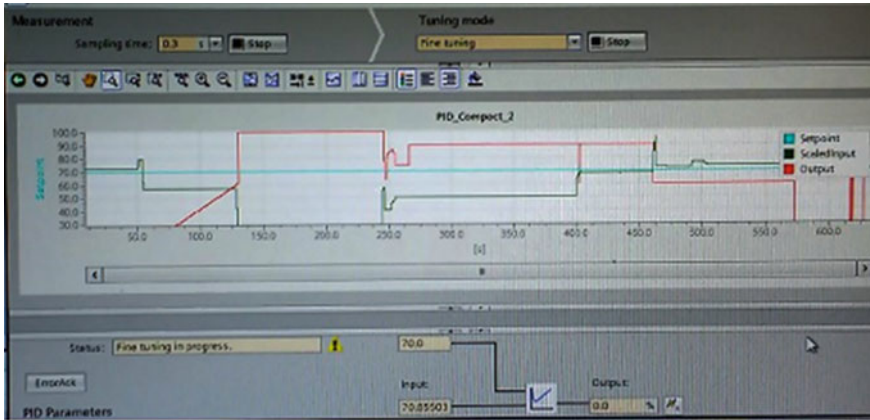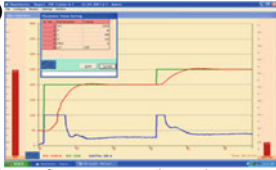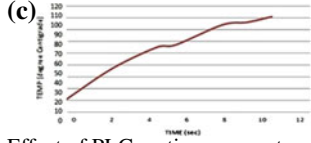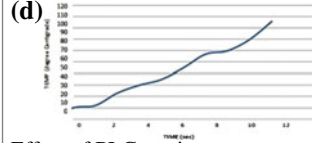
Fig. 5 Graph of PID_Compact with output

4   Comparison

By analyzing the comparison in Table 2 between Neeshionics PID (Graph a) and
PLC-based PID (Graph b), we can say that PLC gives reliable control. Actual system
output agrees on the theoretical result. PLC controller parameters have to be set in
the ladder diagram and can be adjusted in real time accordingly. We have taken
experimental values of PLC which is good enough compared to other controller
because of its discrete low error response characteristic. By comparing the graph of
c and d, we can say that the modified temperature control process trainer gives better
result compared to conventional Neeshionic's temperature control process trainer in
terms of response time for tracking the set-point.

5   Conclusion

The implementation of the model on PLC gave us an idea about the superiority
of PLC over PID. The modelling and simulation of the temperature control trainer
have been discussed in details. Both experimental and simulation results have been
recorded and compared. It is observed that the Neeshionics PID trainer kit is not as
efficient and effective as the PID model designed on the TIA portal of PLC. It is
also noted that the PID controller modelled on PLC via PID_Compact instruction
reduces error effectively as compared to the microcontroller PID controller inbuilt
in the temperature process control trainer. The temperature control kit has further
modified to receive much faster, efficient and robust output. To achieve this, a second
temperature sensor similar in nature to the already present RTD PT-100 sensor is
drilled into the storage tank containing which has a capacity of 30 L. The normal

Table 2 Comparison of Neeshionics microcontroller based PID and PID on PLC [7]

Specifications	Neeshionics PID	PID on PLC using FBD PID_compact
Robustness	Less robust as Neeshionics PID kit is designed with some IC, therefore, in extreme and dusty environment, it behaves in an erratic manner	PID on PLC is more robust than other rut processes
Graphs	(a) Effect of PID control on the Neeshionics PID trainer kit	(b) PID graph on PLC using PID_Compact block
Graphs	(c) Effect of PLC on time versus temp change for temperature control trainer	(d) Effect of PLC on time versus temp change for modified temperature control trainer

operation of the temperature control trainer will not be hampered by the addition of the extra sensor. In fact, once the temperature of the heated water inside the tank starts decreasing, the second sensor will send a signal to the electric heater to start conducting immediately, so that the temperature is maintained at the specific value. This will make the system more efficient and quicker on PLC as well.

We can also design a more robust system by using H-∞ techniques and implementing the concept of non-linear dynamics of the system.

References

1. Neeshionics Operation & maintenance manual for temperature control process trainer pdf. www.neeshionics.com
2. Neeshionics Operation & maintenance manual for PID trainer pdf. www.neeshionics.com
3. Siemens Siemens S7-1200 PLC and KTP-700 BASIC HMI manuals. https://support.industry.siemens.com
4. Engin D, Engin M (2013) Auto-tuning of PID parameters with programmable logic controller. In: IEEE international conference mechatronics and automation (ICMA), pp 1469–1474, 4–7 Aug 2013. ISBN: 978-1-4673-5557-5
5. Gabriela R, Mihai R (2016) Temperature control solution with PLC. In: International conference and exposition on electrical and power engineering (EPE 2016), Iasi, Romania, pp 20–22, Oct 2016

6. Gulpanich S, Krongratana V, Srimuang A, Wongvanich N (2017) PLC-based industrial temperature controller with different response times. In: 17th international conference on control, automation and systems (ICCAS). https://doi.org/10.23919/iccas.2017.8204291
7. Hasan Hafizul Haque SK, Imran Hassan HM, Abul Hossain SM (2014) Comparison of control system using PLC & PID. In: ASEE 2014 Zone I conference, University of Bridgeport, Bridgpeort, CT, USA, 3–5 Apr 2014

An Efficient and Enhanced Mechanism for Message Hiding Based on Image Steganography Using ECC-Cryptosystem

Mili Dhar and Subhasish Banerjee

Abstract With the rapid growth and advancement in computer technology, secure data communication over public networks become a major concern in todays era. Out of the various alternatives, image steganography is one of the key solutions for secure data communication over untrusted network. The main focus of this article is to define an embedded technique to hide the information in a digital image by which any adversary cannot predict the existence of secret information. And, to provide the higher security with smaller key size, ECC-cryptosystem has to be used. In this proposed method, LSB substitution strategies are used for higher imperceptibility. To prove the efficiency and security enhancement, a comparative study with existing schemes is also performed at the end based on the MSE and PSNR values.

Keywords Steganography · Cryptography · Elliptic curve · Circular queue · MSE · PSNR · LSB substitution

1 Introduction

In the current generation, where technology is progressing at a very rapid pace, staying up-to-date in security practices is of high priority. Protecting against editing and replication of digital images and prevention of tampering of multimedia files is an important concern. Transmitted data needs to be safe and secure to prevent undesirable access by unauthorized entities. Only an authorized entity should be able to access using their secret key. Steganography is a technique of hiding secret data into multimedia files like image,video,audio, and text. Instead of keeping the

M. Dhar (✉)
National Institute of Technology Arunachal Pradesh, Yupia, District Papum Pare 791112, Arunachal Pradesh, India
e-mail: milidhar28@gmail.com

S. Banerjee
Department of Copmputer Science and Engineering, National Institute of Technology Arunachal Pradesh, Yupia, District Papum Pare 791112, Arunachal Pradesh, India
e-mail: subhasishism@gmail.com

461

contents of the message secret, the entity from where the message originated is kept secret.

Steganography is formed from the Greek words "stegos" and "grafia" which defines [1] "covered writing". In steganography, a violator is not able to detect the secret message passing through the media channel. There are many techniques for hiding messages in multimedia files [2, 3] (audio, video, image, text, and documents). It can be said that to improve security, cryptography, and steganography are combined where cryptography deals with hiding information in order to protect the contents of the hidden message, whereas steganography uses the composition of these hidden messages.This combination facilitates not only hiding the message but also maintaining the security of content. If we use only cryptography, the secret message is noticeable by an entity that monitors the communication, but it cannot leak the information unless it knows the secret key. But if we add steganography to the mix, the entity will not be able to detect if some secret information is being transferred through the communication channel. Thus, the strengths of data hiding is amplified [4]. Steganography is of four types: audio, video, image and text (Fig. 1). Image is the most used medium for steganography because it is heavily exchanged on the Internet nowadays and another reason is the limited power of human visual system. Image steganography can be divided into two types, namely, spatial domain and frequency domain steganography. An image is a collection of pixels which shows different intensities of brightness. LSB method is useful to hide a message into an image as it does not change the image size, but the image file is noticeably distorted depending on the quantity of the message that is to be concealed inside the image. Based on the calculation of pixel value ranges, one can hide bits of different numbers in LSBs of distinct pixels instead of embedding a constant number of bits in the LSBs of individual pixels [5].
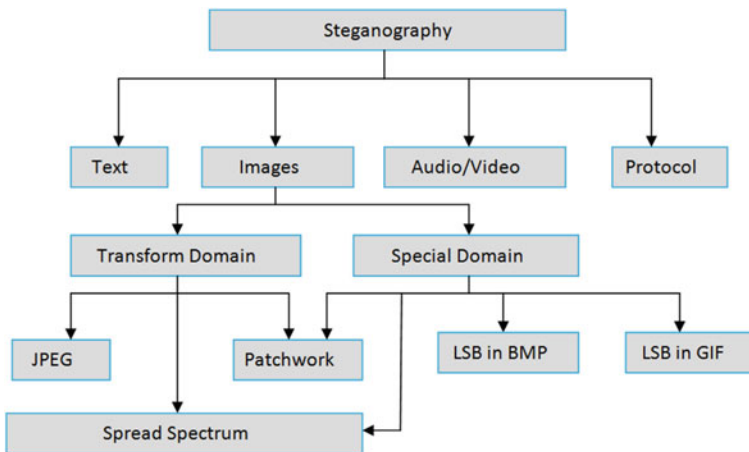

Fig. 1 Categories of image steganography

Wu et al. [6] proposed a palette based steganography scheme that reduces the root mean square (RMS) error iteratively between the original and its stego image. Their scheme is based on palette modification. In their method, one data bit is embedded into each pixel iteratively in a palette based image. Wang and Chen [7] presented a novel image steganography technique by using a two-way block matching procedure. Their method searches for the highest similarity block for data hiding and uses hop embedding scheme to hide secret image data inside a cover image. Swain and Lenka [8] presented steganography based on LSB array. Four arrays are selected and created depending on the range of the secret message. For different secret data, a different array is chosen and mapped on the array where matching is highest, thereby concealing the data. The approach by Younes and Jantam [9] is a new LSB insertion technique to hide data inside an image. Hidden data bits are overwritten into the LSB of each byte within the image. It shows that before and after insertion the correlation and entropy values of the encrypted image is the same. Nag et al. [10] proposed a steganography based on affine cipher algorithm and LSB substitution to provide robust security and imperceptibility to secret information. Nag et al. [10, 11] proposed an image steganography based on adaptive circular queue. In their method, RSA cryptography is used for confidentiality and authentication. For embedding, dynamic adaptive procedure is employed. Their scheme is efficient only if the size of the cover image is increased and the size of the secret data is decreased.

We have proposed a steganography method that gives two levels of security. Data is first encrypted by ECC-Elgamal public key cryptography and after that LSB substitution is employed for embedding encrypted message into the cover image. LSB substitution of the cover image for hiding data does not influence its image quality much and ECC is more secure and faster than other existing encryption algorithms. In our proposed method, a large amount of data can be embedded irrespective of the cover image size.

2 Flow Diagram of Our Proposed Methodology

Proposed algorithm consists of dynamic selection procedure for LSB replacement of the cover image and ECC-Elgamal public key cryptosystem to secure secret message as well as to create a secure crypto–stegno algorithm which is more secure than other systems being used for secretly sending data (Fig. 2).

3 Proposed Algorithm

1. First a grayscale image and a message are selected as a cover image and embedded message in the cover image respectively.
2. To encrypt the secret message, ECC-ElGamal encryption algorithm is used. This is a public key cryptosystem for transmitting the secret message P_m. To send the
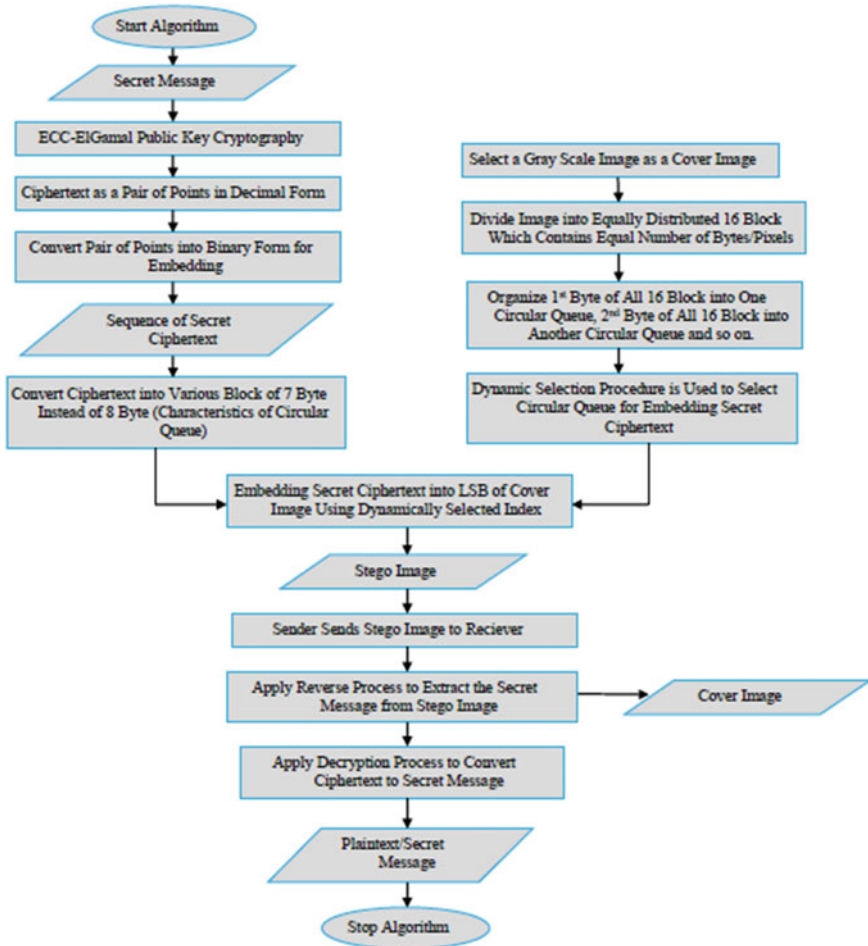
Fig. 2 Flow diagram of proposed work

message P_m to Bob, Alice selects a random key k and sends the pair of points $(k\beta, P_m + k(\alpha_B\beta))$ where $\alpha_B\beta$ is a public key of Bob and β is a base point defined over the elliptic curve (E) finite field F_q i.e. $\beta \in E$. This pair of points will be embedded inside the cover image.

3. Now, this pair of points are converted into binary bits and divided ciphertext into blocks. Each block contains 7 bytes instead of 8 because one slot is always empty for a full circular queue. Hence one slot will not be used further.This is a characteristic of full circular queue.
4. The detailed steps for converting points into binary bits is explain in [12].
5. Now, the cover image will be divided into 16 equal image blocks without altering pixels from 3045 to 4045 because image characteristics may be changed. Each

image block contains an equal number of bytes or pixels. Then, sort the first byte of all image block in one circular queue and second byte of all image block in another circular queue and so on.

6. For embedding message into image, dynamic selection procedure is used for selection of circular queue and data blocks. The first data block is selected dynamically based on the difference value in increasing order after subtracting 255 from the first byte of all the data blocks. After subtracting, the minimum difference value will be taken as the first data block for embedding process. And, for choosing circular queue dynamically, we subtract 255 from all the pixel values of second block which is actually the second pixel of circular queue. Based on the difference in values in decreasing order, the value with the maximum difference is taken as the first element in the circular queue for embedding.

7. Take the integer value of first four bits of selected data block to find the start index for dynamic selection of first embedding index in pixels of the image. And embed 5th–8th bit of LSB position in a pixel for better visual quality of a stego image.

8. To extract and decrypt the message by the receiver, reverse mechanism will be used.

4   Implementation of Proposed Method

Our proposed algorithm is divided into two phase. Phase 1 is for cryptography and phase 2 is for embedding the encrypted data into image. First, an elliptic curve is chosen over a finite field $p = 13$ by the following equation: $y^2 = x^3 + x + 1$
The generating points of elliptic curve on this finite field are shown in Table 1.

Therefore, the chosen curve contains 17 points with $\beta = (1, 4)$ which is the base point or we can say point generator of this EC curve. Suppose, Alice wants to encrypt and embed a message "hello" then she does the following.

4.1   Phase 1

ECC-Elgamal algorithm on elliptic curve domain parameter are (p, a, b, β, n, h) where β is the generating point and h is the cofactor, n is the order of elliptic curve. For example: $p = 13, a = 1, b = 1, n = 17$, and generally $h = 1$.

Table 1 Set of points on Elliptic curve

(1, 4)	(1, 9)	(4, 2)	(4, 11)	(5, 1)	(5, 12)	(7, 0)	(8, 1)	(8, 12)
(10, 6)	(10, 7)	(11, 2)	(11, 11)	(12, 5)	(12, 8)	(0, 1)	(0, 12)	∞

Key generation technique

1. select a random number $d = 3$ that $(d < n)$.
2. compute $\alpha = d * \beta = 3 * (1, 4) = (0, 12)$ which is the public key of Bob.
3. d is the private key and α is the public key of Bob.

Representation of message into elliptic curve points

Convert the message into ASCII values and then converts that values into points on the curve using Koblitzs method. For example,

1. Assuming that the curve parameters are $p = 13, a = 1, b = 1, n = 17, k = 10$ where p is the prime number in which the EC curve is defined over, a and b are the constant values, n is the order of the curve and k is an auxiliary base parameter which is known by both parties.
2. Let the message that is to be encoded is $S =$ 'h' $= 104$
3. Now check $x = Sk + 1, x = Sk + 2, x = Sk + 3$ under the mod p and so on.
   If $x = 0$, y does not exist.
   If $x = 1$, y exists.
4. So, the point will be $P_m = (1, 4)$.
5. Other messages will also be converted by the same manner.

Encryption

1. Alice gets the public key (α) of Bob.
2. Select a random number $e = 2$ which is the private key of the Alice $(e < n)$.
3. Compute $c1 = e * \beta = (8, 12)$ and $c2 = P_m + e * \alpha = (1, 4) + 2 * (0, 12) = (12, 8)$
4. $P_c = (c1, c2)$ is the pair of points of ciphertext.
5. Repeat the steps until all messages are not converted into points.

Generating the binary data sequence

1. β is the point generator with order $n = 17$ and let $m = 4$.
2. Convert the decimal numbers $(0 - n)$ using base-3 sequence generator format, as shown below:

   0000
   0001
   0002
   0010
   0011
   .
   .
   .

3. State a matrix P of (18×4) in the above form same as the matrix described in [12]
4. Circularly right shift each row of P by one element.

5. Encrypted message is $C = (P(8, 12), P(12, 8))$, where $x1 = 8$, $y1 = 12$ and $x2 = 12$, $y2 = 8$.

6. By applying algorithm proposed by [12] scheme, the sequence for the point $P(8, 12) = 2002$ and $P(12, 8) = 2011$ can be calculated. With the application of their algorithm points are converted to ciphertext, which is then transmitted in binary form by converting $0 \Rightarrow 00$, $1 \Rightarrow 01$, $2 \Rightarrow 10$. Therefore, the message to be transmitted is embedded into cover image as 1000001010000101.

4.2   Reverse Process of Phase 1

Reversing the data sequence into points

1. Gets the embedded secret message 1000001010000101 and reverse the binary sequence to get 20022011

2. From ciphertext, extract four digits: 2002 and circularly shifting left by one element: 0022.

3. Converts this sequence in to decimal form and store it in T i.e. 0022 will be in the form of: $0 * 3^3 + 0 * 3^2 + 2 * 3^1 + 2 * 3^0 = 8 = T$

4. To get the point $P(x1, y1)$, calculate $(T + 1)$ which is the index value of generating point matrix i.e. $P(x1, y1) = (8, 12) = c1$ represents $e * \beta$.

5. Other points are recovered by repeating the same steps.

Decryption

1. Compute $M = d * c1 = 3 * (8, 12) = (10, 6)$ where d is private key of Bob.

2. Compute $P_m = c2 - M = (1, 4)$.

3. P_m represents the message.

Representation of elliptic curve points into message

1. For decoding, compute

$$\frac{x - 1}{k} = \frac{1040 - 1}{10} \cong 104 = \text{`}h\text{'}} \tag{1}$$

4.3   Phase 2

In phase 2, cover image is divided into equally distributed 16 blocks with equal number of pixels without disturbing the pixel number 3045–4045 otherwise picture characteristic may be disturbed. Then create multiple circular queue in which put the first byte of all blocks in one circular queue and second byte in another one and so on. Subtract 255 from the values of second circular queues and sort it in decreasing order for dynamically selecting index value for embedding secret message in LSBs of the pixel of cover image.

5 Result and Analysis

In our novel proposed method, different cover image and stego image are considered which are simulated by using MATLAB. Images and their corresponding histogram are shown in Fig. 3. The difference between histogram of stego image and original cover image is very negligible when histogram is considered.

Estimating stego image with original image requires a quality measurement. Mean square error (MSE), peak signal to noise ratio (PSNR) and security for different
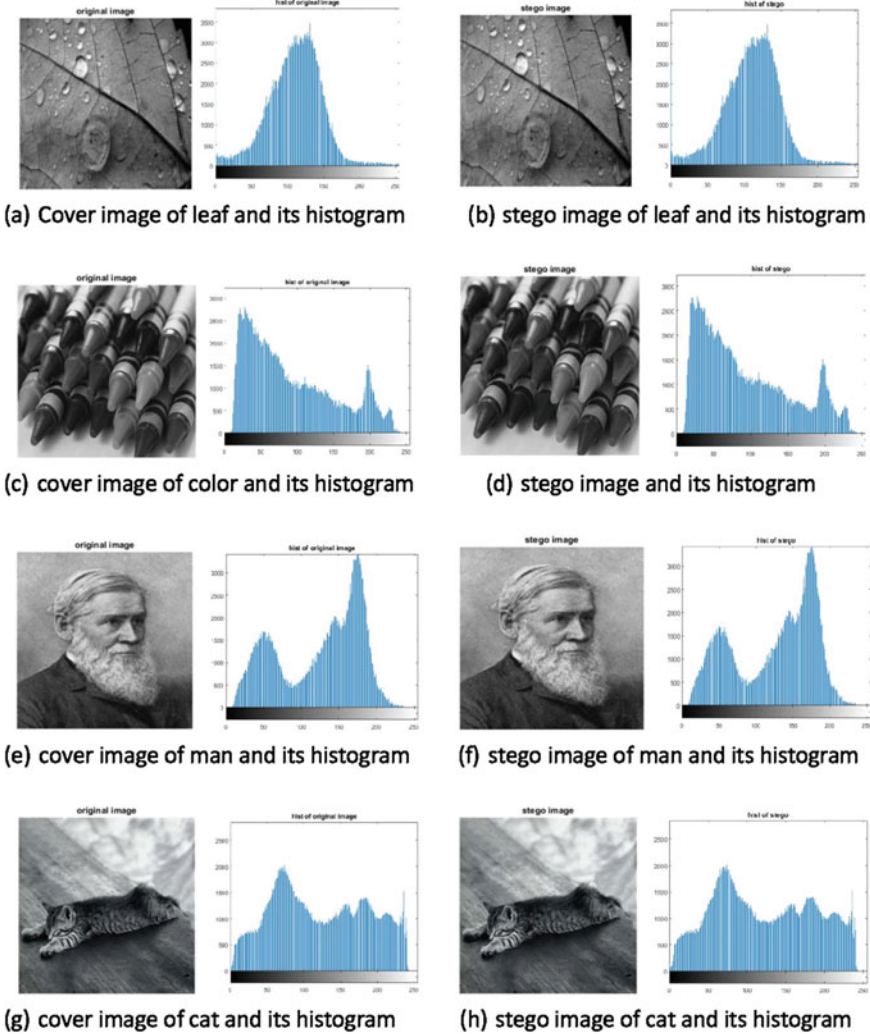

(a) Cover image of leaf and its histogram (b) stego image of leaf and its histogram

(c) cover image of color and its histogram (d) stego image and its histogram

(e) cover image of man and its histogram (f) stego image of man and its histogram

(g) cover image of cat and its histogram (h) stego image of cat and its histogram

Fig. 3 **a**, **c**, **e**, **g** is the original cover images and their histograms and **b**, **d**, **f**, **h** is the stego images and their corresponding histograms

images are commonly used measures for comparing image quality. MSE and PSNR are two error metrices that are used to compare the various image compression techniques.

5.1 Mean Square Error

Cumulative squared error between stego and original image is measured by MSE. The mathematical formula for MSE is:

$$MSE = \frac{1}{mn} \sum_{x=1}^{m} \sum_{y=1}^{n} (OI_{x,y} - SI_{x,y})^2 \tag{2}$$

where $OI_{x,y}$ and $SI_{x,y}$ are the original image and stego image respectively and m,n are the dimensions of the image and x, y are the coordinates of the image.

5.2 Peak Signal to Noise Ratio

PSNR is a measure of peak error which is measured in decibels (dB). In image steganography, generally the message is embedded inside the cover image such that the visual quality of the reconstructed image is not affected much. PSNR value reduces the visual attacks by human eye. Higher PSNR value points better visual quality of image due to lower distortion. A higher PSNR value is good because it means that the ratio of signal to noise is higher. Here signal indicates the original image and noise indicates the stego image. The mathematical formula for PSNR is:

$$PSNR = 10 log_{10} \frac{OI_{max}^2}{MSE} \tag{3}$$

where OI_{max} is the maximum pixel value of the original image. So, if we find a method which is having minimum MSE value and higher PSNR value, we can say that it is a better one.

5.3 Security

Security is used to protect data from unwanted attacks and intruders. Sometimes by active warden images are manipulated such as cropping or rotating that may destroy the hidden message, but as we use dynamic selection procedure secret data is scattered all over the image such that in case of manipulation the loss of data is minimized which improves the security.

Table 2 Observed PSNR and MSE values for our proposed algorithm on the basis of the ECC and RSA algorithm

Cover image	Size of cover image (in KB)	Quantity of ciphertext embedded (in bytes)	MSE value (using ECC based steganography)	PSNR value (using ECC based steganography) in dB	MSE value (using RSA based steganography)	PSNR (using RSA based steganography) in dB
Leaf	263	256	0.1719	69.3954	0.0154	75.4603
	263	880	0.1878	68.3039	0.1535	68.1297
Color	263	256	0.0103	70.4576	0.0020	71.8344
	263	880	0.0133	69.0316	0.0138	68.0343
Man	263	256	0.0028	69.9379	0.0266	70.4063
	263	880	0.0034	68.6841	0.8801	68.5222
Cat	263	256	0.0652	69.9989	0.0042	70.2718
	263	880	0.3122	68.8985	0.2072	66.3775
Leaf	1044	256	0.0022	81.0628	0.0306	61.3133
	1044	880	0.0032	80.2141	0.0922	56.1176
Color	1044	256	0.0091	82.1406	0.0033	70.7104
	1044	880	0.0122	80.4940	0.0126	65.9148
Man	1044	256	0.0018	82.9372	0.0011	81.3320
	1044	880	0.0033	81.0246	0.0029	80.9725
Cat	1044	256	0.0081	80.2097	0.0037	73.6933
	1044	880	0.0112	79.8262	0.0142	63.7023

Table 3 Comparison with other research for large data embedding in different image size

Name of the scheme	Avarage calculated PSNR (dB)	Security	Visual imperceptibility
Wang and chen [7]	44.25	Medium	Good
Nag et al. [10]	30.48	Very poor	Poor
Swain and lenka [8]	59.08	Medium	Good
Jain et al. [11]	67.22	Good	Better
Proposed algorithm	74.56	Very good	Best

After analysis of Fig. 3, one can easily observe that there is no visual artifact between the stago image and original image and their histogram. Result analysis of our proposed algorithm using various parameters is shown in Table 2. Table 3 shows the comparison of the other research. Using Fig. 4, it can be observed that embed small quantity of cipher into images with same size, PSNR value of RSA is better than ECC but embed large amount of cipher into images of same size, PSNR value of ECC is much more better than RSA. If the image size is large, it does not depend on the ciphertext size means PSNR value of ECC is always better than RSA with lower MSE value. For ECC, the difference between PSNR values of large and small quantity of cipher embedded is less compared to RSA. By result analysis, increasing the secret data size, PSNR value is increased up to 80.49 dB and MSE value is decreased by 0.0122. A high-quality image should aspire for 40 dB and above [11]. Higher the PSNR value means higher the performance. Security of secret transmission over the
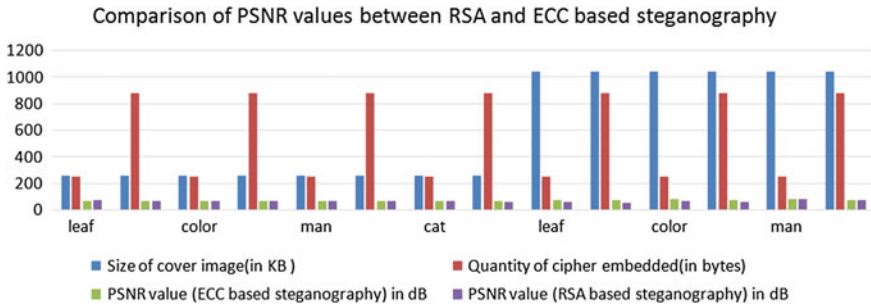
Fig. 4 Result analysis of proposed algorithm

Internet/transmission channel and confidentiality of secret message is also better as proposed algorithm uses ECC cryptography instead of RSA [13] and LSB substitution in a dynamic manner. Our proposed scheme gives a better imperceptibility, higher PSNR value and minimum MSE value which concludes better performance as compared to the other existing algorithms. Compared to other steganographic methods proposed by other researcher, our scheme is stronger and highly secured by ECC algorithm for securing any kind of secret message.

6 Conclusion

In this paper, to enhance the security in data communication, circular queue LSB substitutions based crypto–stegno scheme is proposed. The major objective of the proposed scheme is to provide two label security. First, an efficient ECC-cryptosystem is used to hide content of the message. Second, a strong circular queue LSB substitutions based embedded method are used to hide existence of the message in a digital image. To prove the superiority of the scheme, a comparative study with the similar existing approaches is also performed. After comparison with the experimental results and histogram analysis, it is observed that our scheme, not only maintains the confidentiality but also gets successful to achieve the higher PSNR and low MSE value, than others.

References

1. Cox IJ, Kilian J, Leighton T, Shamoon T (1996) International workshop on information hiding. Springer, pp 185–206
2. Provos N, Honeyman P (2003) IEEE Secur Priv 99(3):32
3. Hariri M, Karimi R, Nosrati M (2011) World Appl Progr 1(3):191
4. Morkel T, Eloff JH, Olivier MS (2005) ISSA, pp 1–11
5. Jain YK, Ahirwal R (2010) Int J Comput Sci Secur 4(1):40

6. Wu MY, Ho YK, Lee JH (2004) Pattern Recogn Lett 25(3):301
7. Wang RZ, Chen YS (2006) IEEE Signal Process Lett 13(3):161
8. Swain G, Lenka SK (2015) Int J Signal Imaging Syst Eng 8(1–2):115
9. Younes MAB, Jantan A (2008) Int J Comput Sci Netw Secur 8(6):247
10. Nag A, Singh JP, Khan S, Ghosh S, Biswas S, Sarkar D, Sarkar PP (2011) International conference on advances in computing and communications, pp 620–627 (Springer)
11. Jain M, Lenka SK, Vasistha SK (2016) Perspect Sci 8:417
12. Amounas F, El Kinani E (2012) Appl Math Sci 6(101):5039
13. Gururaja HS, Seetha M, Koundinya AK, Shashank AM (2014) Int J Appl Innov Eng Res Manag (IJAIEM) 3(1):111

Outdoor Obstacle Detection Module to Assist Visually Challenged

Anuradha Lakshmanan, S. Gayathri Devi, M. Meena Nisha and M. Dhanalakshmi

Abstract Autonomous navigation is of utmost importance for people suffering from visual impairment problems in their daily life. At present, a device known as Electronic Travelling Aids (ETA) is used by the visually impaired to detect static objects. As it is to be held by hand, it limits them from performing some of their activities. The objective of this project is to provide a hands-free module which focusses on dynamic objects (car) in an outdoor environment. This is achieved by using Raspberry Pi module, camera and ultrasonic sensor. The vehicle is captured by the camera in real time and detected using image processing techniques and Haar Cascade classifier. The ultrasonic module scans for the obstacle in three different directions, and then intimates the user to move away from the obstacle using an audio output, which is provided through headphones. The above-mentioned modules are incorporated into a waist belt. This prototype would be a better choice than the existing modules as it is long-lasting, hands-free, detects dynamic objects and helps the user in dauntless navigation in the outdoor environment.

Keywords Visual impairment · Raspberry pi · Haar cascade classifier · Ultrasonic sensor · Dynamic obstacle · Hands-free waist belt

A. Lakshmanan · S. Gayathri Devi (✉) · M. Meena Nisha · M. Dhanalakshmi
Department of Biomedical Engineering, SSN College of Engineering,
Chennai, India
e-mail: gayudevi9999@gmail.com

A. Lakshmanan
e-mail: anu03w@gmail.com

M. Meena Nisha
e-mail: mnisha1122@gmail.com

M. Dhanalakshmi
e-mail: dhanalakshmim@ssn.edu.in

1 Introduction

Vision impairment refers to when you lose a part or all of your ability to see (vision). This cannot be fixed by usual means such as glasses or contact lenses. 2011 WHO statistics show that there are 285 billion people in the world with visual impairment, 39 billion of people are blind, 246 billion are with low vision and 15 million people are blind in India [1]. However, it is estimated that the number of people with vision impairment could triple due to population growth and ageing. For example, by 2050, there could be 115 million people who are blind, up from 38.5 million in 2020. The statistics reveal that the causes of visual impairment are due to cataract (47.9%), glaucoma (12.3%) and age-related macular degeneration (AMD) (8.7%). Mobility is one of the major problems encountered by the visually impaired people in their daily life. By the roll of vision, people are able to move from one place to another. Without the vision, people will feel insecure while walking even in familiar places such as home, workplace. So, they need assistance for their mobility. It is impossible for a person to be always there to assist them. A common assistive device used by the visually impaired are the white cane. It is a mechanical cane which will provide tactile feedback when they sense an obstacle in front of them [2]. Ultrasonic sensors are widely used for the detection of obstacles in front [3–7]. Currently, a device known as Electronic Travelling Aids (ETA) is used to help the blind in managing a daily routine [8, 9]. The main drawbacks of ETA are that they are not long-lasting, they are uncomfortable to hold for long periods of time and they can detect only static, nonspecific objects [10]. Most of the assistive devices that are available in the market fail to provide for a hands-free environment for the blind. Hands-free environment will be convenient for the visually impaired and gives them a more sophisticated and aesthetic environment. Most of the devices are designed in a way that they are to be held in the hand and used indoors [11]. This will make the visually impaired to limit some of their activities as the normal persons do. This project aims to provide an economical, hands-free travelling guide to help the visually impaired get informed about an approaching dynamic object (car) in an outdoor environment [12]. It provides an improved navigation to the user by communicating to the user through voice alert or a buzzer output, thus enabling an easy and comfortable navigation to the user [13].

2 Methodology

Dynamic objects (cars) in the outdoor environment are detected by Raspberry Pi 3, Pi camera and ultrasonic module using image processing techniques. The classifier module used in the detection process is obtained using Haar training. Depending on the distance of the object calculated by the ultrasonic sensors, the user is intimated to move away from the approaching object via buzzer or audio output. These assistive
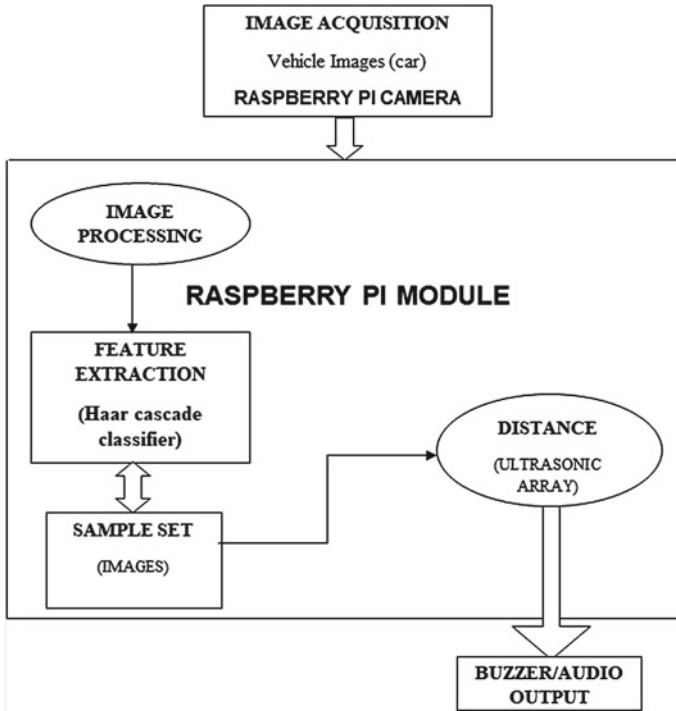
Fig. 1 Flow diagram

modules are incorporated into a waist belt, thus making it a hands-free module. Figure 1 depicts the flow of designing the outdoor obstacle detection module.

2.1 Block Diagram

The Raspberry Pi camera is used to capture the nearing car in real time. The resolution of the captured frames is changed using image processing techniques and it is then detected using Haar Cascade classifiers. On detection, the ultrasonic module is triggered and the distance between the vehicle and the user is calculated. When a specified threshold is reached, the user is provided with an audio output that specifies the direction in which the user has to move in order to avoid the vehicle.

2.2 Image Acquisition

The images used for training are acquired manually by taking pictures of multiple cars in different locations at various angles using a 5 Mp Omni vision Raspberry Pi camera module. The sample set, containing 2033 positive (car) and 1041 negative (non-vehicle) images is fed into the classifier for the training process. The resolution of the captured image is 2592×1944.

2.3 Image Processing

The image has to be processed before detection. On powering up the Raspberry pi, the video of the approaching car is captured in real time using Raspberry Pi camera at 1080 pixels and 30 frames per rate. The captured video is converted into image frames. Then, the grey scale conversion and alteration of resolution is performed to facilitate the detection process.

2.4 Haar Training

Haar feature-based cascade classifiers are a machine learning approach which is extremely rapid in detecting the objects in an image and video [14]. Every region of the image is analysed using a set of classifiers called Haar features that act as a funnel called Haar Cascade [15]. Figure 2 explains the generation of classifier module using the features extracted from an obtained sample data set containing large number of positive and negative images. The preliminary training and the acquiring of the trained file (XML module) are done in the Windows platform and using the Visual Studio 2017 software. It is later incorporated into the Raspberry Pi module using its terminal. The OpenCV computer program and its related modules are used for the real-time computer application.

Fig. 2 Steps involved in Haar training
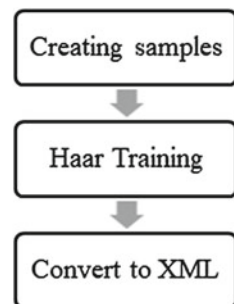
2.5 Distance Measurement

The HC-SR04 ultrasound module is used to measure the distance of the detected car from the user. A pulse of 10 μs is used to trigger it. The transmitter emits a high-frequency ultrasonic sound, which bounces off any nearby solid objects. Some of that ultrasonic noise is reflected and detected by the receiver on the sensor. That return signal is then processed by the control circuit to calculate the time difference between the signal being transmitted and received. This time it can subsequently be used, along with some clever math, to calculate the distance between the sensor and the reflecting object. Depending on the distance measured, the output is intimated to the user, either through buzzer or headphones.

3 Device Set up

The prototype to detect dynamic objects (car) in an outdoor environment employs the Raspberry Pi 3, Raspberry Pi camera, Ultrasonic sensors, buzzer module and any type of headphones.

3.1 Raspberry Pi Module

The Raspberry Pi 3 Model B is used which is a single-board computer with wireless LAN and Bluetooth connectivity. It is basically a mini computer which works just as fast as a home PC depending on the model. It has a Quad Core 1.2 GHz Broadcom BCM2837 SoC (system on a chip) which includes an ARM Cortex A-53 64-bit CPU and a 40-pin extended GPIO. A 1 GB RAM memory range is available on board. It consists of four ports, namely Full-size HDMI, CSI Camera port for connecting Raspberry Pi camera, DSI Display port for connecting a display, a MicroSD port for loading the OS and storing the data. In Raspberry Pi Model B, the Ethernet port is provided by a built-in USB Ethernet.

3.2 Raspberry Pi Camera

The Raspberry Pi Camera Board plugs directly into the CSI connector on the Raspberry Pi. It is able to deliver a crystal clear 5 MP resolution image, or 1080p HD video recording at 30 fps. The camera board attaches to the Raspberry Pi via a 15-way ribbon cable. There are only two connections to make: the ribbon cable needs to be attached to the camera PCB and the Raspberry Pi itself.

3.3 Hc-Sr04 Ultrasonic Module

The HC-SR04 ultrasound module is used for the purpose of calculating the distance between the user and the detected car. Pulse duration is the full time between the sensor outputting an ultrasonic pulse, and the return pulse is being detected by the sensor receiver [16]. The Python script measures the pulse duration, and then calculate the distance (Eq. 1) from this. This pulse duration is multiplied with speed of ultrasonic and then divided by two. Mathematically,

$$\text{Distance (m)} = \frac{\text{Time duration (s)} \ \times \ 34300 \ (\text{m/s})}{2} \tag{1}$$

where
Speed of ultrasound in air $= 34300$ m/s

3.4 Buzzer/Audio

SB-01 buzzer is initially connected to Raspberry Pi. Based on the obstacle detected by single ultrasonic, the buzzer module is activated. This emits a shrill sound on detection. In case of a single ultrasonic sensor, a buzzer output is produced. Once the car has been detected, the ultrasonic module determines its distance from the user. The buzzer then emits a loud beep thus alerting the user.

4 Results and Discussions

Conventionally, people suffering from visual impairment problems used walking sticks to navigate. At present, Electronic Travel Aids (ETA) is used to detect static objects. The aim of the project is to provide a hands-free module to detect dynamic objects (car) in an outdoor environment. This is done using Raspberry Pi, camera and Ultrasonic sensors.

4.1 Real-Time Output

The working of the prototype is tested in the outdoor environment using the Raspberry Pi and its camera module. This uses real-time streaming video, which means that the video capturing and the detection process are carried out simultaneously.
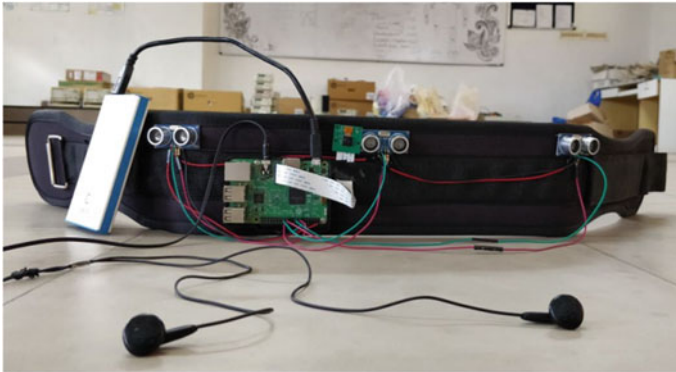
Fig. 3 Final prototype

4.2 Ultrasonic Output

Using the ultrasonic module, the distance of the detected obstacle is found. Once the car is detected, the ultrasonic module is triggered. The module detects the car up to a range of 5 m. Three ultrasonic sensors at three different angles ($\pm180°$ and $+90°$) in the form of an array are used. For single ultrasound, once the car is detected, the ultrasonic module determines its distance from the user. The buzzer then emits a loud beep thus alerting the user. In this case, the user is not given any useful commands to avoid the obstacle which is the drawback with this setup. So, the ultrasonic array setup is used. In the case of an ultrasonic array, an audio output given via headphones is used. This tells the user to move in a particular direction in order to avoid the car (Fig. 3).

4.3 Buzzer/Audio Output

The buzzer emits loud beeps as a moving car passes, provided it is within the ultrasonic range. This output is further improved using an audio output (Table 1) via headphones.

Table 1 Audio output

Object position	Sensor reading (m)	Reading output
Left	0–5	Move right
Straight	0–5	Move right or left
Right	0–5	Move left

4.4 On-Field Detection

All the supporting modules (Raspberry Pi module, Raspberry Pi camera and Ultrasonic sensors) explained in the methodology is incorporated into a waist belt (Fig. 3). The ultrasonic sensors are fitted onto the belt and focussed in such a way that all the three major directions ($\pm 180°$ and $+90°$) are covered.

During the on-field detection, the user is made to stand in an outdoor environment (road) with the belt worn around his/her waist. It is observed that for every car that approached the user, detection occurred and the audio output is heard. It states the proper direction that the user has to move in order to avoid the vehicle. It is also observed that the detection of other obstacles like static objects, people in motion and bicycles are ignored. The detection speed for multiple cars is found to be quick and there is minimum or no time lag.

5 Conclusion

The detection of cars in an outdoor environment is performed using Raspberry Pi module, camera and Ultrasonic array. The classifier module is generated by the features extracted from a manually obtained sample data set including positive (car) and negative (non-vehicle) images. It is initially tested on virtual platform followed by real-time detection along with reduced misclassifications. Single ultrasonic sensor and buzzer are interfaced with Raspberry Pi to detect and calculate the distance of the detected car in front of the user. An ultrasonic array of three ultrasonic sensors and headphones are interfaced to the Raspberry Pi to detect and calculate the distance of the car in three directions surrounding the user. An audio output intimates the user whether to move right (or) move left (or) move left or right based on the direction of the approaching car.

Further work is to add on to the level of accuracy in detection by using ultrasonic sensors with higher range and to integrate three Raspberry Pi cameras or pan and tilt setup to improve direction specificity. The algorithm may be extended to differentiate various dynamic obstacles in the outdoor environment so as to provide a more precise output.

References

1. Agarwal A, Kumar D, Bhardwaj A (2015) Ultrasonic stick for blind. Int J Eng Comput Sci 4(4):11375–11378. ISSN 2319-7242
2. Gurubaran GK, Ramalingam M (2014) A survey of voice aided electronic stick for visually impaired people. Int J Innov Res Adv Eng (IJIRAE) 1(8):342–346
3. Dambhara S, Sakhara A (2011) Smart stick for Blind: obstacle detection, artificial vision and real-time assistance via GPS. Int J Comput Appl (IJCA) 31–33

4. Faria J, Lopes S, Fernandes H, Martins P, Barroso J (2010) Electronic white cane for blind people navigation assistance. In: World automation congress, pp 1–7
5. Gayathri G, Vishnupriya M, Nandhini R, Banupriya MM (2014) Smart walking stick for visually impaired. Int J Eng Comput Sci 3:4057–4061. ISSN 2319-7242
6. Saaid MF, Ismail I, Noor MZH (2009) Radio frequency identification walking stick (RFIWS): a device for the blind. In: 5th international colloquium on signal processing & its applications (CSPA), pp 250–253
7. Niitsu Y, Taniguchi T, Kawashima K (2014) Detection and notification of dangerous obstacles and places for visually impaired persons using a smart cane. In: IEEE 7th international conference on mobile computing and ubiquitous networking (ICMU), pp 68–69
8. Dakopoulos D, Bourbakis NG (2010) Wearable obstacle avoidance electronic travel aids for blind: a survey. IEEE Trans Syst Man Cybern 40(1):25–35
9. Badre V, Chhabria R, Kadam T, Karamchandani K (2016) Ultrasonic blind walking stick with voice playback. Int Res J Eng Technol 03(04):1948–1951
10. Ando B, Graziani S (2009) Multi strategies to assist blind people: a clear-path indicator. IEEE Trans Instrum Meas 58(8):2488–249
11. Nakajima M, Haruyama S (2013) New indoor navigation system for visually impaired people using visible light communication. EURASIP J Wirel Commun Netw Springer Open J 697–706
12. Koley S, Mishra R (2012) Voice operated outdoor navigation system for visually impaired persons. Int J Eng Trends Technol 3(2):153–157
13. Nilsson ME, Schenkman BN (2015) Blind people are more sensitive than sighted people to binaural sound-location cues, particularly inter-aural level differences. Hearing Research 332:223–232
14. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Accepted conference on computer vision and pattern recognition, pp 133–145
15. Pavani SK, Delgado D, Frangi AF (2010) Haar-like features with optimally weighted rectangles for rapid object detection. Pattern Recogn. Elsevier 43(1):160–172
16. Kaur M, Pal J (2015) Distance measurement of object by ultrasonic sensor HC-SR04. Int J Sci Res Dev 3(05):503–505

Design of PID Controller for Maximum Power Point Tracking for PV Energy Systems

Saurabh Bhowmik, Raval Parth Pradip and Chitrangada Roy

Abstract Modeling of PID control system in DC–DC boost converter with variable input DC voltage at varying irradiation levels in the solar PV system has been implemented in this paper. The proposed model of PID controlled DC–DC boost converter prevents overshoot and oscillations of the output voltage obtained from it and tracks maximum power. Design and simulation of a PID control system is to soothe the output voltage of the DC–DC boost converter. The implementation results using constant irradiation and variable irradiation level with constant temperature are presented. Result and analysis supports the validity and advantages of the PID method.

Keywords DC–DC boost converter · PID controller · Irradiation · Maximum power point tracking · Solar cell

1 Introduction

Solar energy is the renewable form of energy, which can be utilized and stored in the form of voltage. However, due to its very low efficiency and unregulated characteristics makes it underrated. In the past few years, many new inventions and methods were proposed to overcome this problem. Solar Photovoltaic (PV) cell generates unregulated voltage from the solar radiations but it is very less [1]. To get the maximum power from the given system, its maximum power should be known by maximum power point tracking methods. DC–DC converters help to enhance or

S. Bhowmik · R. P. Pradip (✉) · C. Roy (✉)
Electrical and Electronics Engineering Department,
Sikkim Manipal Institute of Technology, Rangpo, Sikkim, India
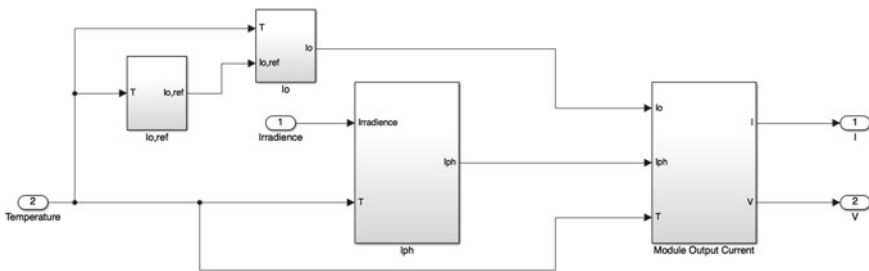e-mail: parthraval31@yahoo.com

C. Roy
e-mail: chitrangada.r@smit.smu.edu.in

S. Bhowmik
e-mail: nick2604888@gmail.com
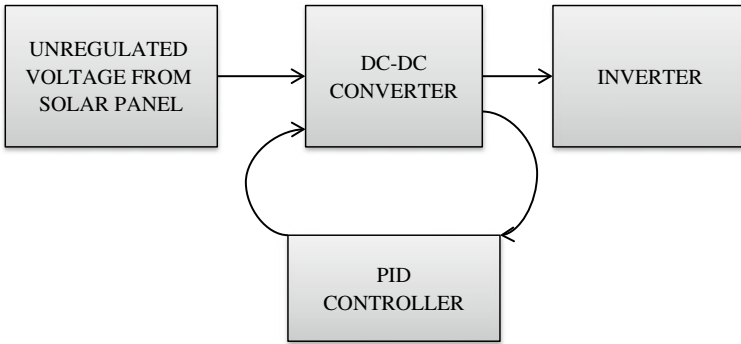
Fig. 1 General block diagram for control of DC–DC converter

Fig. 2 Basic modeling of solar cell

reduce the voltage that is generated from the solar cell to acquire maximum power at specific voltage. To get maximum power with the help of DC–DC converters different types of controllers are used. Here, PID and normal pulse generator are used to generate the specific voltage for maximum power (Fig. 1).

2 Modeling of Solar Cell

Modeling of solar cell for different irradiation and temperature is proposed in Fig. 2 [3]. Now accordingly, the generated voltage from the solar cell is fed to the DC–DC converter. This solar cell can be tuned for a specific purpose.

3 Designing of DC–DC Boost Converter

DC–DC converter increases or decreases the unregulated voltage from the solar panel [2]. Here, the main focus is on increasing the voltage, so the boost converter modeling
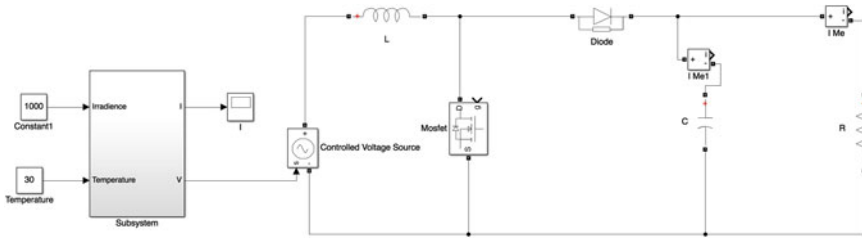
Fig. 3 Basic circuit for DC–DC boost converter

is proposed. A general model of the boost converter is shown in Fig. 3. Unregulated DC input voltage from solar cell, L is the induction, C is the capacitance, and output voltage will be taken across resistor R. The switching of MOSFET is given by the PID controller [3].

4 Designing of PID Controller

To control the switching of MOSFET in DC–DC boost converter, different types of controllers are used like P, D, PI, PD, PID, many more, and I. PID controller controls the switching of MOSFET to get the desired output voltage from DC–DC boost converter. Proportional, Derivative and Integral operations are used to reduce the error and for stability of step response in the system. These can be achieved by controlling the gains K_p, K_i, and K_d. The general equation for PID controller is given in Eq. 1 [4].

$$U(t) = K_p\, e(t) + K_i \int_0^t e(t)dt + K_d \frac{d\, e(t)}{dt} \ldots \tag{1}$$

Here, e(t) is the error in voltage output with respect to reference voltage. There are some fixed rules gain values, which is given in Table 1.

The generalized diagram for PID controller for DC–DC boost converter is shown in Fig. 4.

Here, constant 1 comes from output voltage, constant 2 comes from capacitor current, and constant 3 goes to gate of MOSFET in the boost converter.

Table 1 Rules for selecting gain in PID controller

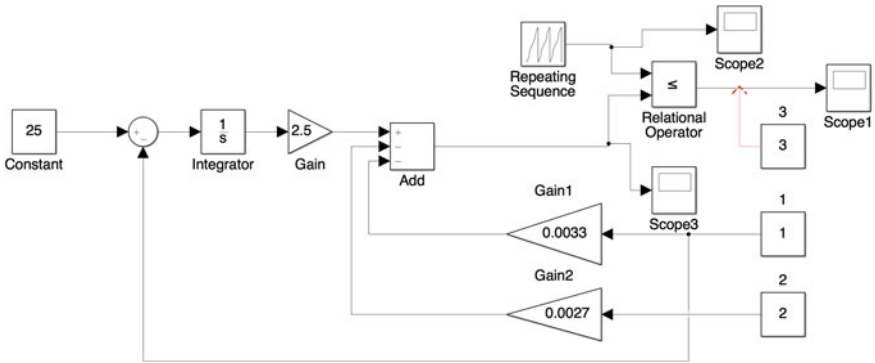| Parameter	Response speed	Stability	Output accuracy
K_p	High	Decreases	Good
K_i	Low	Decreases	Good
K_d	High	Increases	No effect

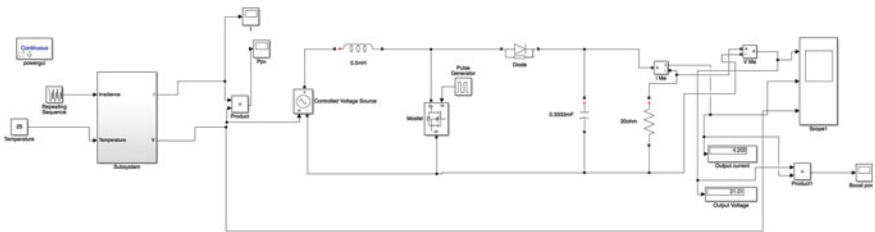**Fig. 4** Generalized diagram of PID controller for DC–DC boost converter

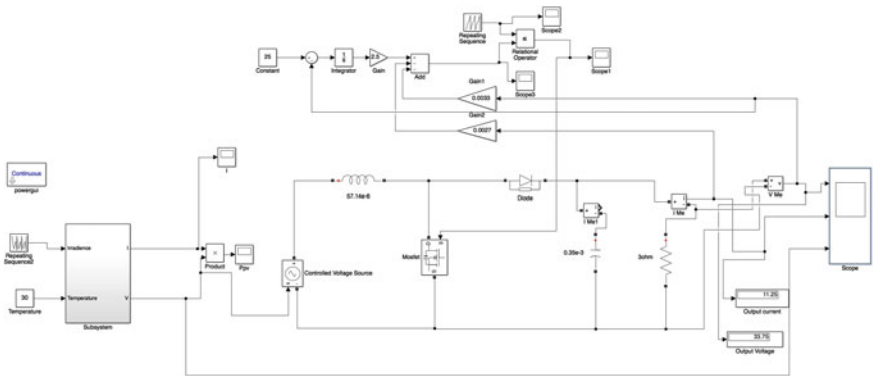**Fig. 5** DC–DC boost converter without using PID controller

Fig. 6 DC–DC boost converter with using PID controller

5 Simulation Results

For modeling and simulation purpose, we have used MATLAB Simulink software. Figure 5 shows the diagram of system without using PID controller and Fig. 6 shows the system with PID controller.

Table 2 System parameters of variables

System parameters	Corresponding values
Irradiation	[700, 500, 450, 550, 600, 430] W/m^2
Temperature	30 °C
Inductance	57.14e–6 H
Capacitance	0.35e–3 F
Constant	25
K_p	0.0033
K_i	2.5
K_d	0.0027
Input voltage	4–8.5 V
Output voltage	25 V

5.1 Parameters Values

Following are the values useful for the system in Table 2.

5.2 Simulation Results

Figure 7 shows the output waveform of system without using PID controller and Fig. 8 shows the output waveform system with PID controller.

The Fig. 7 gives three-output waveform at different irradiation levels. The third graph gives the output voltage of the solar cell which is varying in the range of
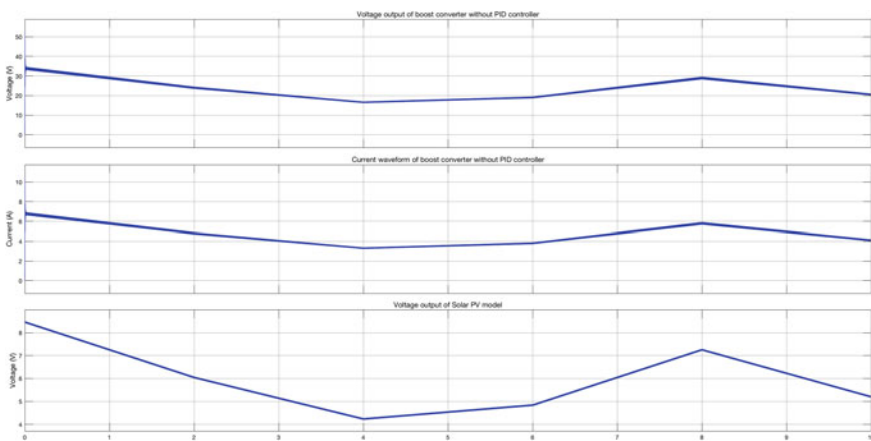
Fig. 7 Output waveform of the system without using PID controller
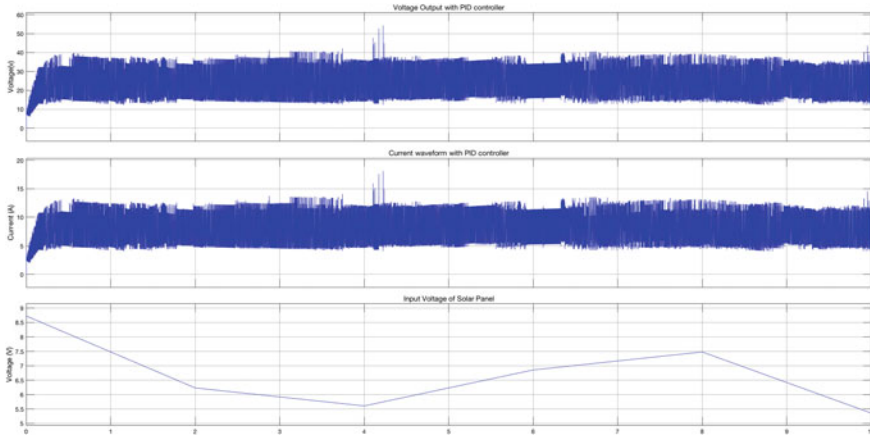
Fig. 8 Output waveform of the system with using PID controller

4.2–8.5 V at different irradiation levels and further, it is fed as the input of the boost converter. The unregulated variable voltage is because of the frequent changes in irradiation levels. The second graph gives output current of DC–DC converter with respect to time at different irradiation levels. The first graph gives the output of DC–DC boost converter with respect to time without using PID controller in the boost converter. Where it is observed that output voltage of the boost converter has increased and is varying in the range of 25–40 V at five different irradiation levels. In order to obtain maximum power irrespective to any available voltage from the solar cell, it is required to track the voltage to 25 V (as per calculation). Thus, the PID controller has been introduced to provide gate pulse and to drive the MOSFET of the boost converter, so to obtain maximum power.

The Fig. 8 gives three-output waveform at different irradiation levels using PID controller. The third graph gives the output voltage of the solar cell which is varying at different irradiation levels. The second graph gives output current of DC–DC converter which is almost linear with respect to time at different irradiation levels. The first graph gives the output of DC–DC boost converter with respect to time using PID controller in boost converter. Where it is observed that output voltage of the boost converter has increased and is almost linear (average of 25 V) at five different irradiation levels which is required to obtain maximum power from the system.

6 Discussion and Conclusion

As we can see that because of the variable irradiation levels that have been given at different instant of time, the voltage level changes with that. Now with normal pulse generator, the output voltage changes with great difference in voltages. Now with

the help of PID controller, we are able to achieve less overshoot and more constant voltage. The output waveform was magnified for better results and can be observed that it has less difference in output voltage.

Now with the help of comparing the results, it can be concluded that PID controller is more suitable for the linear system. It has overshoots and zero state errors but in minimal. We can achieve good results with less error with the help of PID controller for solar cell.

References

1. Design aspects of DC-DC boost converter in solar PV by MPPT algorithm—chapter-3
2. Fulmali V, Gupta S, Firoz Khan Md (2015) Modelling and simulation of boost converter for Solar_PV energy system to enhance its output. In: IEEE international conference of computer, communication and control (IC4-2015)
3. Bellia H, Youcef R, Fatima M (2014) A detailed modeling of photovoltaic module using MATLAB. NRIAG J Astron Geophys 3:53–61
4. Kumar N, Sharma N (2016) Improve performance of PV system by PID controller. Int J Sci Eng Technol 4(1)

Fleet Tracking Using Internet of Things

Yogesh Goyal, Spandan Buragohain, K. Srihari and Arun Kumar Singh

Abstract Vehicle tracking is a very crucial aspect of organizing and optimizing the fleet under consideration. The real-time tracking is done with the help of a hardware system, which uses GPS to acquire the data is then transmitted to a remote server. This data then can be retrieved by the user with the help of the mobile application.

Keywords NodeMCU · GPS · Fleet management · Google maps

1 Introduction

Vehicle has become one of the most crucial parts of the society, and this vehicle can be considered as private vehicles, fleet for transportation of goods, or the means of public transport. Vehicle becoming the crucial part also brings into the question of their security and optimized uses in the case of public transport and fleets. For this purpose, a tracking system is suggested which will provide the real-time location

Y. Goyal (✉) · A. K. Singh
Department of Electronics and Communication Engineering,
Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Majitar, Rangpo 737136, Sikkim, India
e-mail: gtkgoyal.16@gmail.com

A. K. Singh
e-mail: arunsingh.smit@gmail.com

S. Buragohain
Department of Computer Science Engineering,
Sikkim Manipal Institute of Technology,
Majitar, Rangpo, Sikkim, India
e-mail: spandanbg@gmail.com

K. Srihari
Department of Electrical and Electronics Engineering,
Sikkim Manipal Institute of Technology,
Majitar, Rangpo, Sikkim, India
e-mail: shrihari.hari420@hotmail.com

of the vehicle to the requesting user as per their needs. Vehicle tracking system is electrical system which are installed in the vehicles, and this system uses Global Positioning System (GPS) to get the location of the vehicle; a system to analyze this data, and then transmit it to the server unit [1–5]. Vehicle tracking helps in enhancing the security by constant monitoring of its location and can decrease the misuse of it. Fleet owner using the vehicle for transportation can use the tracking system to optimize the vehicle use by considering the availability of the vehicle at a given time as well as the progress of goods being transported. Vehicle tracking can also be used in public transportation or transportation provided by large organizations, a user can check the status of the vehicle at any time which can help a user the user to use their time effectively [6–8].

Vehicle tracking system has basically two components at the hardware end, first the GPS unit, this is used to get the location along with the time detail of the vehicle. The other includes the microcontroller unit which receives raw data from the GPS system, conversion of the data and then transmitting of data.

1.1 Global Positioning System (GPS)

GPS is a space-based satellite navigation system that provides the location and time of a person or vehicle or any device in every weather and anywhere on the Earth 24 hours a day. It is a constellation of about 24 satellites orbiting the earth at altitudes of approximately 11,000 miles (Fig. 1).

GPS satellites are orbited high enough to avoid the problems associated with land-based systems. It is a pillar for the global vehicle tracking system. The GPS receiver will receive the signal information from the satellite and with the help of triangulation; the exact location of the vehicle is traced. The GPS receiver must receive the signal with at least three satellites to calculate the 2D position and tracking movement. If GPS receiver is made to receive the signal for four or more satellites, then the
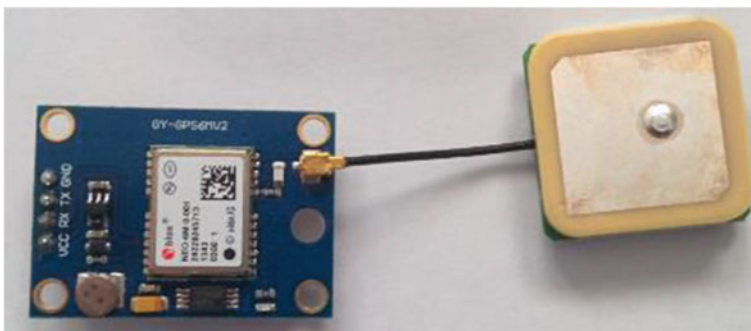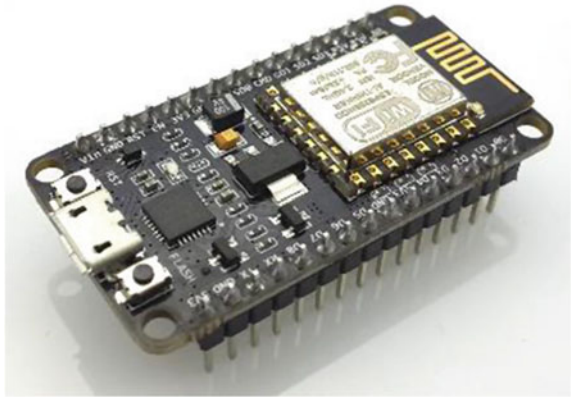

Fig. 1 GPS module NEO 6M with antenna [11]

Fig. 2 NodeMCU SOC [12]

receiver can calculate and the 3D position of the user is determined. Uncorrected positions determined from GPS satellite signals produce accuracies in the range of 5–100 m [1]. In this work, the GPS module used is U-Blox NEO 6 M by U-Blox, and this has low-power consumption and high level of integration capability with flexible connectivity [9]. When the GPS module receives data, it is then processed by the processing unit inbuilt or that to any other hardware.

1.2 NodeMCU

NodeMCU is an open source IoT platform. It is an eLua-based firmware for the ESP8266 Wi-Fi SOC from Espressif (Fig. 2).

It is interactive and programmable, and it helps in reducing the redundant work for configuration and manipulating hardware. It also integrates GPIO, PWM, IIC, 1-Wire, and ADC. This development kit has an FCC-certified Wi-Fi module, with PCB antenna [10]. Wi-Fi feature is provided by ESP8266 which enable the user to connect to Wi-Fi and to establish a TCP/IP HTTP communication with the server.

2 Methodology

2.1 Block Diagram

Block diagram in Fig. 3 describes the working of the system. Here, the processing unit is NodeMCU development kit. The code for this is written with the language C and was compiled using the Arduino IDE. The NodeMCU unit was connected to the GPS module U-Blox NEO 6M.
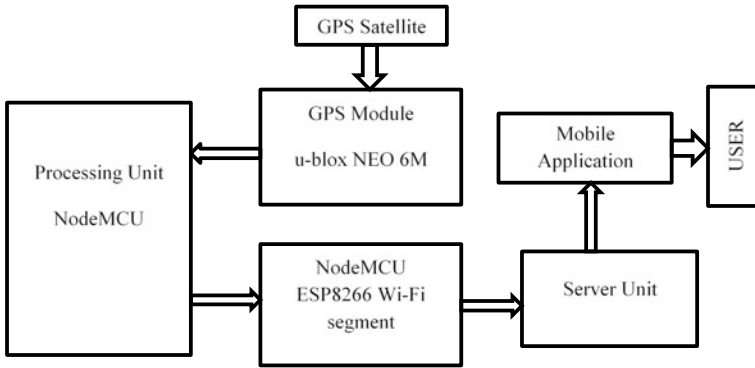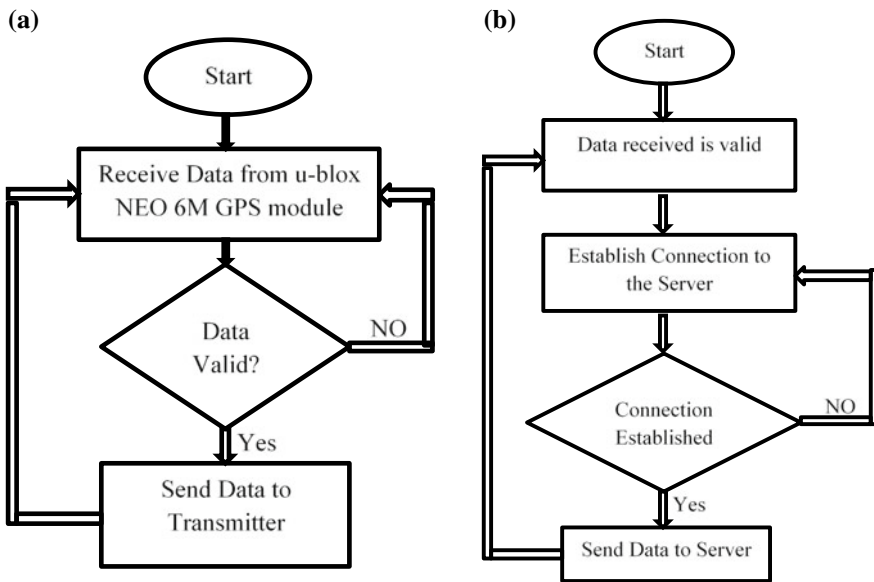
Fig. 3 Tracking device

Fig. 4 Flowchart **a** for data reception from GPS, **b** for data transmission to server

The GPS module gets connected to the GPS satellite so as to obtain the location information and time. This data is then transferred to the NodeMCU. NodeMCU then using the Wi-Fi module ESP8266 connect to a Wi-Fi hotspot and establishes the TCP/IP HTTP connection to an independently made server. After the connection is established, the data is sent to the server using the GET method. The system then waits to receive the acknowledgement from the server. Figure 4 represents the flow of the process.

2.2 Back-End: Server and Application

The server was made independently. When the server receives the request for the hardware unit to establish a connection, a PHP script authenticates the hardware unit, and this authentication is done with the help of a variable key. Once the process of validation is complete, the data is received from the system and the location data along with the time of data being received is stored in the database.

This stored data can be viewed by the user using a mobile application. This mobile application upon execution requests the server for the latest location received. Server upon receiving the request also authenticates that the user has the authorization to access the given set of data or not. This process of authentication is also done by a PHP session after which it queries the database for the newest entry of the location data and the time. This data is then phrased to create a JSON object and is sent to the requesting application. When this is received by the application, it parses the object to extract the longitude, latitude and time data. The application then adds a mark with the obtained set in the Google Map API and updates the UI.

2.3 Practical Model and Result

Figure 5 shows the data as observed on the server. Figure 6 shows the practical model developed for the purpose of data transmission and Fig. 7 shows the result obtained on the user side mobile application.
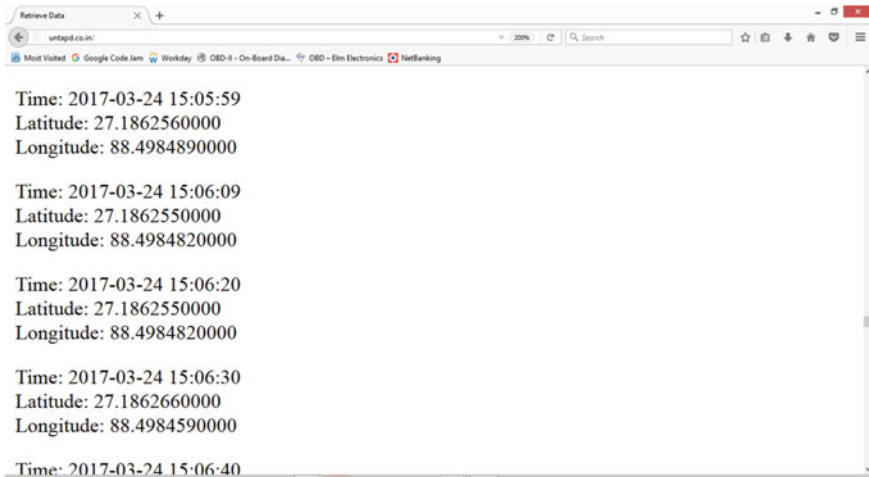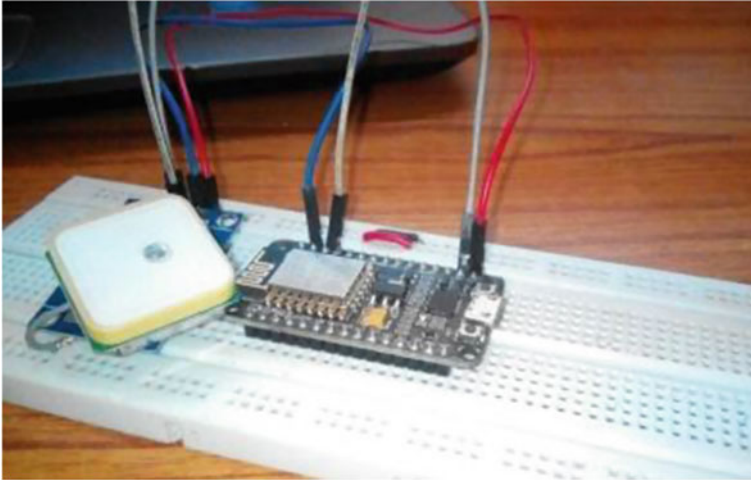

Fig. 5 Data as observed on the server

Fig. 6 Practical model of the transmitting unit

Fig. 7 Data set plot of the received data, user end

3 Conclusion

The practical system for the purpose of vehicle tracking was designed and implemented in the project. The GPS module at times may provide incomplete, insufficient data, as observed in Fig. 6. Where there are no dots representing the change of vehicle location. This may also be attributed to the loss of connectivity between the server and the hardware unit.

References

1. Sahitya S, Swetha N (2014) Real time vehicle tracking system using GPS and GPRS. Int J Res Comput Commun Technol 3(10):1329–1333
2. Dukare SS, Patil DA, Rane KP (2015) Vehicle tracking, monitoring and alerting system: a review. Int J Comput Appl 119(10):39–44
3. Rahman M, Mou JR, Tara K, Sarkar I (2016) Real time Google map and Arduino based vehicle tracking system. In: International conference electrical, comput & telecommunication engineering, vol 2
4. Kodavati B, Raju VK, Rao SS, Prabu AV, Rao TA, Narayana YV GSM and GPS based vehicle location and tracking system. Int J Eng Res Appl 1(3):616–625
5. Verma P, Bhatia JS (2013) Design and development of GPS-GSM based tracking system with Google map based monitoring. Int J Comput Sci, Eng Appl 3(3):33–40
6. Jayanth A, Hemachandra S, Suneetha B, Prasad BG (2013) Vehicle tracking system using GPS. Int J Sci Res 2(5):104–105
7. Lee S, Tewolde G, Kwon J (2014) Design and implementation of vehicle tracking system using GPS/GSM/GPRS technology and smartphone application. In: 2014 IEEE world forum on internet of things (WF-IoT), pp 353–358
8. Pooja S (2013) Vehicle tracking system using GPS. Int J Sci Res 2(9):128–130
9. UBlox NEO 6M. https://www.u-blox.com/en/product/neo-6-series
10. NodeMCU. https://nodemcu.readthedocs.io/en/master/en/
11. GPS Module NEO 6M with antenna. https://developer.mbed.org/media/uploads/edodm85/board_gps.jpg
12. NodeMCU SOC. https://images-na.ssl-images-amazon.com/images/I/71GUeTwFn5L._SL1088_.jpg

IOT-Based Navigation for Visually Impaired Using Arduino and Android

**Aashis Kumar, Monika Jain, Rahul Saxena, Vidyanshu Jain,
Siddharth Jaidka and Tushar Sadana**

Abstract Like many other projects, ultrasonic sensors have been used in devices that enable a visually impaired person to tackle obstacles. However, little has been done to expand the navigation in a broader context, and to enable the visually impaired person to go about hands-free and connect with the world like normal people do. This paper is derived from the in-depth study of the real-life scenarios and problem faced and proposes an advancement of the existing smart stick technology. The system is designed to act like a navigation guide through voice outputs consisting of ultrasonic sensors, Arduino Uno R3, HC05 Bluetooth module, and an Android application. The Android application is used as a navigation tool to calculate the shortest distance from the source and the location, and the ultrasonic sensors attached to the person's knees look out for obstacles while the blind person walks as per the directions narrated to him on an earphone, and upon detection of an obstacle, the HC05 triggers a warning and the Android application describes the environment to the person. With this project, we seek to take a step closer in narrowing the gap between blind people and us while keeping it within the reach of even the poorest.

Keywords Blind · Arduino · Ultrasonic sensor · Android application ·
Navigation · Environment description

1 Introduction

Visually impaired people unlike us have a hard time interacting with the environment, especially because of their inability to judge their relative position as a reference in an area. They face a hard time getting employment and have to hugely rely on their friends and family for even the smallest of the task.

A. Kumar (✉) · M. Jain · R. Saxena · V. Jain · S. Jaidka · T. Sadana
Manipal University Jaipur, Jaipur, India
e-mail: aashiskumar986@gmail.com

India houses to more than half the blind population in the globe [1], and the stats are ever rising. Science has made a lot of progress but has failed to come up with a satisfactory and feasible solution to cater the need of blind people. IoT brought about many devices that sought to bridge the gap between the blind and the normal people. Some of the research led to the development of smart sticks that allowed them to tackle hurdles, while some used beacons to let the blind know what they are touching or what is around them.

The beacon method proposed by Korial and Abdullah [2] proposed a method to implement beacon system that a smartphone can capture and let the blind person know the things around them. The proposal was to install low-power consumption beckons in an indoor environment. This method had a major flaw, as time passes by the person gets acquainted with the things around the room and navigates around with a lot fewer difficulties. Also, the time taken to configure the beacons and their low scalability would make anyone give a second thought.

Shoval et al. [3] developed a NavBelt that was a portable wearable computer designed to help the blind avoid obstacles indoor. The information was conveyed to the blind by different audio sounds. The method did not clearly distinguish between sounds, though an ideal method to navigate indoor this method did not convey the momentary position of the person in the room nor made the sounds easily distinguishable.

Smart stick using infrared sensors for indoor navigation was an idea proposed by Innet and Ritnoom [4]. Upon detection of the obstacle, the stick produced different vibrations, and the vibration pattern was hard to distinguish and moreover made the person unable to use both his hands freely.

All the recent developments were either for indoor use or if suitable for outdoor used probes that enabled the blind person to use both of his hands freely. Our project aims to implement a cheap navigation tool that is easy to scale and can be used indoor and outdoors easily. We aim to use the power of smartphone and microcontroller at its fullest to enable GPS navigation, obstacle detection, and allow the blind person to get a description of the environment or what exactly is in his way through his smartphone, that now even the poorest can afford.

2   System Design

The system consists of an ultrasonic sensor attached to the blind person's feet about 25 cm from the ground. The sensor has a physical connection to a sling bag coated with reflective tape which makes it easier for the vehicles or other people to notice the blind person. The sling bag houses an Arduino and his Android phone is communicating through Bluetooth. Upon detection of an obstacle, the Arduino sends an alarm to the Android device and a voice command can be heard by the blind person that a danger is present forward. The blind person can use his phone camera (triggered by a click of earphone's button) to take the picture of his surroundings, which then returns a statement in simple languages as audio commands. The blind
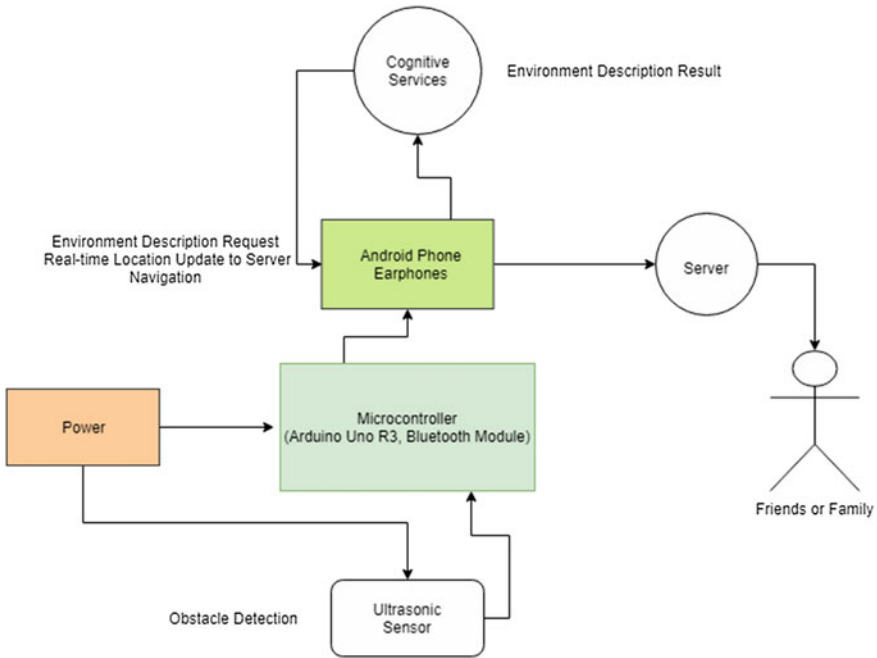
Fig. 1 Proposed model

person can use his phone to hear audio commands to the destination he wants to go to. The phone also broadcasts real-time location of the person to the server which would be visible to his friends and family. Also, the app broadcast an SOS message upon triple click of earphone's button to everyone on his SOS list.

2.1 System Configuration

The proposed system follows the architecture as shown in Fig. 1.

2.2 Microcontroller

Arduino Uno is based on ATmega328P. The board consists of 14 input/output pin, 6 analog inputs, a 16 MHz and a quartz crystal [5]. The microcontroller can be connected directly to the computer. The microcontroller has a low power consumption of 5 V and a flash memory of 32 KB. Arduino easily interfaces with Android phones (Figs. 2, 3).

Fig. 2 Arduino Uno R3

2.3 Obstacle Detection Unit

<u>Ultrasonic Sensor</u>: The module sends 40 kHz waves and detects for any pulse signal deflected back. The module used, HC-SR04 provides distance measurement of 2–400 cm with an accuracy of 3 mm. The module is attached to the person's leg. About 25 cm above the ground, it continuously looks for obstacles and sends a trigger message as soon as there is an obstacle detected based on the flowchart shown in Fig. 4. The obstacle is decided based on the following formula: Test distance $=$ (high $-$ level time velocity of sound (340 M/S)/2.

2.4 Warning Unit

<u>Bluetooth Module</u>: The HC-05 module provides a transparent wireless serial connection setup. The module is fully qualified Bluetooth V2.0 + EDR 3 Mbps Modulation with 2.4 GHz radio transceiver and baseband. The module stays in constant communication with the Android phone and sends a warning signal to the phone as soon as the ultrasonic sensor detects an obstacle (Fig. 5).

<u>Android Application</u>: The Android application stays in constant communication with the Bluetooth device. Upon detection of an obstacle, the message is communicated through the Bluetooth. A warning message saying "obstacle detected" can be heard over the earphone. The app captures an image and converts it into input stream upon the click of the button of the earphone. The input stream is sent to the API discussed in Sect. 2.5 which returns a string in plain English. The string is again played on the earphone. A service runs in the background detecting for triple taps of the earphone's button. An SOS message is immediately sent to the person's loved ones if the triple tap service is triggered.
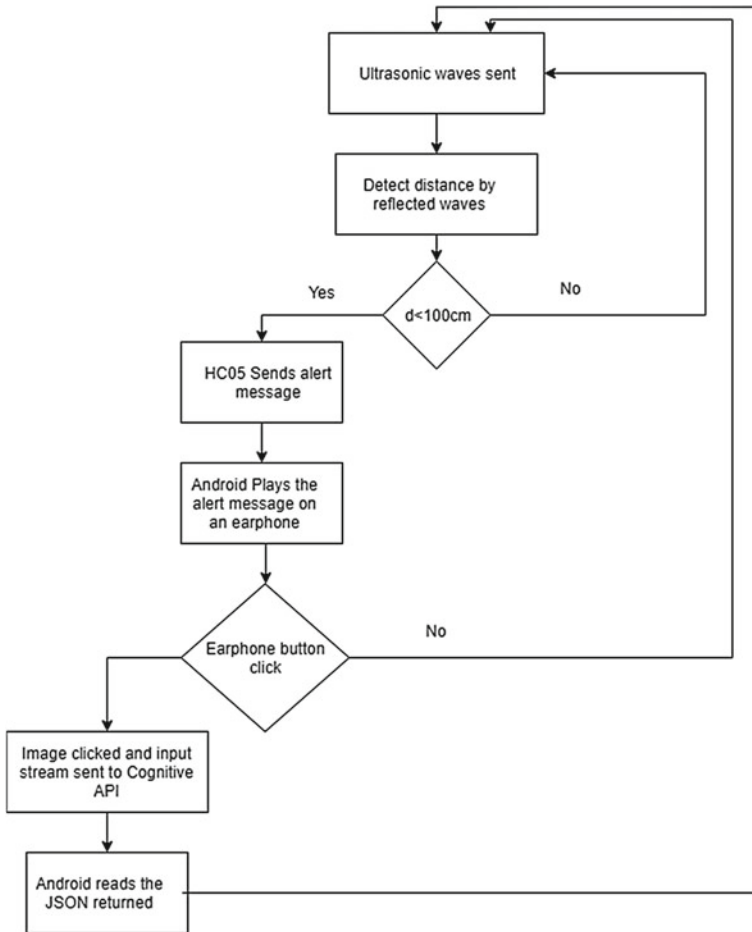
Fig. 3 Logical diagram of the obstacle detection

2.5   APIs and Database

Object Detection API: Microsoft Vision API [6] takes an input stream from the Android device and returns the description and tags of the image recognized. The blind person can click on his earphone's button to trigger the camera and take a photo. The app then sends the bitmap to the API, which then returns to the description of the image (Figs. 6, 7).

Navigation API: Indoor Atlas [7] uses radio signals, geomagnetic field, inertial sensor data, barometric pressure, camera data, and other details from the embedded sensors of the smartphone to provide precise data of the location of the user that the

Fig. 4 HC-05 Bluetooth module
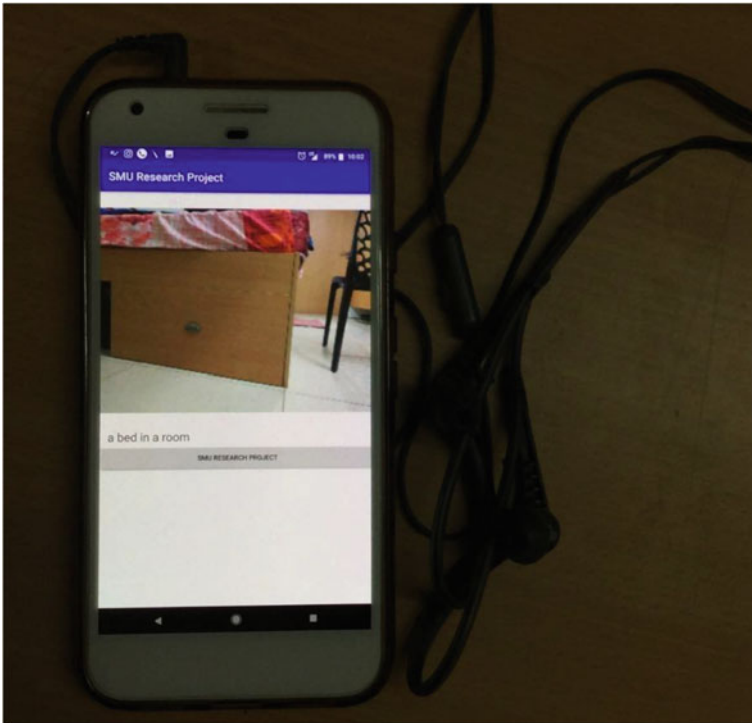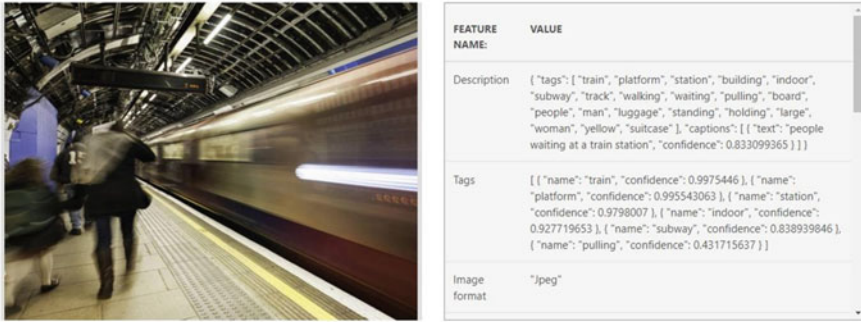

Fig. 5 ANDROID device running the app

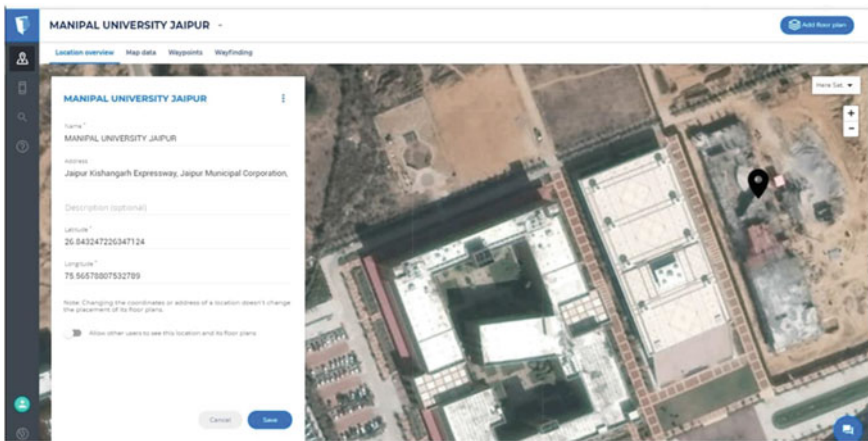**Fig. 6** Microsoft vision API [6]

Fig. 7 Screenshot of indoor atlas console [7]

normal positioning system fails to do. The API uses the floor plan of the area and needs to be trained with providing waypoints for the detection to start working. It provides a high ability to scale the system with a much greater accuracy as compared to tradition GPS with no hardware cost (Fig. 8).

Database: DeepStreamHub [8] is a real-time database providing 3 ms response time. The app stays in constant communication with DeepStreamHub to update real-time coordinates to the loved ones of the blind person. Anyone who is added to the trusted list of the blind person can view the location and receive SOS messages.
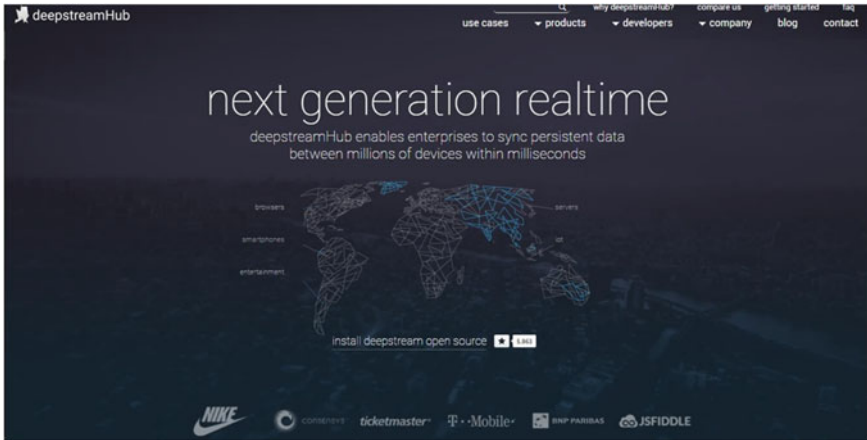
Fig. 8 Screenshot of DeepStreamHub [8]

3 Results

The ultrasonic sensors are attached to both the left and right feet of the blind person above 25 cm of the ground. The ultrasonic on both the left and right feet reduces the chances of collision from obstacles when one of the feet is behind the other. Also, attaching the ultrasonic sensors to the feet gives the blind person's hand mobility to move freely while traveling. The Arduino, Buzzer, and the Bluetooth sensor are kept in a sling bag whose straps are covered with reflective tape. The reflective tapes promote visibly for any car or person far away from the blind person, thus reducing the chances of the cars or other persons with the blind person (Fig. 9).

The blind person starts to walk while the ultrasonic sensor looks out for obstacles. The setup can successfully detect chairs, trees etc. above the ground level. A buzzer is triggered along with an output voice message that an obstacle has been detected by the ultrasonic sensor. The blind person can then take out his phone and with the click of his earphone's button click a picture which returns the description of the image.

To navigate through the app, the location needs to be fed to the android application. The application uses a pretrained blueprint to increase its efficiency in location prediction. The location is predicted with an accuracy of 1 m. The blind person then puts the phone in its pocket and listens to the navigation commands sent by the API, while the ultrasonic sensor looks for the obstacles in his way. And thus, the blind person is able to successfully move and reach his required destination.
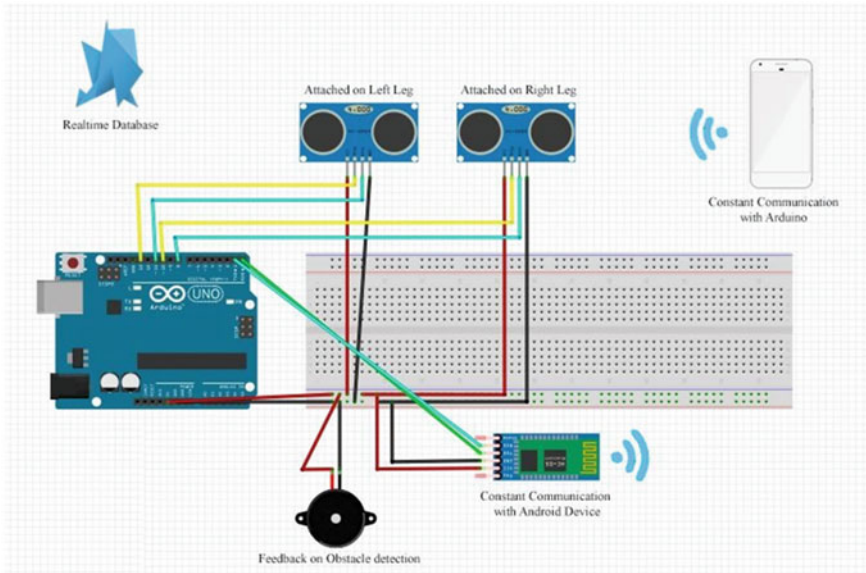
Fig. 9 Connections

4   Conclusion and Future Work

The suggested method showed a great improvement over the previous methods. It helped the blind go from one location to the other without the use of external probes like a stick. A separate application for accessing the GPS location of the person was implemented which showed the location of the person indoor or outdoor with an accuracy of 1 m. The results for indoor and navigation within a small area showed little to no errors and helped the visually impaired a lot. Further, this idea can be worked on various other platforms like iOS and Windows. Also, the location setting can be extended to work through voice commands.

References

1. Sinha K (2007) India has largest blind population Oct 11, 2007 (https://timesofindia.indiatimes.com/india/India-has-largest-blind-population/articleshow/2447603.cms)
2. Korial AE, Abdullah MN (2016) Novel method using Beacon and smart phone for visually impaired/blind people. Proc Int J Comput Appl (0975–8887) 137(1)
3. Shoval S, Borenstein J, Koren Y (1994) Mobile robot obstacle avoidance in a computerized travel aid for the blind. In: Proceedings of the IEEE international conference on robotics and automation, May 1994

4. Innet S, Ritnoom N (2008) An application of infrared sensors for electronic white stick. In: 2008 international symposium on intelligent signal processing and communication systems, Bangkok, Thailand
5. Saxena R, Jain M, Jaidka S (2018) Building a smart and energy efficient world: an era of internet of things. The stances of e-Government: policies, processes and technologies, p 143
6. Microsoft Azure for Research Overview (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/microsoft-azure-for-research-overview.pdf)
7. Magnetic Positioning the Arrival of 'Indoor GPS' (https://www.indooratlas.com/wp-content/uploads/2016/03/magnetic_positioning_opus_jun2014.pdf)
8. Getting started with Android (https://deepstreamhub.com/tutorials/getting-started/android/)

Implementation of Lightweight Crypto Processor Using Logistic Map for Wireless Sensor Network

Monjul Saikia and Md. Anwar Hussain

Abstract Use of a suitable cryptographic algorithm for wireless sensor networks is important due to the limitations of energy, computation capability and storage resources of the sensor nodes. Among two basic types of cryptographic techniques, namely asymmetric and symmetric key cryptography, symmetric cryptography technique is considered to be efficient over other in terms of computation cost. In symmetric key encryption, the secret key is known prior to encryption and decryption process. Therefore, in a wireless sensor network where keys are pre-distributed with Key Pre-distribution Schemes (KPS), can be used by both sides, i.e. sender and receiver nodes. In this paper, we will discuss symmetric key encryption techniques that can be efficiently used in wireless sensor network and explain the design possibilities and computation cost of using symmetric key cryptographic method. We have discussed the design of a crypto processor using a well-known chaotic map called Logistic map. Also, we have performed some experiments on the crypto processor and the results were stated.

Keywords Symmetric key encryption · Wireless sensor nodes ·
Encryption standards · Block cipher · Stream cipher · Security ·
Key pre-distribution scheme

M. Saikia (✉)
Department of CSE, North Eastern Regional Institute of Science and Technology,
Nirjuli, Arunachal Pradesh, India
e-mail: msk@nerist.ac.in
URL: https://www.nerist.ac.in

M. A. Hussain
Department of ECE, North Eastern Regional Institute of Science and Technology,
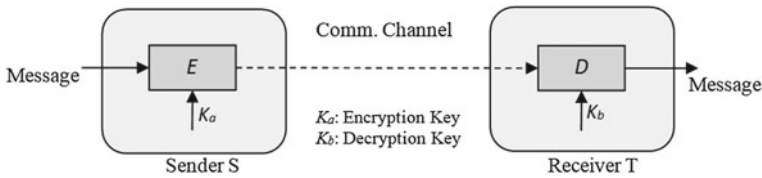Nirjuli, Arunachal Pradesh, India
e-mail: ah@nerist.ac.in

Fig. 1 Symmetric key encryption setting

1 Introduction

Because of the limitations of node function, it is preferable to use symmetric key encryption technology [1, 2] over the use of public key encryption technology [3]. A series of security mechanisms for wireless sensor network is studied under security protocol IEEE802.15.4, based on Advanced Encryption Standard (AES) algorithm [1]. The Security Protocols for Sensor Network (SPINS) is established based on the symmetric key system, and a more practical security scheme for sensor networks in the security system is put forward. Simply symmetric key ciphers are used to communicate certain secret data among two parties over communication channel. The both parties need to agree on a same key called symmetric key (say K_a and K_b; $K_a = K_b$) and uses an encryption algorithm E at the sender side and similarly a decryption algorithm D at destination end. The process can be shown diagrammatically as in shown Fig. 1.

In general, the communication channel is assumed to public and therefore, there is always chance of eavesdropper who tries to sneak into the communication channel being his objective to understand the messages. Eavesdropper knows the algorithm E and D, but not K_a or K_b at the same time, which made it difficult for him to decrypt the actual message being sent. The symmetric key encryption is categorized into two types block ciphers and stream ciphers. In block ciphers, a block of data is encrypted at a time whereas in stream cipher process 1 bit at a time. In wireless sensor network application point of view stream cipher is considered to be more useful where a sensor node detects an event in real time and the same needs to be forwarded to Base station. Essentially, block cipher will be mostly applicable when there is a need of deal with tons of data. Due to advancement in processing power block cipher is used widely, where n-bit block of plaintext is processed by the encryption algorithm to produce a ciphertext block and same n-bit block is decrypted by the decryption algorithm to get back the plaintext. Block cipher may be of two types, namely Transposition cipher and Substitution cipher. The transposition cipher is rearrangement of the bits using a transposition function, whereas substitution cipher substitute number of bits with different set of bits.

Although cryptography is said to be a method for secure communication, still there is a possibility of attack. Typically, the objective of attacking an encryption system is to recover the key in use rather than simply to recover the plaintext of a single ciphertext. There may be cryptanalytic attacks where it tries to gather knowledge of

the algorithm from sample plaintext and ciphertext and attempts to find out the key used. Another type of attack may be Brute-force attack where it tries every possible key to obtain plaintext. Therefore, it is important to estimate the amount of effort or time required to cryptanalysis successfully before using it in a specific application. As we have stated earlier that wireless sensor nodes are less capable of complex computation and therefore light version of cryptographic method is to be adopted. Stream cipher takes one byte of plaintext as input to the encryption algorithm and produces a ciphertext. It uses pseudorandom number at each step of encryption. The algorithm is simple but yet robust against attack. In rest of the paper, we discuss about stream cipher in details and its applicability option in wireless sensor network environment.

2 Stream Cipher

Typical stream cipher encrypts plaintext one byte at a time; although a stream cipher may be designed to operate on one bit at a time or on units larger than a byte at a time. In stream cipher, a key is input to a pseudorandom bit generator [4] that produces a stream of 8-bit numbers that are apparently random. The output of the generator, called a keystream, is combined one byte at a time with the plaintext stream using the bit-wise exclusive-OR (XOR) operation. For example, if the next byte generated by the generator is 01101100 and the next plaintext byte is 11001100, then the resulting ciphertext byte is $11001100 \oplus 01101100 = 10100000$ (ciphertext bitstream). Decryption requires the use of the same key bitstream sequence $10100000 \oplus 01101100 = 11001100$ (plaintext bitstream).

The stream cipher is similar to the one-time pad but the difference is that a one-time pad uses a genuine random number stream, whereas a stream cipher uses a pseudorandom number stream [5]. Figure 2 is a representative diagram of stream cipher structure.

The basic idea is to generate a key stream $z = z_1 z_2 \ldots$ and use these bitstream to encrypt a plaintext string $x = x_1 x_2 \ldots$ according to the rule in Eq. 1.
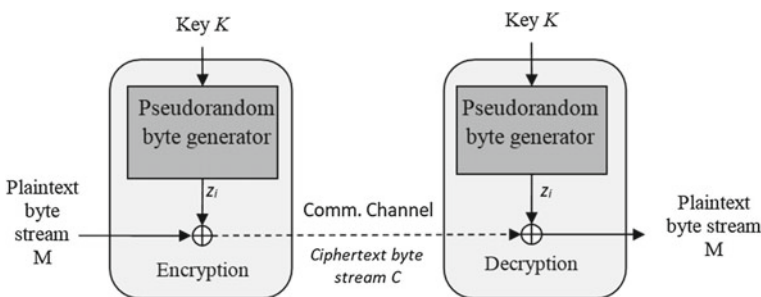

Fig. 2 Stream cipher setup

$$y = y_1 y_2 \ldots = e(z_1)(x_1)e(z_2)(x_2)\ldots \tag{1}$$

Definition 1 A synchronous stream cipher is a tuple (P, C, K, L, E, D) together with a function g, such that the following conditions are satisfied.

- P is a finite set of possible plaintext.
- C is finite set of possible ciphertext.
- K, the keyspace, is a finite set of possible keys
- L is a finite set called keystream alphabet
- g is the keystream generator. g takes a key form set K as input and generates a infinite string $z_1 z_2 \ldots$ called the keystream, where $z_i \in L$ for all $i \geq 1$.
- For each $z \in L$, there is an encryption rule $e_z \in E$ and a corresponding decryption rule $d_z \in D$. $e_z : P \rightarrow C$ and $d_z : C \rightarrow P$ are functions such that $d_z(e_z(x)) = x$ for every plaintext element $x \in P$.

As the pseudorandom byte generator is the key function in the stream cipher, we will discuss the possibility of design of random number generator which is applicable in wireless sensor network in the next section.

3 Principles of Random Number Generations

Random bitstream generation is an important cryptographic function. These random bits streams are used in both key generation and encryption process. While in generation and assignment of keys to sensor nodes, we used a specific key pre-distribution scheme [6, 7] to preload the keys to a sensor node from a key pool of keys generated using random number generator. In case of encryption we can design a pseudorandom bitstream generator that produce bitstream which is used for encryption in latter phase.

In essence, there are two fundamentally different strategies for generating random bits or random numbers. One strategy, which until recently dominated in cryptographic applications, computes bits deterministically using an algorithm. This class of random bit generators is known as pseudorandom number generators (PRNGs) or deterministic random bit generators (DRBGs). The other strategy is to produce bits nondeterministically using some physical source that produces some sort of random output. This latter class of random bit generators is known as true random number generators (TRNGs) or nondeterministic random bit generators (NRBGs).

Definition 2 Let k, l be positive integers such that $l \geq k + 1$. A (k, l)-bit generator is a function $f : \mathbb{Z}_2^k \rightarrow \mathbb{Z}_2^l$ that can be computed in polynomial time (as a function of k). The input $s_0 \in \mathbb{Z}_2^k$ is called the *seed*, and the output $f(s_0) \in \mathbb{Z}_2^l$ is called generated bitstream. It will always be required that l is a polynomial function of k.

The function f is deterministic; therefore the bitstream $f(s)$ is dependent only on the seed.

3.1 Logistic Map as Sequence Generator

The logistic map is a type of recurrence relation which is a polynomial mapping of degree 2 and chaotic behaviour can arise from very simple nonlinear dynamical Eq. 1 [8]. In the year 1976 biologist Robert May [9] discussed the logistic equation as Simple mathematical models with very complicated dynamics.

Mathematically, the logistic map is written as

$$x_{n+1} = rx_n(1 - x_n) \tag{2}$$

where x_n is a number in the between [0, 1] and the parameter r are those in the interval [0, 4]. Algorithm for Logistic Map sequence generator is given in Algorithm 1.

Algorithm 1 LogisticMap()


1: $x_0 \leftarrow seed$ $\triangleright\ 0 < seed < 1$
2: $r \leftarrow val$ $\triangleright\ 0 < val \le 4$
3: **for** $i = 0 \rightarrow \infty$ **do**
4: $x_{n+1} \leftarrow rx_n(1 - x_n)$
5: **end for**

3.2 Chaotic Behaviour Analysis of the Logistic Map

Chaotic behaviour does not exist for all values of r. It can be seen from the following experiments by setting different values of r. At first we set values of $r = 2.8$ and we consider initial seed as $x_0 = 0.35$. It is seen that it doesn't give any randomness after some iteration Fig. 3a. Increasing the value of r to $r = 3.0$ and setting same initial seed as $x_0 = 0.35$ just oscillates the sequence between two values as shown in Fig. 3b.
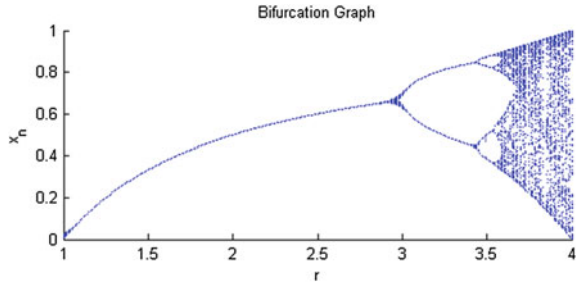

(a) $x_0 = 0.35$ and $r = 2.8$ (b) $x_0 = 0.35$ and $r = 3.2$
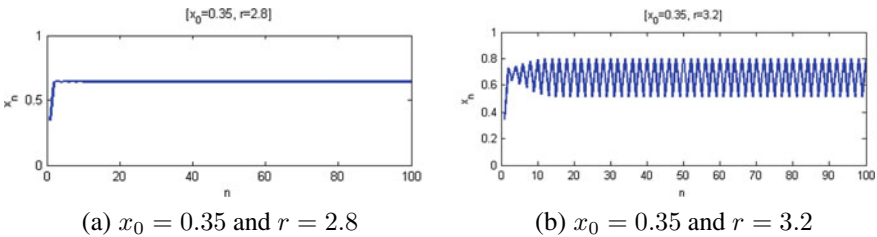
Fig. 3 Chaotic behaviour analysis: plot of first 100 points of the orbit

Fig. 4 Plot of first 50 points of the orbit [$x_0 = 0.35$ (blue), $x_0 = 0.70$ (red) and $r = 3.5$]
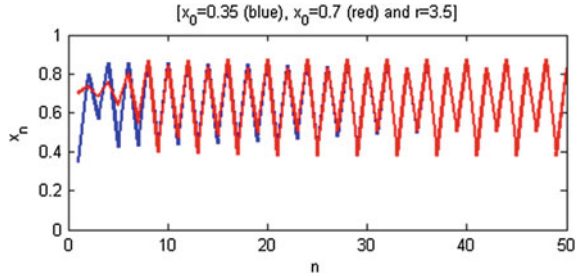
Fig. 5 The bifurcation graph for logistic map

In the next experiment we take two cases, *case*1 with the initial value of $x_0 = 0.35$ and *case*2 with initial value $x_0 = 0.7$ and set the value of $r = 3.5$. It is seen from Fig. 4 that both converge rapidly to a stable period of orbit 4.

It can also be shown that convergence occurs for any initial condition in the interval $(0, 1)$ while we consider value of $r = 3.5$.

The general behaviour of the logistic map depends critically on value of parameter r, as we have already seen in the previous examples. If we produce graphic that captures the change in behaviour as a function of r and $r \in [0, 4]$, we get a graph called *bifurcation graph* as shown in Fig. 5.

From various experiments, it is seen that when parameter r in the range of $3.5699 < r \leq 4$, the numbers generated in successive iterations of the mapping become chaotic in nature.

Therefore we consider two nearly identical $x_0 = 0.35$ and $x_0 = 0.3501$ as initial seed conditions with the parameter $r = 4$. This shows the chaotic behaviour, which is often thought of as 'sensitive dependence on initial conditions'. In this scenario, even though the orbits are nearly identical at the start, after 100 points or so, there's no way to detect, either statistically or by looking at Fig. 6, any such correlation between the two orbits.

Fig. 6 Plot of two nearly identical seed [$x_0 = 0.35$ and $x_0 = 0.3501$]
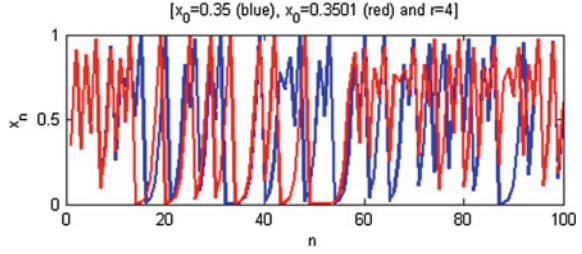
# 4 Hardware Implementation of Logistic Sequence Generator

As random number generator is the basics for stream cipher, an efficient processor that can compute the bitstream is often desirable. Considering the limitations of a sensor node in wireless sensor network, design of lightweight hardware architecture [10, 11] for generation of random sequence is an important issue. The function $f(x)$ for Logistic map can be implemented in two type of architecture by taking the basic equation in the form of the following equations.

$$x_{n+1} = 4x_n(1 - x_n) \tag{3}$$

$$x_{n+1} = 4x_n - 4x_n x_n \tag{4}$$

Design makes simple while we set $r = 4$. With minimal requirement of multiplier, adder, shift registers and subtractor blocks a Logistic Map sequence generator module can be designed as shown in Fig. 7a, b using the said equations. The initial seed is feedback by module Z^{-1} to the module Z^0, where it is subtracted from constant 1. Then, multiplication with $r = 4$ is done by *2-bit* shift register, the resultant value then passed to multiplier to get $4x(1 - x)$. Then, MUX selector selects this output to the output register, which is the generated Logistic sequence.

In the second design two 2-bit shift registers are used. The first shift register produces $4x$ by taking value of x from the feedback and second shift register finds $4xx$ from the output xx produced by the multiplier.

(a) Design with equation (3)   (b) Design with equation (4)
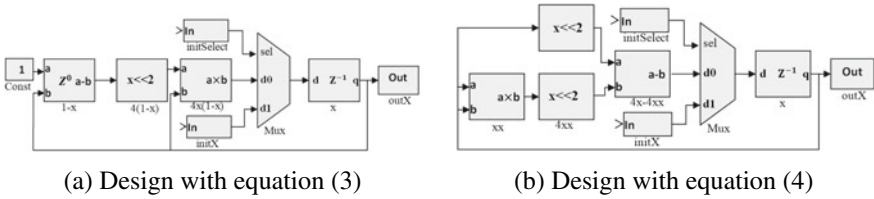
Fig. 7 Simulink block design for logistic sequence generator

The requirement of hardware for designing a Logistic bitstream depends of number of precession considered.

5 Model for Stream Cipher Using Logistic Sequence Generator

The basic model for stream cipher in wireless sensor network is depicted in Fig. 8. The key pre-distribution scheme discussed in [7], assigns a set of keys $K_i = \{k_1, k_2, \ldots\}$ to the sensor nodes i, which were generated using a Logistic sequence generator. Sensor nodes having a shared key may undergo message exchange using that key. The key is passed to Logistic sequence generator, which produce random byte stream then and performs the encryption over plaintext. Similarly, the receiver sensor node uses the same key for its key chain to decrypt the message. If it requires further forwarding of the message, then the same strategy is being used, which is shown in Fig. 9. Figure 9 shows process of message transmission in case of multiple hop
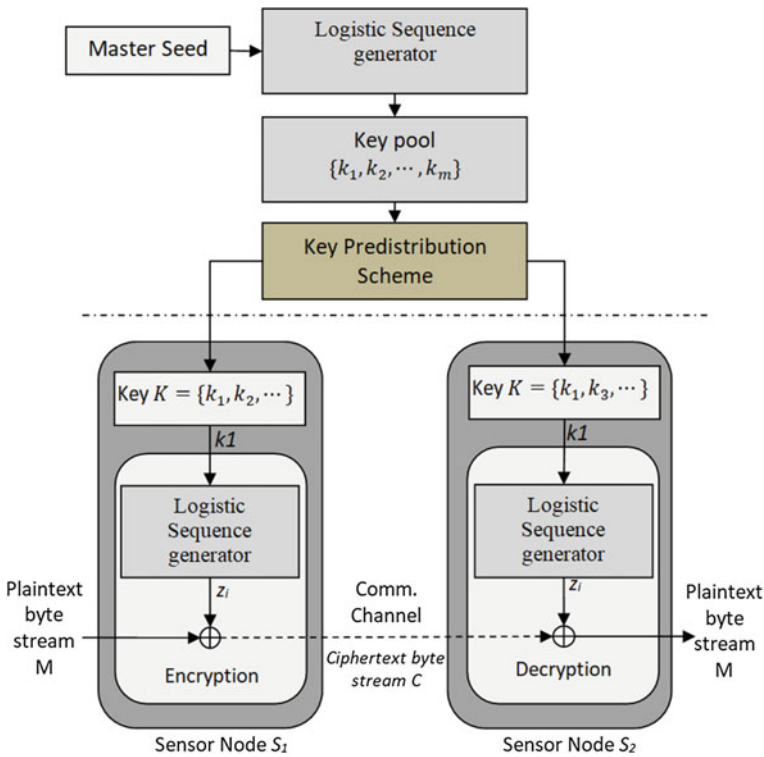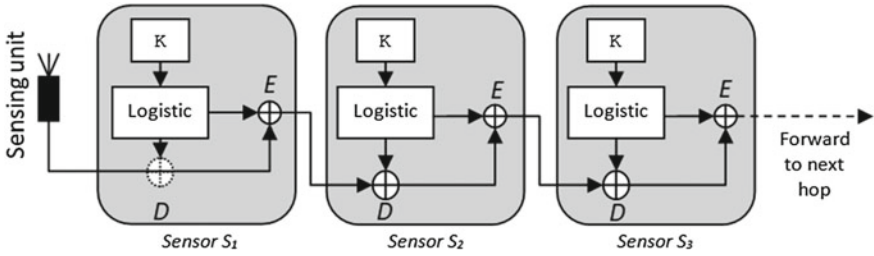

Fig. 8 Stream cipher setup for WSN

Fig. 9 Secure message forwarding in WSN multi-hop path

paths. Sensor S_1 encrypts the sensing data using shared key of sensor S_2, sensor S_2 then decrypts the data and encrypts the same with the shared key of sensor S_3 and so on.

6 Experimental Results

For our experiments, we take simple 8-*bit* logistic sequence generator. We take eight registers to store the 8 successive sequence produced by the generator. 8-*bit* precision is the initial seed to the logistic generator. The generator produces 8-*bit* sequences which are stored in 8 registers successively. Bitstreams are fetched from each significant bit of the registers. Figure 10a shows an example of the method used, where 16 numbers of bitstream outputs were shown by taking $x_0 = [1011010]$.

Encryption process is simply the XOR with the sequence of plaintext bitstream, which is as shown clearly in Fig. 10b. For the experiment, we take 8-bit stream
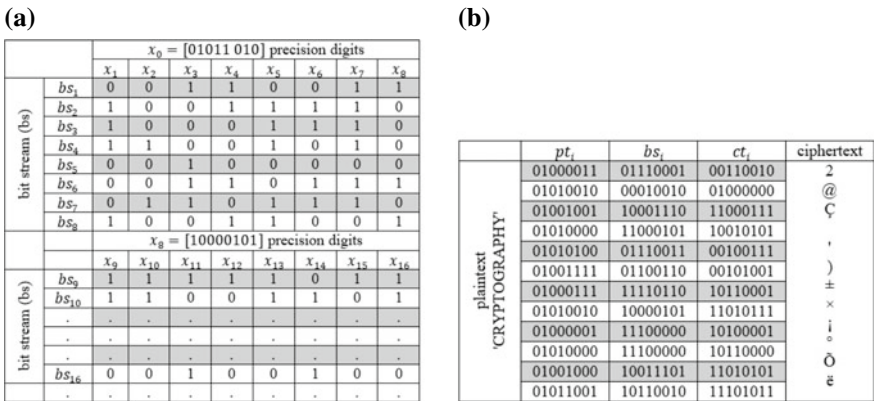
(a)

$x_0 = [01011\ 010]$ precision digits

		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
	bs_1	0	0	1	1	0	0	1	1
	bs_2	1	0	0	1	1	1	1	0
	bs_3	1	0	0	0	1	1	1	0
	bs_4	1	1	0	0	1	0	1	0
	bs_5	0	0	1	0	0	0	0	0
	bs_6	0	0	1	1	0	1	1	1
	bs_7	0	1	1	0	1	1	1	0
	bs_8	1	0	0	1	1	0	0	1

$x_9 = [10000101]$ precision digits

		x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
	bs_9	1	1	1	1	1	0	1	1
	bs_{10}	1	1	0	0	1	1	0	1
	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
	bs_{16}	0	0	1	0	0	1	0	0
	.	.	.	.	.	.	.	.	.

bit stream (bs) / bit stream (bs)

(b)

plaintext 'CRYPTOGRAPHY'

pt_i	bs_i	ct_i	ciphertext
01000011	01110001	00110010	2
01010010	00010010	01000000	@
01001001	10001110	11000111	Ç
01010000	11000101	10010101	•
01010100	01110011	00100111	'
01001111	01100110	00101001	)
01000111	11110110	10110001	±
01010010	10000101	11010111	×
01000001	11100000	10100001	¡
01010000	11100000	10110000	°
01001000	10011101	11010101	Õ
01011001	10110010	11101011	ë

Fig. 10 **a** A bitstream generation process, **b** encryption of plaintext '*CRYPTOGRAPHY*' using 8-bit sequence generator

Fig. 11 Overlapping plot of plaintext and cipher text bitstream
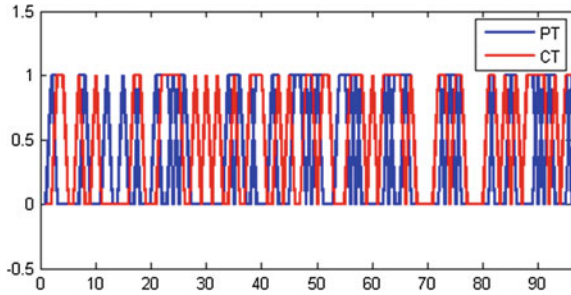
generator and plaintext '*CRYPTOGRAPHY*' is converted to ASCII character, then converted to binary 8-bit and perform *bit-wise XOR* operation to get ciphertext.

Correlation between the plaintext and ciphertext is found to be 0.0161. A overlapping plot with plaintext (Blue line) and ciphertext (Red line) is shown in Fig. 11. Similarly, decryption can be done in the similar manner by the receiver sensor.

7   Conclusion

In this paper, we have discussed requirement of symmetric cipher as a security tool for wireless sensor network. We have discussed importance of lightweight crypto processor for a sensor node. Symmetric cipher is a simple but yet effective technique of encryption. Implementation of symmetric cipher requires random bitstream generator, which is discussed here. Hardware implementation of Logistic Map based sequence generator is discussed with its minimum requirement of logic units in two possible ways. Wireless sensor network uses key pre-distribution scheme prior to deployment of sensor node into target field, the same logistic sequence generator can be used to generate a key pool, from where keys are preloaded to the sensor nodes. In case of multi-hop path, the same crypto processor can be used for encryption and decryption processes, until data reached at its final destination.

References

1. Delfs H, Knebl H (2007) Symmetric-key encryption. Introduction to cryptography. Springer, Berlin, pp 11–31
2. Agrawal M, Mishra P (2012) A comparative survey on symmetric key encryption techniques. Int J Comput Sci Eng 4(5):877
3. Wander, AS et al (2005) Energy analysis of public-key cryptography for wireless sensor networks. In: Third IEEE international conference on pervasive computing and communications, 2005. PerCom 2005. IEEE
4. Zulfikar Z (2014) FPGA implementations of uniform random number based on residue method. Trans J Rekayasa Elektrika 11(1)

5. Lehmer DH (1954) Random number generation on the BRL high speed computing machines, by M. L. Juncosa. Math Rev 15:559

6. Eschenauer L, Gligor VD (2002) A key-management scheme for distributed sensor networks. In: Proceedings of the 9th ACM conference on computer and communications security. ACM

7. Boeing G (2016) Visual analysis of nonlinear dynamical systems: chaos, fractals, self-similarity and the limits of prediction. Systems 4(4):37. https://doi.org/10.3390/systems4040037

8. May RM (1976) Simple mathematical models with very complicated dynamics. Nature 261(5560):459467. https://doi.org/10.1038/261459a0

9. Dabal P, Pelka R (2011) A chaos-based pseudo-random bit generator implemented in FPGA device. In: 2011 IEEE 14th international symposium on design and diagnostics of electronic circuits & systems (DDECS). IEEE

10. Blum L, Blum M, Shub M (1986) A simple unpredictable pseudo-random number generator. SIAM J Comput, no. 2

11. Pathan A-SK, Lee H-W, Hong CS (2006) Security in wireless sensor networks: issues and challenges. In: The 8th international conference on advanced communication technology, 2006. ICACT 2006, vol 2. IEEE

12. Zulfikar Z (2012) Novel area optimization in FPGA implementation using efficient VHDL code. Trans J Rekayasa Elektrika 10(2) (2012)

13. Walters JP et al (2007) Wireless sensor network security: a survey. Secur Distrib Grid Mob Pervasive Comput 1:367

14. Hoang T (2012) An efficient FPGA implementation of the advanced encryption standard algorithm. In: 2012 IEEE RIVF international conference on computing and communication technologies, research, innovation, and vision for the future (RIVF). IEEE

15. Intel Corp (2012) Intel digital random number generator (DRNG) software implementation guide, 7 Aug 2012

16. Ma D, Tsudik G (2010) Security and privacy in emerging wireless networks. IEEE Wirel Commun

17. Park S, Miller K (1988) Random number generators: good ones are hard to find. Commun ACM

18. Chan H, Perrig A, Song D (2003) Random Key predistribution schemes for sensor networks. In: Proceedings of the 2003 IEEE symposium on security and privacy. IEEE

19. May RM (1976) Simple mathematical models with very complicated dynamics. Nature 261:459

20. Tanaka H, Sato S, Nakajima K (2000) Integrated circuits of map chaos generators. Analog Integr Circuits Signal Process 25(3):329–335

Application of IoT-Enabled Smart Agriculture in Vertical Farming

Sutanni Bhowmick, Bikram Biswas, Mandira Biswas, Anup Dey, Subhashis Roy and Subir Kumar Sarkar

Abstract Vertical farming is an unconventional farming technique that has gained relevance in recent years, as existing agricultural lands fail to meet the needs of the growing population. Smart monitoring of the ambient parameters in vertical farming can improve the productivity and quality of the crops. A system has been proposed to develop sensor arrays that can measure the ambient parameters and upload the data onto the ThingSpeak Cloud, using the Intel Edison wireless module. The web-based application can be used to analyze and monitor the light, temperature, humidity, and soil moisture of the vertical farming stacks. Using the Virtuino app, a SMS can be sent if the parameters fall below a threshold value.

Keywords Vertical farming · Intel Edison · Sensor arrays · ThingSpeak cloud

S. Bhowmick (✉) · B. Biswas · M. Biswas · A. Dey · S. Roy · S. K. Sarkar
Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata 700032, West Bengal, India
e-mail: sutannibhow94@gmail.com

B. Biswas
e-mail: bikram94biswas@gmail.com

M. Biswas
e-mail: mbiiswas@gmail.com

A. Dey
e-mail: anupetce@gmail.com

S. Roy
e-mail: subhashisaec@gmail.com

S. K. Sarkar
e-mail: sksarkar@etce.jdvu.ac.in

1   Introduction

Increase in the worldwide population combined with the rise in per capita consumption of modern man has led to increased demand for food. To meet the growing demands and to feed the masses, more agricultural land is required. This has led to increased pressure on existing farmlands (which are being harvested multiple times in a year) as well as necessitated the acquisition of more land for agriculture. But an increased understanding of the adverse effects of deforestation has led man to think of alternative ways to optimize land use in agriculture. One such solution is vertical farming. This technique involves growing crops in a vertically stacked manner in an indoor controlled environment. Vertical farming is environment-friendly and pesticide free. But this technique is highly critical as it requires constant monitoring of the ambient parameters. Vertical farming is usually practiced in large warehouses which can be several storey's tall. The sunlight is replaced by LED lights. The controlling parameters for vertical farming are mainly ambient light, ambient temperature, soil moisture, and humidity. Several automation techniques have been previously suggested for optimizing resource use, such as accommodation of actuation by automated irrigation systems [1, 2], but wireless monitoring using IoT in the field of vertical agriculture has not been previously suggested.

The ambient parameters must be monitored for each stack along the vertical columns. This requires a set of sensors for each vertical stack. The large amounts of data generated by these sensors need an efficient data management system. Using the Internet of Things (IoT), we can efficiently manage the sensor data, identify redundant sensor readings, and visualize it with the help of web-based applications [3]. Thus, a convenient IoT-based system may be developed for vertical farming. The concept of vertical farming has been shown in Fig. 1 where the levels represent the vertical farming stacks.

Internet of Things (IoT) is an emerging concept and a paradigm that seeks to connect a variety of things or objects that are able to interact and coordinate with each other with the help of wireless and wired connections [4, 5]. Aided by the internet and the advances in cloud computing technology, IoT has extensive scope of application in multiple areas such as transport, health care, security, communication, infotainment, and smart cities. From a functional point of view, the IoT needs a set of basic components or modules to facilitate the connection of "things":

- A module for interaction with local IoT devices (smart sensors). This module acquires observations from sensors and forwards it to remote servers for analysis and storage.
- A module for interaction with remote IoT devices.
- Module for application specific data analysis and processing.
- User interface module (web or mobile).

From the data flow point of view, the IoT comprises of the following:

(a)   End Devices—These are the sensors or actuators which directly interact with the surrounding environment to acquire some information (sensors) or per-
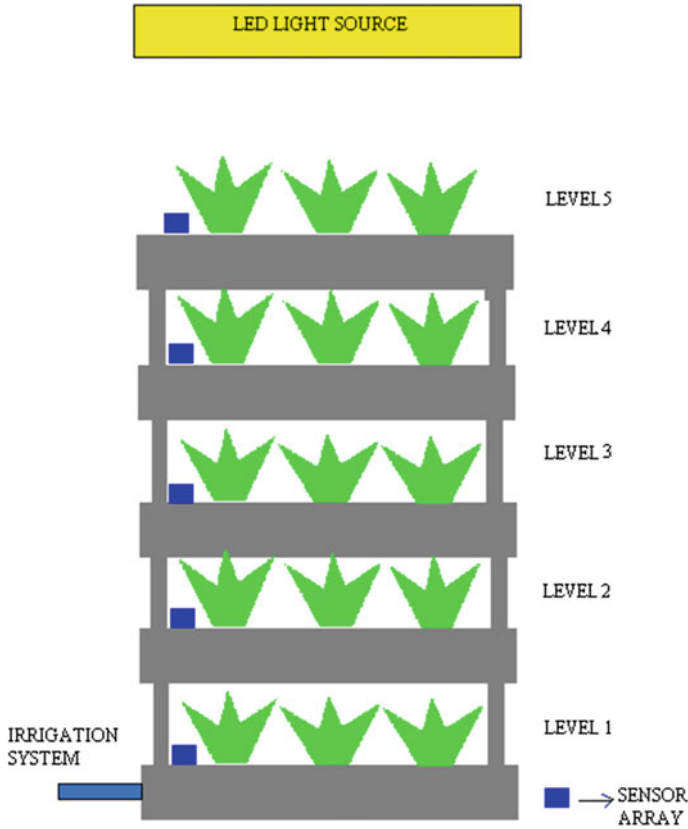
Fig. 1 Diagrammatic representation of vertical farming aided by ambient sensor arrays

form some changes on the surroundings (actuators). These devices must be lightweight and power efficient. They are not OS based, and are often referred to as dumb devices as they have no decision-making capabilities.

(b) Propagation Node—The propagation node gathers and stores data. These are based on the operating system. The nodes check the efficacy of the incoming sensor data and prunes redundant data. Data from the propagation node is periodically bundled and sent to the cloud via ipV6 or ipV4.

(c) Filter Gateway and Machine Learning—Once the data is uploaded through the internet onto the cloud, it is sent for analysis and storage.

The proposed system comprises of (1) the set of sensors that facilitate the monitoring of the ambient conditions; namely, ambient light sensor, soil moisture sensor, humidity sensor, and ambient temperature sensor, (2) a data transmission pathway through which the sensor data can be communicated wirelessly using a microcontroller and a wireless module, and (3) a web-based application to visualize the sensor data.
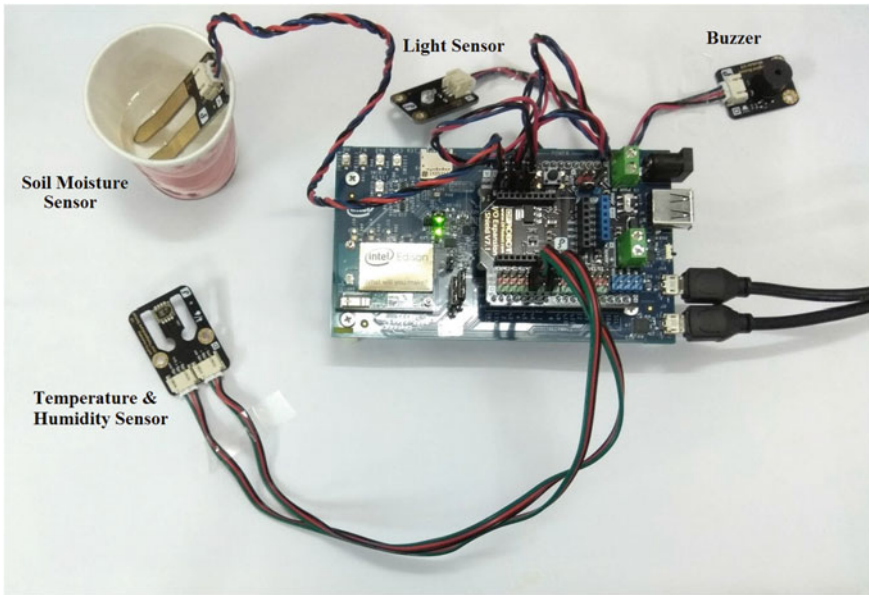
Fig. 2 Prototype of the proposed system

2 System Overview

See Fig. 2.

3 System Design

Vertical farming is mostly practiced indoors and hence, there is no sunlight. The sunlight is simulated with the use of LED lights that enable the plants to carry out photosynthesis. Thus, monitoring the ambient light is necessary to provide optimum photonic energy for proper photosynthesis. Hence, our system uses an ambient light sensor. Temperature and moisture are also critical for the health of the plants in an indoor environment [6]. This creates the need for using temperature and humidity sensors. Finally, monitoring the soil moisture constantly and generating an alarm if the moisture level falls below a set threshold, is required. For this, we have used soil moisture sensor [7]. The four sensors together form a sensor array. The data from the sensors must be constantly monitored in real time in order to initiate necessary actions that will ensure the health of the growing crops. The proposed system uses Intel Edison wireless module that is mounted on an Arduino prototype. The Intel Edison is a computer-on-module that has found extensive use as a development system for wearable devices and IoT-based applications. Computer-on-Modules (COMs) are

complete embedded computers built on a single circuit board. The board runs on Linux OS with developmental support for Arduino IDE, Eclipse, and Intel XDK. In our system, the board has been programmed to receive the readings from the aforementioned sensors, generate an alarm when soil moisture falls below optimum threshold, and upload the sensor readings onto the cloud. The program is pushed through the serial port of the Arduino prototype and is available in the application section. The backend service calls the .exe of this application. Since, it is Linux based, the PuTTY tool is used to configure the Edison Board with the Wireless Local Area Network (WLAN) to facilitate communication with the Internet.

To facilitate the monitoring of the sensor data remotely, a web-based user interface has been made with the help of ThingSpeak. ThingSpeak is an open-source IoT application that facilitates the storage and retrieval of data from "things" using the HTTP protocol over a local area network or the internet. A channel was created on ThingSpeak platform and was named "Farm". The API Keys of the channel were used on the programming end so that the sensor data could be viewed in the channel. Four fields were created for the purpose of visualizing the data from the four sensors. Hence, with the help of ThingSpeak, remote monitoring of sensor value is possible. For the mobile-based monitoring interface, Virtuino was used. Virtuino is an Android app which shows the sensor data graphically and also allows sending of SMS when the sensor reading crosses a threshold (which can be programmed).

The components of the sensor array are as follows:

- Ambient Light Sensor V2 (SKU: DFR0026)—this analog sensor is used to measure the light intensity. The illumination range varies from 1 to 6000 lx.
- SHT1x Humidity and Temperature sensor (SKU: DFR0066)—the humidity sensor is pre-calibrated in a precision humidity chamber. The sensor output is in %humidity or relative humidity. The temperature sensor provides the room temperature value in degree Celsius.
- Soil Moisture sensor (SKU: SEN0114)—the soil moisture has two probes to pass current through the soil. The resistance value obtained reflects the moisture content of the soil. Presence of water improves the soil conductivity and vice versa.

4 Results and Discussion

Figure 3 shows the graphical reading of the ambient temperature sensor in degree Celsius. Figure 4 shows the Virtuino graphs for temperature and humidity sensors. The Virtuino can be programmed to send SMS when the temperature and humidity values fall above or below predetermined thresholds. The threshold values depend upon several factors like the type of crop grown, etc., Figures 5 and 6 show the Virtuino and the ThingSpeak graphs for the soil moisture sensor, respectively. The soil moisture sensor is calibrated in terms of moisture levels. If the soil moisture falls below a certain predefined level, the buzzer rings an alarm that can notify the farmers to switch on the irrigation system for the concerned vertical stack.
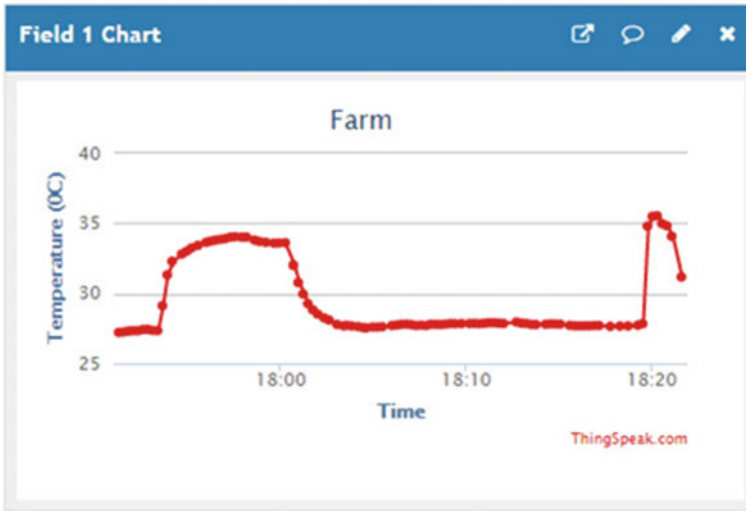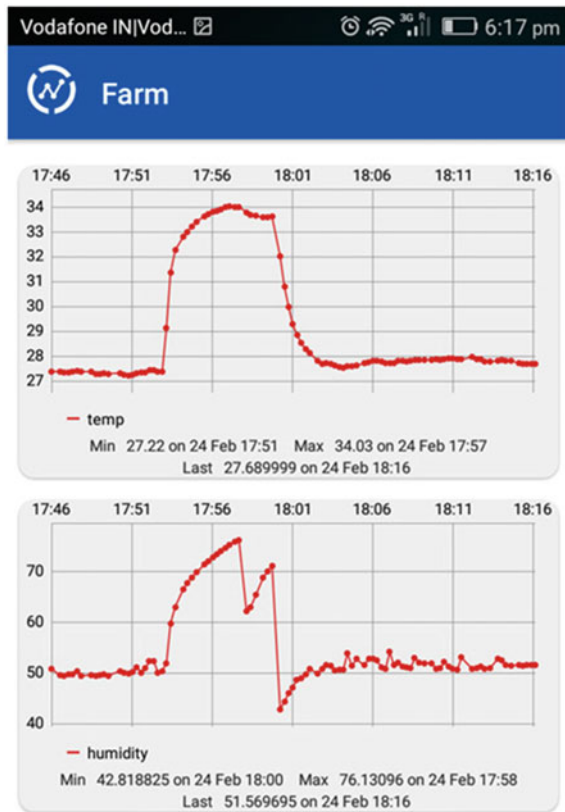
Fig. 3 Results for the temperature sensor obtained from ThingSpeak

Fig. 4 Results for the
temperature and humidity
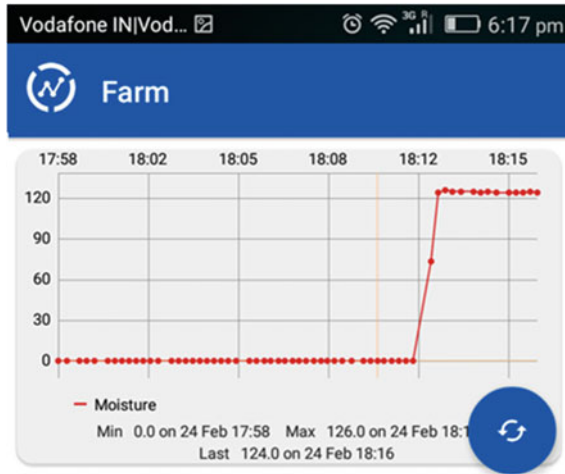sensor obtained from
Virtuino

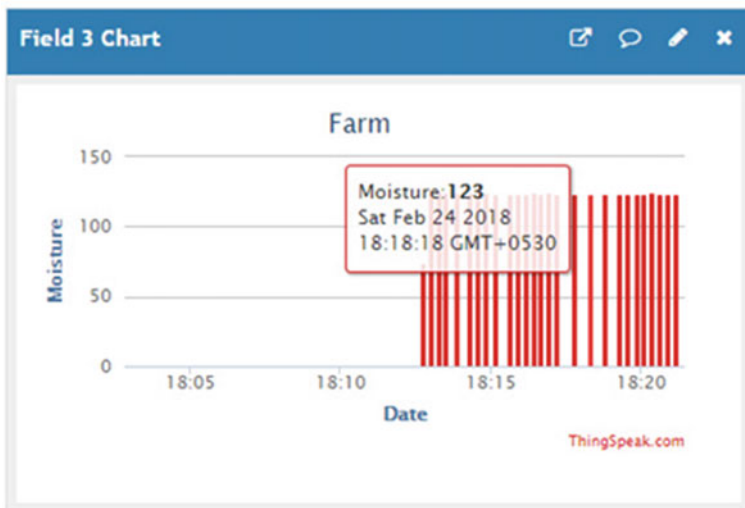Fig. 5 Results from the soil moisture sensor obtained from Virtuino

Fig. 6 Results from the soil moisture sensor obtained from ThingSpeak

5 Conclusion

The proposed system when implemented on the vertical farming warehouses, can allow users to achieve proper monitoring and control of the ambient parameters. As the demand for agricultural land increases, and the trend shifts toward unconventional farming methods, such as vertical farming, the proposed system can be extensively used to achieve smart agriculture. Future work includes integration of the system with

RFID technology such that each vertical stack can be uniquely identified. The system can be further improved by accommodating actuation such as automated irrigation system and temperature control. Incorporation of pH sensors and gas sensors into the existing system can further address the criticality of the vertical farming technique.

References

1. Kotamaki N, Thessler S, Koskiaho J, Hannukkala AO, Huitu H, Huttula T, Havento J, Jarvenpaa M (2009) Wireless in-situ sensor network for agriculture and water monitoring on a river basin scale in Southern Finland: evaluation from a data users perspective. Sensors 4:2862–2883
2. Gutiérrez J, Villa-Medina JF, Nieto-Garibay G, Porta-Gándara MA (2014) Automated irrigation system using a wireless sensor network and GPRS module. IEEE Trans Instrum Meas 63:166–176
3. Gondchawar N, Kawitkar RS (2016) IoT based smart agriculture. Int J Adv Res Comput Commun Eng 5:12446–12450
4. Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 15:2787–2805
5. Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Context aware computing for the internet of things: a survey. Commun Surv Tutor IEEE 1:414–454
6. Liu H, Meng Z, Cui S (2007) A wireless sensor network prototype for environmental monitoring in greenhouses. In: International conference on wireless communications, networking and mobile computing, pp 21–25
7. Wang Q, Terzis A, Szalay A (2010) A novel soil measuring wireless sensor network. IEEE Trans Instrum Meas, 412–415

UVM-Based Verification IP of AMBA AXI Protocol Showing Multiple Transactions and Data Interconnect

Shivani Malhotra and Neelam Rup Prakash

Abstract Many advanced protocols are emerging and are being used widely in the industry these days. One such protocol is the AMBA Advanced Extensible Interface Protocol by ARM processors. This protocol provides significant attributes which did not exist in the previous Advanced High-Performance Protocol. This paper presents practical results of some of those attributes of this protocol. Now, it should be noted that verification of such protocols is a tedious task and consumes a lot of time. Hence, a Verification IP should be created which acts as a catalyst in the verification process. A Verification IP already contains the necessary mechanisms for testbench generation and can be easily integrated to other tools.

Keywords Advanced microcontroller bus architecture (AMBA) · Advanced RISC machine (ARM) · Advanced extensible interface (AXI) · Advanced high-performance bus (AHB) · Verification intellectual property (VIP) · Universal verification methodology (UVM)

1 Introduction to AMBA

The Advanced Microcontroller Bus Architecture was developed by ARM Ltd. in 1996 and is a registered trademark of ARM Ltd. It is actually a family of protocol guidelines which is open to the public and sets out a policy for the interlink and control the practically working structures that constitute a SoC. It also helps to design the embedded processors who possess one or more CPUs and several peripherals. It

S. Malhotra (✉) · N. R. Prakash
VLSI Design, PEC University of Technology, Sector 12, Chandigarh, India
e-mail: shivani.malhotra610@gmail.com

N. R. Prakash
e-mail: neelamrprakash@pec.ac.in

S. Malhotra · N. R. Prakash
Electronics and Communication Department, PEC University of Technology, Sector 12, Chandigarh, India

529

defines a common backbone for Soc blocks that accolades a multipurpose planning approach [1].

Several buses are described in the AMBA specifications which are based on different generations:

- Advanced High-performance Bus (AHB)
- Advanced System Bus (ASB)
- Advanced Peripheral Bus (APB)
- Advanced Extensible Interface (AXI)
- Advanced Trace Bus (ATB)

The abovementioned protocols are, in fact, a predefined standard for embedded processors as they are registered skillfully and as such, no authorities are entailed to utilize them. These specifications have been obtained to fulfill the following significant conditions:

- It ensures that the growth of embedded microcontroller products is made smooth and easy with just one or sometimes more than one signal processors.
- Other necessity is to become technology maverick and make sure that multipurpose outlying libraries get drifted across a wide ranging of IC processes and be apt for various design implementation technologies [1].
- Also to promote customizable system structure to outdo the freedom of processor, issuing a progress enchiridion for improved cached CPU fundamentals plus the growth of outlying libraries.
- Another important condition is to reduce the necessary silicon substructure in order to carry out on-chip and off-chip transmission efficiently for both operation and manufacturing test.

2 Introduction to AXI Protocol

The latest generation of AMBA that is AXI protocol targets high performance and is used in high-frequency system designs. This latest generation protocol has an edge over previous generations like APB, AHB as it enables high-frequency operation without using complex bridges. And if backtracking is done, it would be compatible with the existing AHB and APB protocols. This VIP [2] can be used to verify the AXI bus in any kind of project and supports many simulators like Aldec Riviera PRO, Cadence Incisive Enterprise Simulator, Vivado Simulator, Mentor Graphics Questa Prime, and Synopsys VCS (Fig. 1).
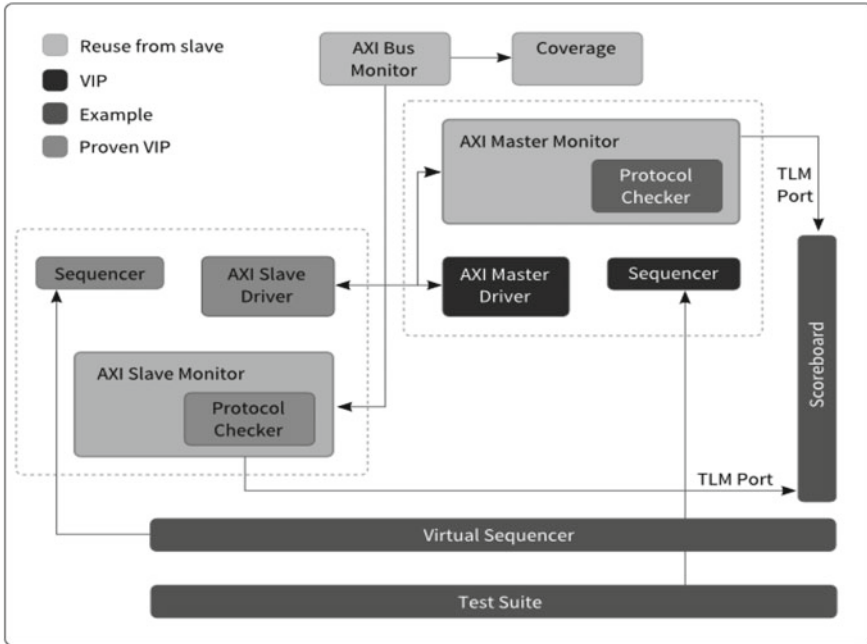
Fig. 1 AXI verification IP

3 Key Features of AXI Protocol

3.1 Individual Address/Control and Data Phases

The main features of AXI protocol is the presence of independent and discrete address and data buses for transmission of data between the master and slave. Several masters and slaves can be present in AXI.

3.2 Unaligned Transfers

For the bursts which are formed of data transfers exceeding one-byte boundary, it is not necessary that the initial bytes which have to be ingressed needs to be aligned with the built-in width threshold. It happens when the value of the address (AxADDR) is not a multiple of the size of the data being transferred (AxSIZE) [3].

3.3   Multiple Outstanding Addresses

AXI protocol utilizes transaction ID tags for delivering manifold outstanding addresses. There are no restrictions on order of data packets sent during the read and write transactions and can be completed in any order. Hence in the AXI protocol, we can complete the out-of-order transactions by permitting every port to work as multiple ordered ports. These features enhance the performance of interconnects and increase the efficiency of the system [4]. We can order only those transactions which have a valid ID and there is no restriction even if they have different IDs. There are five transaction IDs present in AXI which are:

AWID—This ID tag is used for the write address group of signals.
WID—This can be defined as the write ID tag for a write transaction.
BID—This ID tag is used for the write response.
ARID—This is the ID tag for the read address group of signals.
RID—This can be defined as the read ID tag for a read transaction.

3.4   Out-of-Order Transaction Completion

This protocol supports the accomplishment of transactions which are out of order. In this, an ID tag is provided to every transaction across the interface. It is necessary for the protocol to complete the transactions with the same ID tag in order but this does not apply for transactions with different ID tags, they can be completed in any order.

4   Analysis and Results

In this paper, simulation results of two basic yet most important attributes which clearly help this protocol to stand aside from previous generation protocols have been explained. These results have been obtained from DVE Synopsys tool. So, given below are those features along with their VPD as a result of DVE Synopsys which is basically a waveform viewer by Synopsys.

4.1   Separate Address/Control and Data Phases

We know that the identity of information that will be conveyed is explained by the transaction on address channel containing address and control data [2]. The interchange of data between master and slave is performed with help of a write data channel to the slave or a read data channel to the master. In write transactions,
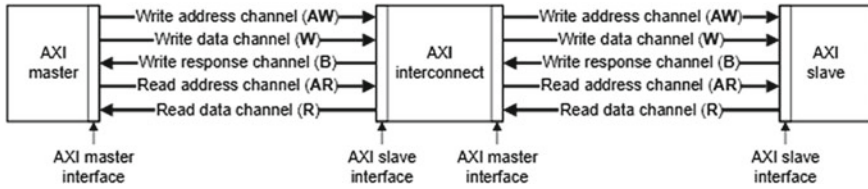
Fig. 2 AXI master slave interface

in which whole data get transferred from master to slave, a supplementary write response channel is there in AXI protocol that allows a write transaction completion signaling from slave to master.

All the five individual channels contain a set of data signals and utilize a two-way VALID and READY handshake process (Fig. 2).

Read/write address channels [5]

All the read and write transactions possess their individual address channels. The necessary address and control data is transferred by a suitable address channel.

Read data channel

The data pair comprising read data and its response is transmitted by the read channel from slave to the master. This channel comprises of:

- A data bus, which can be of the width ranging from 8 to 1024 bits.
- A read response signaling the fulfillment report of the read transaction.

Write data channel

The write data is transmitted from master to slave through this channel. It involves a data bus, which can be of the width ranging from 8 to 1024 bits.

Write response channel

This channel issues various procedures for the slave acknowledge the write transactions. All write transactions use accomplishment indications. These indications are in form of signals which occurs for single time only for individual bursts.

Following are the simulation results referring to individual address/control and data buses for read and write transactions (Figs. 3, 4, 5 and 6).

Fig. 3 Simulation result showing the read address/control signals and Arid value of 3 which is of 4 bits

Fig. 4 Simulation result showing the read data of 64 bits with the same Arid value

Fig. 5 Simulation result showing write data of 64 bits along with the strobe signals

Fig. 6 Simulation result showing write data, strobe signals and the write response with Arid of 3 and 4 in write address channel

4.2 Out-of-Order Completion

The AXI Protocol supports the completion of transactions which are not in order. In this, the protocol provides an ID tag to every transaction present across the interface. This ID tag plays a great role in deciding the order of transactions. It is the requirement of this protocol that the transaction with same ID tags need to be accomplished in same sequence, whereas the ones with different ID tags can be accomplished out of order [6].

The inclusion of this feature has improved the system performance to a great extent in two ways:

(1) The transactions with fast-responding slaves can be enabled by the interconnect to move forward of the prior transactions having slow going slaves.
(2) Read data which is out of order can be returned by complex slaves. For example, a data item which has to be accessed later might be available before the data item which is to be accessed earlier [1].

Following are the simulation results showing the out-of-order completion attribute of AXI Protocol (Figs. 7 and 8).
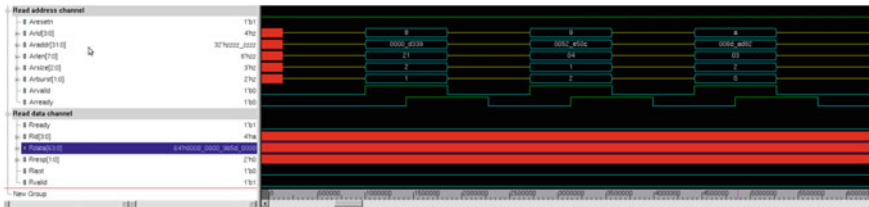
Fig. 7 Simulation result showing read data and Arid of decimal values 8, 9, and 10
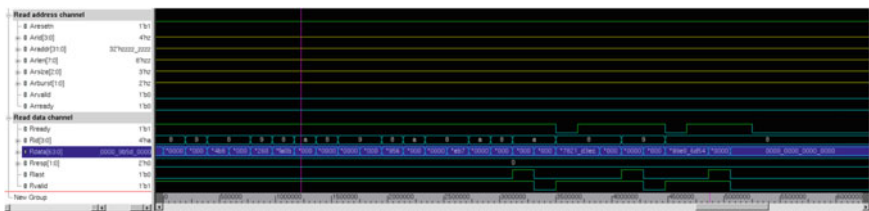
Fig. 8 Simulation result showing out-of-order completion

5 Conclusion

The verification IP has been successfully designed and verified by using DVE Synopsys tool. Test cases concerning Multiple Transactions and Data Interconnect were run and the results were obtained. In the future, more work can be carried out based on some other key features of AXI protocol such as unaligned transfers, atomic accesses, etc. Also, some efforts can be made in the field of AXI Bridge to add external connectivity.

References

1. ARM, AMBA AXI Protocol Specification [Online]. Available http://www.arm.comARM
2. Silberschatz A (1983) A case for non-two-phase locking protocols that ensure atomicity. IEEE Trans Softw Eng SE-9(4)
3. The AMBA SOC Platform. http://www.springer.com/in/book/9781461478638
4. https://www.xilinx.com/products/intellectual-property/axi_systemcache.html
5. Synopsys VCS https://www.synopsys.com/verification/simulation/vcs.html
6. Wang HW, Lai CS, Hwang SA, Lin Y-H (2008) On chip interconnection design and soc integration with open core protocol (OCP). In: Proceedings of the IEEE international symposium, VLSI Design 2008, pp 25–28
7. AMBA AHB Protocol Specification [Online]. Available http://www.arm.com
8. Chang C-Y, Lee K-J (2014) On deadlock problem of on-chip buses supporting out-of-order transactions. IEEE Trans Very Large Scale Integr Syst 22(3)
9. http://www.testandverification.com/solutions/vip/axi-vip/
10. Mahesh G, Sakthivel SM (2015) Verification of memory transactions in AXI protocol using system verilog approach. In: International conference on communication and signal processing (ICCSP) 2015, Melmaruvathur India

Prospect of Improving Internet of Things by Incorporating Software-Defined Network

Rohit Kumar Das, Arnab Kumar Maji and Goutam Saha

Abstract Internet of Things (IoT) is gaining a significant amount of importance in the field of networking. It uses low-cost devices to provide the application to the users. IoT devices are resource-constraint device, so they suffer from the constraints of availability, reliability, flexibility, and security. Lots of research attention has been given to minimize these challenges. By integrating the features of Software-Defined Network (SDN), most of the IoT issues can be overcome. This paper presents a detail investigation of IoT issues which can be improved by incorporating SDN. The resources used for SDN are equipped with huge computational power. This particular feature can help IoT devices to overcome the constraint problem. The virtualization concept of SDN will help to resolve the scalability issues of IoT. Similarly, many other issues can be resolved at a very low cost. Researchers are finding that SDN-based IoT is much powerful than that of the exiting IoT system.

Keywords Internet of things · Software-defined network · Integration · Challenges and issues

1 Introduction

Internet of Things (IoT) is an evolutionary paradigm shift in the world of networking where each and everything is getting connected to each other and also to the internet. The conversion of objects from traditional to smart is done by using the technologies of computing, embedded and sensor systems, Internet protocols, communication mechanisms, etc. [1].

R. K. Das (✉) · A. K. Maji · G. Saha
Department of Information Technology, North-Eastern Hill University, Shillong, India
e-mail: rohitdas.it.13@gmail.com

A. K. Maji
e-mail: arnab.maji@gmail.com

G. Saha
e-mail: dr.goutamsaha@gmail.com

537
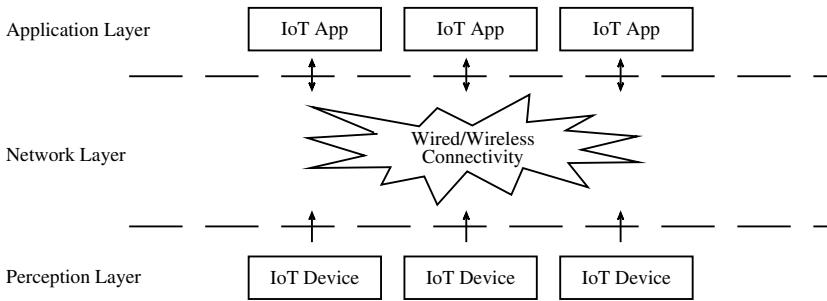
Fig. 1 Three-layer IoT architecture [1]

As shown in Fig. 1, the bottom Perception Layer, the middle Network Layer, and the upper Application Layer together form the basic IoT architecture. They support and communicate with each other to convey the expected administrations to the end clients. An IoT device consists of physical objects and sensor devices which are the basic components of Perception Layer. Network Layer is responsible for providing connectivity to the sensor device(s) from the bottom layer to the upper Application Layer. Data transmissions achieved by using technologies like RFID, Wi-Fi, Bluetooth Low Energy, Infrared, ZigBee, etc. The Application Layer provides the required services to the end users.

The Internet of Things can be viewed as a global network which allows human-to-things and things-to-things communication by providing a unique identity to each and every object [2].

Software-Defined Network (SDN) with high-level language and Application Programming Interface (API) can provide dynamic and automatic management of network devices, including services, topology, traffic paths, packet handling policies and Faults, Configuration, Accounting, Performance, and Security (FCAPS) control the network operator [3]. By incorporating the SDN concept in IoT, some critical issues of IoT like availability, flexibility, scalability, data management, security, and hardware dependencies can be handled relatively with lesser resources [4].

The SDN architecture as shown in Fig. 2, consists of Infrastructure Layer, Control Layer, and Application Layer. Infrastructure Layer also known as data plane which consists of network devices that forward the data based on the decisions made by the controller. Control Layer is regarded as the control plane of SDN where the controller performs overall functions of the network. South-Bound Application Interface (SBI) provides the communication interface between the data plane and control plane. Application Layer provides the required services to the end user by communicating with the controller through North-Bound Application Interface (NBI).

In contrast to the traditional network, the control plane and the data plane are separated in SDN structure. The control plane is placed in a central location, whereas the data plane is distributed over the network. The controller resides in the control plane which can control the overall network performance. The data plane is regarded
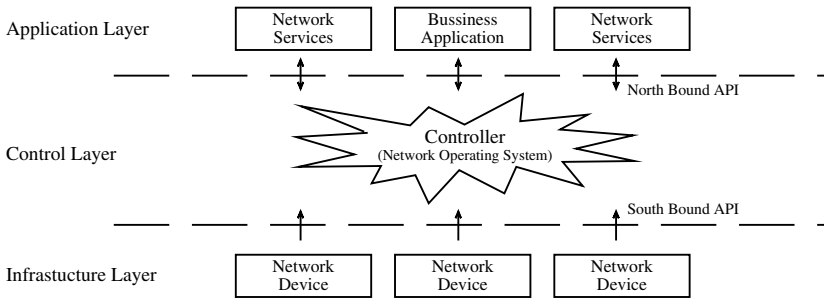
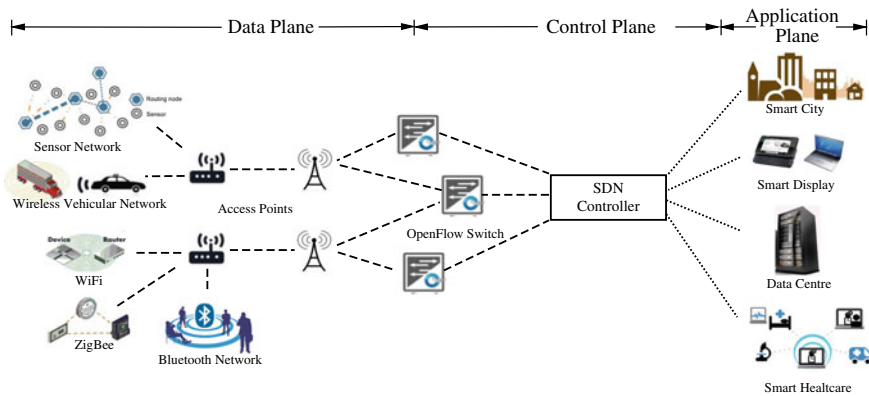**Fig. 2** Three-layer SDN architecture [5]

**Fig. 3** Incorporation of SDN controller in IoT network

as simple forwarding elements (switches), which direct the packet according to the rules provided by the controller [5].

OpenFlow allows SDN functions to be implemented both in hardware and software. The existing infrastructure with OpenFlow as a protocol, new concepts, and experiment can be carried out.

SDN with IoT can propose an organized administration structure, which can enhance the reconfigurable and adaptable feature of the system. The application scenario of SDN can be extended with IoT as shown in Fig. 3, where different IoT applications can be controlled by the SDN controller which gets the data from the data plane that passes through the OpenFlow switch to the SDN controller. SDN can present new conceivable outcomes for organizing administration and setup techniques [6].

The existing IoT system can be augmented easily with less hardware maneuvering. With SDN controller, abstractions can be created and modified. As IoT is still evolving, integration of SDN can help in network management in a more efficient manner. To incorporate SDN with IoT, the focus has to be made on the SDN con-

Fig. 4 SDN and IoT
integration
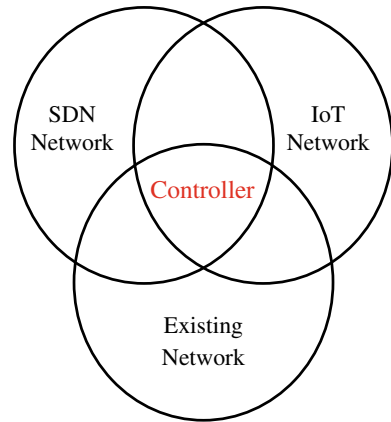
troller. This can be implemented by integration of SDN and IoT together in one control plane that will also look after the existing legacy network infrastructure as shown in Fig. 4.

2 IoT Key Challenges and Opportunities with SDN Solutions

There are several challenges for SDN-based IoT systems which are discussed in the next subsection.

Both SDN and IoT have their own limitations which make a more challenging job for the combination of both the systems. Some of the challenges of IoT available in the literature are listed below which we can be overcome by SDN-based IoT system:

   i. **Availability**: IoT should provide the anywhere anytime facility in terms of both hardware and software [2]. This can be well addressed by SDN, where the availability of the hardware will be monitored and controlled by the controller in the Control layer and application from the Application layer [7].

  ii. **Reliability**: Reliability refers to the proper working of the system based on its specification. Reliability is even more critical and has more stringent requirements when it comes to the field of emergency response applications. In order to have an efficient IoT, the underlying communication must be reliable [2]. SDN can improve this challenge of IoT by providing a distributed environment for controllers. Each controller will constantly supervise the lower perception layer devices and the upper application layer. In the case of emergency response applications, SDN controller can allocate all the underlying available channel bandwidth to the emergency dedicated line.

iii. **Interoperability**: IoT consolidates an arrangement of devices. Consequently, interoperability between different devices is a difficult issue [8]. Home system slicing [9] is a mechanism where different devices work together to provide the necessary action based on OpenFlow and Simple Network Management Protocol (SNMP) to control and configure the network elements. This approach can be utilized with Smart Home or Smart City, where different applications work together in a uniform way. Path Computation Element (PCE) can also help in gradual or partial migration to SDN [10].

iv. **Heterogeneity**: Factors that lead to heterogeneity problem are operating conditions, functionality, resolutions, hardware platforms, service pattern, implementations, and interaction modes [11]. This challenge of IoT can greatly reduce the overall performance of the network. Multi-network Information Architecture (MINA) [12] is an adaptable middleware with reflective self-observing that reallocates application flow and maintaining proactive plans approach to address this problem of heterogeneity for multi-network IoT environment by providing a centralized global view much like SDN for overlay structure and can perform analysis of the network state [13].

v. **Scalability**: Adding new devices, services, or functions to the network without affecting the existing network is one major challenge [14]. Scalability is not a major issue in SDN. Moving whole network control over the control plane of the network (SDN) makes it much simpler to handle the network. The controller of the network gets the network wide view which empower itself to control policies and applications [15].

vi. **Security**: IoT is made up of millions of devices where communication between them is done by exchange of millions of data in which providing security measure becomes important [10]. SDN can increase the security of the system. Only the controller of the system allows for the analysis and approval of connections and traffic flow. During the connection establishment phase between the controller and switches, the controller can first check for the tagger (authentic label) of the switch and according to its police accept or reject the particular switch [16].

vii. **Performance**: The performance of IoT is depended not only to the services it is providing but also to the underlying technologies been used to deliver those services. Enhancement in the field of protocols and technologies are needed to provide upgradation in the performance of IoT [2] which is a challenging task. SDN can enhance performance in controlling the IoT devices by adding multiple controllers [17].

3 Existing Literature

There have been few works done on the combined approach of IoT and SDN in recent years. Some of them are listed below:

Qin et al. [12] proposed a Multi-network Information Architecture (MINA), which is an IoT deployment framework for facilitating heterogeneous access network with the help of SDN controller that will translate services requirements into low-level network requirements and handle the traffic flows of the abiding certain QoS constraints.

Jararweh et al. [18] proposed a framework based on Software-Defined Internet of Things (SDIoT). They have presented a comprehensive software-defined framework for managing the IoT processes and applications. SDIoT integrates the network, store, and provide security from one control model.

Wu et al. [16] developed a framework for IoT multi-networks development based on SDN (UbiFlow) for monitoring ubiquitous flow control and mobility management in multi-networks. UbiFlow achieves distributed mode of operation by partitioning geographical locations and setting up a controller for each partition where switches from different partitions are partially connected.

Flauzac et al. [17], a cluster-based IoT for SDN scheme has been proposed by the authors where each cluster head is considered as a coordinator (controller) and each cluster represent an SDN domain. The cluster head is responsible for monitoring and controlling the traffic flow inside its domain only. The communication between two different clusters is done via cluster heads only which use inter-domain link. By consider clustering method, it also emphasis the distributive mode for SDN.

Zhang et al. [19] describe the implementation for software-defined network IoT controller which can execute IoT applications. The IoT controller can cause the IoT device to operate based on the functions provided by the IoT applications. The controller can load IoT applications program interface associated with the set of functions.

Han and Ren [20] have considered using OpenFlow protocol in Wireless Sensor Network. The authors have used the centralized concept of SDN in Wireless Sensor Network (WSN) by having a master node (controller) programmed by NOX as Network Operating System (NOS) and center nodes (OpenFlow switches).

De Oliveira et al. [21] have proposed a framework based on TinyOS for SDN in WSN, which enables for multiple controllers. It consists of SDN-enabled sensor nodes in the data plane and SDN controller node in the control plane. The SDN-enabled sensor nodes find an SDN controller for communication. The controller keeps the track of topology change, applications, apply policies, and manage the traffic flow.

4   Research Issues

After going through the literature, it is found that there are several research issues to provide interoperability among heterogeneous IoT devices that may crop up presently. They are as follows:

   i. SDN is a centralized-based approach where the controller is regarded as a central entity which controls the overall function of the network. The centralized architecture of SDN cannot be considered as a reliable and efficient one as it can prompt issues like a bottleneck, load, failure, security, etc. Research is been carried out to make control plane more flexible and distributive with more than one controller, which can likewise solve the scalability problem of the controller.

  ii. The most common protocol for IoT is 6LoWPAN and for SDN is OpenFlow. Most of the characteristic of these two protocols are the same. Research can be carried out for the development of hybrid protocol that can be used for IoT-based SDN network. There can be a possibility of incorporating the features of the SDN protocol with IoT protocol to form the hybrid protocol.

 iii. SDN utilize OpenFlow and IoT depends on IEEE 802.15.4 specialized standard. This is a prominent area where research can be carried out on how to use these technologies with the existing infrastructure.

 iv. IoT devices are required to be authenticated, the data transmitted by them must be securely presented, and the applications provided by IoT based on the data collected from the IoT devices must be validated so the integrity is maintained. This is a prominent area of research, where both IoT and SDN security mechanism can be further implemented on the system.

  v. The network should be able to provide connectivity to all the resource-constrained devices which are usually low powered. The current state-of-the-art for IoT devices is that it should transmit data up to the cloud. This usually consumes a huge amount of energy. Researchers are focusing on placing the computing power nearer to the IoT device as an edge network (Fog Computing).

5 Conclusion

In this paper, we have ventured to incorporate the SDN concept in IoT so that the existing difficulties of IoT like availability, scalability, reliability, interoperability, heterogeneity, hardware dependencies, and others can be handled relatively with lesser resources and in a more efficient manner. The existing IoT system can be augmented easily with less hardware maneuvering. This particular endeavor has not attained its maturity yet. This resulted in various open research issues which are discussed in this paper. The integration of SDN feature for IoT network opens more interesting research areas but can enhance the overall performance of the network. In this paper, we have tried to provide an overview of how SDN and IoT can be merged together so that it can provide better network functionalities.

References

1. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M (2014) Internet of things for smart cities. IEEE Internet Things J 1(1):22–32
2. Al-Fuqaha A, Guizani M, Mohammadi M, Aledhari M, Ayyash M (2015) Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun Surv Tutor 17(4):2347–2376
3. Kreutz D, Ramos FM, Verissimo PE, Rothenberg CE, Azodolmolky S, Uhlig S (2015) Software-defined networking: a comprehensive survey. Proc IEEE 103(1):14–76
4. Qin Z, Denker G, Giannelli C, Bellavista P, Venkatasubramanian N (2014) A software defined networking architecture for the internet-of-things. In: Network operations and management symposium (NOMS), 2014 IEEE, pp 1–9
5. Hu F, Hao Q, Bao K (2014) A survey on software-defined network and OpenFlow: from concept to implementation. IEEE Commun Surv Tutor 16(4):2181–2206
6. Haque IT, Abu-Ghazaleh N (2016) Wireless software defined networking: a survey and taxonomy. IEEE Commun Surv Tutor 18(4):2713–2737
7. Jarraya Y, Madi T, Debbabi M (2014) A survey and a layered taxonomy of software-defined networking. IEEE Commun Surv Tutor 16(4):1955–1980
8. Shah PAA, Habib M, Sajjad T, Umar M, Babar M (2016) Applications and challenges faced by internet of things—a survey. In: International conference on future intelligent vehicular technologies, pp 182–188
9. Yiakoumis Y, Yap K-K, Katti S, Parulkar G, McKeown N (2011) Slicing home networks. In: Proceedings of the 2nd ACM SIGCOMM workshop on home networks, pp 1–6
10. Sezer S, Scott-Hayward S, Chouhan PK, Fraser B, Lake D, Finnegan J, Viljoen N, Miller M, Rao N (2013) Are we ready for SDN? Implementation challenges for software-defined networks. IEEE Commun Mag 51(7):36–43
11. Hussain MI (2016) Internet of things: challenges and research opportunities. CSI Trans ICT, 1–9
12. Qin Z, Iannario L, Giannelli C, Bellavista P, Denker G, Venkatasubramanian N (2014) MINA: a reflective middleware for managing dynamic multinetwork environments. In: Network operations and management symposium (NOMS), 2014 IEEE, pp 1–4
13. Puthal D, Nepal S, Ranjan R, Chen J (2015) DPBSV—an efficient and secure scheme for big sensing data stream. Trustcom/BigDataSE/ISPA, 2015 IEEE, vol 1, pp 246–253
14. Kim H, Feamster N (2013) Improving network management with software defined networking. IEEE Commun Mag 51(2):114–119
15. Yeganeh SH, Tootoonchian A, Ganjali Y (2013) On scalability of software defined networking. IEEE Commun Mag 51(2):136–141
16. Wu D, Arkhipov DI, Asmare E, Qin Z, McCann JA (2015) UbiFlow: mobility management in urban-scale software defined IoT. In: 2015 IEEE conference on computer communications (INFOCOM), pp 208–216
17. Flauzac O, Gonzalez C, Nolot F (2016) Developing a distributed software defined networking testbed for IoT. Proc Comput Sci 83:680–684
18. Jararweh Y, Al-Ayyoub M, Benkhelifa E, Vouk M, Rindos A et al (2015) SDIoT: a software defined based internet of things framework. J Ambient Intell Humaniz Comput 6(4):453–461
19. Zhang Z, Banks JA, Leung LA (2015) Software-defined IoT controller. US Patent App. 14/688,852, 16 Apr 2015
20. Han Z, Ren W (2014) A novel wireless sensor networks structure based on the SDN. Int J Distrib Sens Netw 10(3):1–7
21. De Oliveira BT, Gabriel LB, Margi CB (2015) TinySDN: enabling multiple controllers for software-defined wireless sensor networks. IEEE Lat Am Trans 13(11):3690–3696

Smart Healthcare Systems Using Cloud Computing Environments

Jayashree Agarkhed, Ramegowda Ashalatha and Siddarama R. Patil

Abstract The involvement of Information and Communication Technology (ICT) for the adoption of smart cities in India is in high demand today. The healthcare service solution uses cloud computing based future technology which uses the coming Internet of things technology to access the patient's details and their confidential healthcare information. The rapid circulation of mobile cloud computing model is associated with massive amounts of energy spent by the cloud data centers. So enhancing the energy efficiency of those data centers, has to they turn out to be a leading challenge in the healthcare industry. Mobile cloud computing provides services by bringing the plentiful resources of medical data in cloud computing to the nearness of mobile devices to authorize the performance of the application and preserve the battery life.

Keywords Mobile computing · Offloading · Smart city · Health care

1 Introduction

The concept of cloud computing is playing a salient role in the medical industry. Due to heavy workload, the pharmaceutical sector often struggles with active movement. Patients' access to their report from all places is quite complicated because of network traffic. The healthcare industry requires fast and effective access to medi-

J. Agarkhed · R. Ashalatha (✉)
Department of C.S.E, Poojya Doddappa Appa College of Engineering, Kalaburagi,
Karnataka, India
e-mail: ashalatha.dsce@gmail.com

J. Agarkhed
e-mail: jayashreeptl@yahoo.com

S. R. Patil
Department of E.C.E, Poojya Doddappa Appa College of Engineering, Kalaburagi,
Karnataka, India
e-mail: pdapatil@gmail.com

cal resources access such as patient medical reports and scan copies from the cloud server. To solve the problem, the proposed model is effective since the proxy server is accessing the stochastic process. For example, suppose scenario is that, a patient is accessing the resources from the particular location. These resources have been placed in another place whose position has distance. The proposed solution suits perfectly in this type of situation. The proxy server is installed near a typical location of the patient. If the patient tries to access the resources of their own from the mobile devices, then the resources are located on the proxy server rather than the cloud server. It makes the easy way to get rid of network traffic and makes offloading effective. The stochastic process makes the network bandwidth very effective one [1]. Mobile cloud provides a robust and ubiquitous resource pool for mobile devices. In the meantime, the Central Processing Unit (CPU) in mobile systems remains the most power-hungry component which consumes battery lifetime consistently and steadily. A power estimation model provides vision regarding the power breakage concerned with hardware components. The experimental result demonstrates that although energy consumption changes drastically based on the workload, the monitor and the CPU are the two most significant power consuming components. Various wireless mobile network technologies and their cooperation are undergoing rapid development and deployment. The ultimate goal is to facilitate mobile handsets to efficiently run applications that usually run on the more powerful desktop or laptop Personal Computers (PCs) while allowing users to enjoy the freedom of mobility [2]. However, due to their miniaturization, even the most potent mobile handsets today are unable to compete against their desktop siblings concerning resources, especially battery life and storage or network capacity. Mainly because of working, entertaining, and living environments, the resources are abundant and often mostly idle on desktop PCs. This article will investigate how an offloading process provides service and for resource-constrained mobile handsets in cloud systems [3]. One of the challenging techniques adopted in mobile cloud computing is the offloading process. This technique allows common computational offloading to different devices that include mobile devices, cloud, and proxy server. A universal and flexible architecture maximizes the computational gain concerning variant metrics such as minimizing the response time, dropping the energy consumption, and growing the lifetime in the network. The novel design is used to systematize computation offloading of various medical resources on the interrupted network connections because of the nature of mobile devices. This technique assists the pharmaceutical industry to avoid the slowness in patients to access their details in mobile from the cloud.

In the existing work, the goal of the virtual machine is right-sizing the cloud virtual data centers which can be made possible through consolidating the multiple workloads into a small number of servers. The dynamic nature of data centers on Virtual Machines (VMs) arrives at the system from time to time and releases the resources when finishing their tasks. Because of its dynamic nature, queuing theory and stochastic process are especially useful in this scenario. It allows us to understand the behavior of a complex system and to guide the intention of resource management policies in cloud data centers. But each time, the user will access the cloud, and some drawback may occur such as bandwidth that will be increased based on the number of

users to access the cloud. The communication cost is raised and then the web traffic may happen at this time. Therefore a technique as called offloading can be used [4]. The offloading process means the arrangement of enormous resources like memory, CPU, and communication cost which includes bandwidth as the primary resource and aims to preserve these precious smart devices resources as much as possible. The objective is to cover the reduced response time for a user application running on resource-constrained intelligent devices. Less resource consumption usually leads to more energy savings in smart machines. Here, the cloud proxy directly accesses the cloud securely. The cloud proxy shares the resources with the data users based on offloading. The remaining structure of the paper organized as follows. The following section presents the literature survey. Section 3 gives the system architecture. Section 4 presents the result analysis part. Section 5 presents the conclusion part.

2 Related Work

The personal health records of patients are required to store in the form of domains. Depending upon the level of the necessary security, these medical records need to be secure in the private or public area [5]. The mobile health monitoring system has to preserve the cloud to present high-quality service [6]. The cloud system has been secured using a hybrid system based on smart card technology. The users are authenticated to increase the confidentiality in healthcare system using cloud technology. The medical and confidential data can be made secure by making use of near-field communication [7]. Scaled data can be processed using mapReduce technique on the cloud network. The health monitoring made practiced using IoT technology. It also helps in the remote monitoring system of the patients in the hospital [8]. Health monitoring scheme makes use of the sensors for high security. It uses the cryptographic measures in the cloud to overcome various types of attacks [9]. Mobile healthcare systems need to secure the privacy of the user's data. They provide end-to-end encryption facility for patient's confidential data over the private clouds [10]. The health monitoring system's privacy is made secure from unauthorized parties. The system uses cloud storage for ensuring privacy preservation of healthcare systems [11]. The healthcare systems require privacy for the mobile data [12]. The encryption technique has fattribute method for providing access control using auditing issues in cloud systems [13]. The health monitoring system's privacy made secured from unauthorized parties uses cloud storage for ensuring privacy preservation of healthcare systems [14].

3 System Architecture

Doctors from the hospital send medical reports of the patients to the cloud. Their port has to be encrypted before uploading it to the cloud. The doctor defines access

control policy in the form of access control tree on reports to assign which attributes the cloud proxy should obtain if they want to access specific medical reports [15]. The medical reports are encrypted using attribute encryption method. The access control policy is embedded in the ciphertext of the symmetric key. The cloud proxy sends the request to the trusted authority. Trusted authority can get the request to access the file. Once trusted authority approves the file request from the proxy cloud, who obtains the attribute keys that fulfill the access control policy can make use of the ciphertext and retrieve the file. Then the uploaded files are stored in the different data centers. Consider the case where a cloud proxy wants to share data with a set of patients in offloading. It will help to minimize the communication costs and energy usage. Moreover, predictive pushing reduces the number of redundant transfers and reduces the storage costs of the mobile devices. We deal with that although it is difficult to predict the file consumption of a user with a high degree of accuracy; low prediction accuracy will still generate a significant impact because of the increase in content availability due to content replication on the patients [16].

Figure 1 gives the overall process flow for the security system in cloud computing. Significant entities involve the doctor, trusted authority, cloud proxy server and third-party auditor who are responsible for auditing the data resources. Cloud data center requires some cloud servers for cloud data storage. Doctors are responsible for generating the secret key for their secret files and then encrypting and uploading them to the server. The data users share their resources through wireless access point in the cloud data centers. The data auditing method is performed by the cloud auditor by making use of data file cipher text.

3.1 Offloading Process

The offloading is one of the techniques for balancing the incoming load in distributed systems. Mobile network cloud computing delivers the possibility to protect energy through the accomplishment of massive load computations remotely in cloud systems. The offloading is one of the primary features of mobile cloud computing to improve the energy consumption of the mobile devices. The technique is used to enhance the performance of application's execution environment for faster processing [17]. Offloading aims to enhance the performance and save the energy before making the offloading decision, we have to identify the parts of an application that are be migrated to the cloud. It is usually achieved by partitioning the application into separate independent components. Many algorithms are used to partition computations between a mobile system and a cloud server [18]. The resource access get slower in the network traffic mainly because of more number of users access in the network. All the files are obtained for the sake of getting single type resources. If the user wants to access any of the required documents, they have to access all the files present. It makes very difficult for the user as well as for the network. To resolve this problem efficiently, we are using the technique called stochastic process. It moves apart from one format. It only processes the particular asked form of data [19]. The

Fig. 1 Process flow
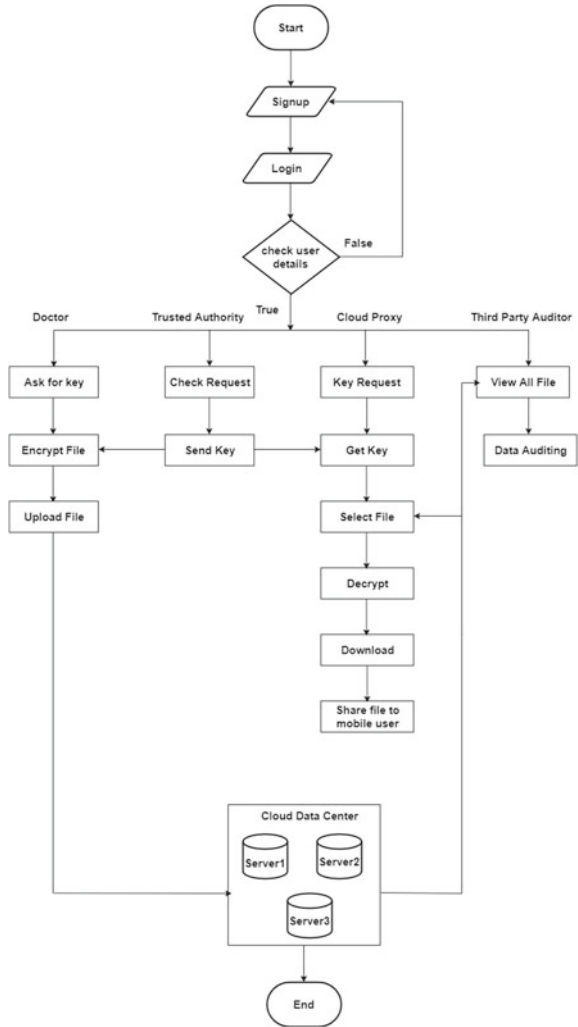
bandwidth maintenance is the primary issue to deal with for quicker access to the resource from the cloud server as well as from the proxy server. Stochastic processes are one of the fundamental solutions which divide the data into different sort of types as per type of the file. It shares majorly into video, audio, document, and images. Even in the report, it is separated into excel, word document and so on. It helps to avoid the high bandwidth to access the file. Since the bandwidth is lessened, the accessing time is also getting reduced, so it makes user-friendly process [20].
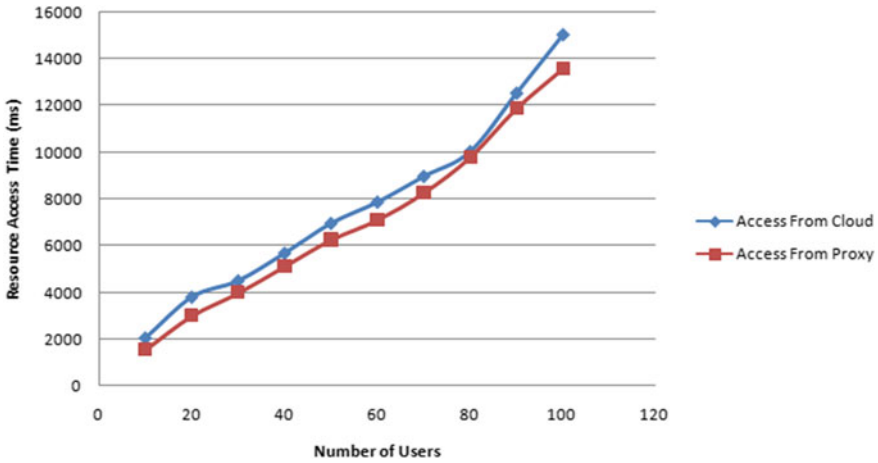
Fig. 2 Resource access time

4 Result Analysis

4.1 Improving Data Storage Capacity and Processing Time

Mobile cloud computing allows the users to store and access the data on the cloud through wireless networks so. Here the users can protect the constrained quantity of energy and storage volume on their mobile devices. The multimedia applications that need massive processing capabilities are transferred to the storage cloud system to execute. On completing the execution, results are returned to the mobile device [21]. The access time of the cloud data can be reduced comparatively by the usage of the proxy server than the cloud server as shown in Fig. 2. The graph depicts the resource access time in milliseconds with a number of the cloud users accessing the data files from the cloud servers. It says that the time taken to access the data from cloud server is more than the time taken to obtain from the cloud proxy server.

5 Conclusion

The security, privacy, and management of healthcare systems within smart cities deliver computing services and resources in cloud computing technology. The cloud adoption in the healthcare management system for smart city improves health services for public lives. Mobile cloud computing is one of the evolving mobile technology trends that combine the advantages of both mobile computing and cloud computing. Mobile cloud computing technology dramatically reduces the energy consumption in mobile devices. In the mobile cloud computing with massive computations, the

tasks are transferred to the cloud proxy server and only the results are sent back to the mobile devices. Thus, it reduces the processing time as well as the consumed power. An efficient technique known as offloading process has been used to reduce the energy consumption of mobile cloud computing in this work. Therefore, to better utilize the resources in the cloud, a trade-off between transmission energy cost and local execution energy cost gets executed. Though many tasks have made an effort to find the balance point, only a subset of those studies focuses on energy efficiency, while in many cases they merely focus on response time as well as resource consumptions. A large part of the researches uses modeling and simulation. The security establishment in the offloading process is an area of concern in the future work. The best possible way to accomplishing the security in an offloading process is the authorization of the resources which is accessed from the cloud proxy server. The approval for accessing the resources between the cloud proxy and patients is a significant way to avoid the leakage of unnecessary data resources to other users. It also prevents the colliding of data resource among the patients. It also makes sure that the reports of the patients have reached the right patients.

References

1. Rana J, Bajpayee A (2015) HealthCare monitoring and alerting system using cloud computing. Int J Recent Innov Trends Comput Commun 3(2):102–105
2. Yang K, Ou S, Chen HH (2008) On effective offloading services for resource-constrained mobile devices running heavier mobile internet applications. IEEE Commun Mag 46(1)
3. Azhar M, Laxman M (2014) Secured health monitoring system in mobile cloud computing. Int J Comput Trends Technol 13(3):138–142
4. Shen D, Luo J, Dong F, Fei X, Wang W, Jin G, Li W (2015) Stochastic modeling of dynamic right-sizing for energy-efficiency in cloud data centers. Future Gener Comput Syst 48:82–95
5. Islam MR, Habiba M, Kashem MII (2017, January) A framework for providing security to personal healthcare records. In: 2017 international conference on networking, systems and security (NSysS), pp 168–173. IEEE
6. Lin H, Shao J, Zhang C, Fang Y (2013) CAM: cloud-assisted privacy preserving mobile health monitoring. IEEE Trans Inf Forensics Secur 8(6):985–997
7. Moudgil K, Maheshwari R, Parekh HB, Devadkar K (2017) Cloud-based secure smartcard healthcare monitoring and tracking system. In: 2017 second international conference on electrical, computer and communication technologies (ICECCT), pp 1–8. IEEE
8. Vibha MB, Rakshitha Kiran P, Raju Gondkar R, Nataraja P (2017, February) Load balancing streamed healthcare data using MR technique. In: 2017 international conference on innovative mechanisms for industry applications (ICIMIA), pp 641–645. IEEE
9. Jadhav A, Bhiksham V, Vishwapathi P (2014, September) Mobile health monitoring technique using cloud computing. Int J Recent Technol Eng (IJRTE) 3(4):10–12
10. Prasad DH, Srikanth M (2015) Mobile healthcare monitoring system in mobile cloud computing. Int J Comput Technol Appl 6(1):43–46
11. Manjunath S, Koteswara Rao N (2014) Privacy towards cloud based mobile health monitoring services. Int J Sci Res Comput Sci (IJSRCS) 2(6):1–6
12. Wang B, Song W, Lou W, Thomas Hou Y (2017) Privacy-preserving pattern matching over encrypted genetic data in cloud computing. In IEEE Conference on Computer Communications, IEEE INFOCOM 2017, pp 1–9. IEEE

13. Tong Y, Sun J, Chow SSM, Li P (2014) Cloud-assisted mobile-access of health data with privacy and auditability. IEEE J Biomed Health Inform 18(2):419–429
14. Aledhari M, Marhoon A, Hamad A, Saeed F (2017, July) A new cryptography algorithm to protect cloud-based healthcare services. In: Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies, pp 37–43. IEEE
15. Kurle AS, Patil KR (2015) Survey on privacy preserving mobile health monitoring system using cloud computing. Int J Electr Electron Comput Syst (IJEECS) 3(4):2347–2820
16. Bahwaireth KS, Tawalbeh L, Basalamah A, Jararweh Y, Tawalbeh M (2015, November) Efficient techniques for energy optimization in mobile cloud computing. In: 2015 IEEE/ACS 12th international conference of computer systems and applications (AICCSA), pp 1–8. IEEE
17. Bruneo D (2014) A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems. IEEE Trans Parallel Distrib Syst 25(3):560–569
18. Dattatraya PY, Agarkhed J (2016) Offload computation in cloud assisted wireless sensor networks. In: 2016 2nd international conference on advances in electrical, electronics, information, communication and bio-informatics (AEEICB), pp 660–664. IEEE
19. Xia Y, Zhou M, Luo X, Pang S, Zhu Q (2015) A stochastic approach to analysis of energy-aware DVS-enabled cloud datacenters. IEEE Trans Syst Man Cybern Syst 45(1):73–83
20. Ma X, Cui Y, Wang L, Stojmenovic I (2012, December) Energy optimizations for mobile terminals via computation offloading. In: 2012 2nd IEEE international conference on parallel distributed and grid computing (PDGC), pp 236–241. IEEE
21. Agarkhed J, Mundewadi S, Patil SS (2016, March) Mobile health monitoring system using cloud computing. In: International conference on wireless communications, signal processing and networking (WiSPNET), pp 1301–1305. IEEE

Cloud-Based Multilayer Telemedicine Architecture: A Case Study

Snigdha Mishra, Samten D. Bhutia, Nadeem Akhtar and Sourav Dhar

Abstract Telemedicine is a life-changing product of this era, which is used to distribute health and medical information and services over a given area. Low-cost sensors, wireless systems, and internet of things (IoT) based tele-health care brings about constant care and a comprehensive monitoring capability both at home and at work. Real-time monitoring and timely responses have helped to cater a wide array of patient needs. Telemedicine provides to patients, living in secluded areas, an opportunity to receive care from specialists practising far away without paying a physical visit to them. Hybridization of technologies can provide a universal platform to healthcare experts across the globe to share information and discuss patient issues irrespective of their geographical location. This paper aims to establish and simulate cloud-based integrated multilayer telemedicine diagnosis software which takes in certain symptoms from the user and provides a diagnosis by consulting databases at the local, regional and the national level.

Keywords IoT · QoL · Cloud computing · Telemedicine · Differential diagnosis · k-Nearest neighbours' algorithm

1 Introduction

IoT-based telemedicine has broken the distance barriers and improved the medical facilities that would often not be easily available in remote areas [1]. Saving of lives in critical care and emergency situations has become possible to a large extent due to the implementation of telemedicine [2]. It permits exchanges between the long-suffering patients and health workers with both dexterity and conformity. The transmission of medical, imaging and health informatics data from one site to another is also taken care of. Telemedicine enables a better quality of life (QoL) and a cost-effective

---

S. Mishra · S. D. Bhutia · N. Akhtar · S. Dhar (✉)
Department of Electronics & Communication Engineering, Sikkim Manipal Institute Technology, Sikkim Manipal University, Majitar, Rangpo 737136, Sikkim, India
e-mail: sourav.dhar80@gmail.com

553

care for patients with life-threatening diseases, such as chronic respiratory diseases, diabetes, cardiovascular disease and even cancer [3]. It also comprises applications for fit people, who want to maintain/ improve their health by constant monitoring. The telemedicine will be advantageous from different perspectives viz. the patient's perspective, the provider's perspective and the economic perspective.

From a patient's perspective, access to the high-quality and modern healthcare infrastructures in underserved areas, such as rural communities, is amongst the significant and most important promised benefits of telemedicine. It also allows the patient to obtain a 'second opinion' diagnosis from diverse sources [4].

As can be seen from a provider's perspective, telemedicine provides prompt access to statistics of an individual patient or a particular topic that can be vital or even lifesaving [2]. By the use of 'tele-assistance' (e.g. tele-consultation with specialists) [5–7], it is expected that the doctors will also be beneficial to get expert advice which can help further in reduction of medical errors.

From economic development and QoL perspective, telemedicine provides advancement in delivery of services, keeps money in the local economy, helps business employment and retention, leads to development of workforce and provides better QoL and longevity [4].

2 Proposed Architecture

Authors, earlier, have introduced single layer architecture of telemedicine in their previous work [8]. Extending the work of [8], Fig. 1 depicts the multilayer telemedicine network. Here, the patient can be monitored continuously with help of a set of sensors and the measured data may be fed to the diagnosis software either automatically through wireless sensor network (WSN) or manually. The diagnosis software at the patients end is interfaced with a control unit in the local/rural hospital or health centre (layer-I) via link 1 which is a wireless technology for a low-range communication. The rural health centre has a database consisting of common diseases, their symptoms, tests to be undertaken to make a diagnosis and a differential diagnosis. The rural health centre is wirelessly connected to the regional hospital/ district hospital database (layer-II) via link 2. The district hospital has a database comprising of common along with typical diseases, their symptoms, tests to be undertaken to make a diagnosis and a differential diagnosis. The typical diseases are more severe than common diseases and require more extravagant examination for which facilities may not be available in district hospital. Thus, the district hospital is further connected to some metro hospital (Layer-III) via link 3. The metro hospital has a huge database comprising of common, typical as well as critical diseases, their symptoms, tests to be undertaken to make a diagnosis and a differential diagnosis. The critical diseases involve life-threatening diseases that need to be attended on priority and those patients are to be monitored round the clock. Hence, this cloud-based architecture will provide quick diagnosis.
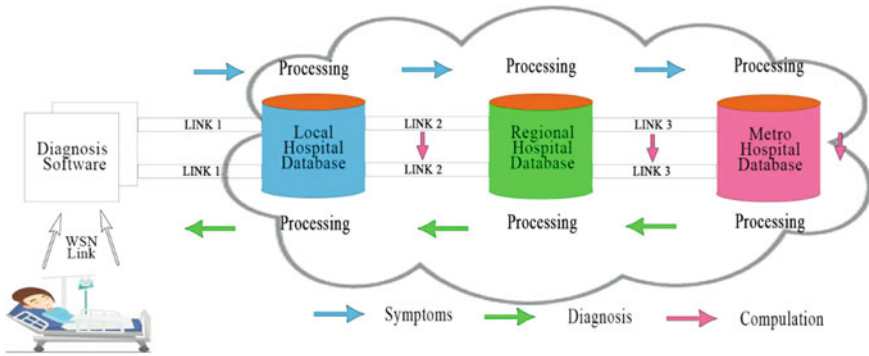
Fig. 1 Multilayer telemedicine network architecture

3 Creation of Database

Three different databases have been created, each for the local/ rural, regional/district and metro hospitals respectively. The databases comprise the diseases, their symptoms, diagnostic tests along with differential diagnosis. These diseases have been classified as common, typical and critical and shown in tree diagram in Fig. 2.
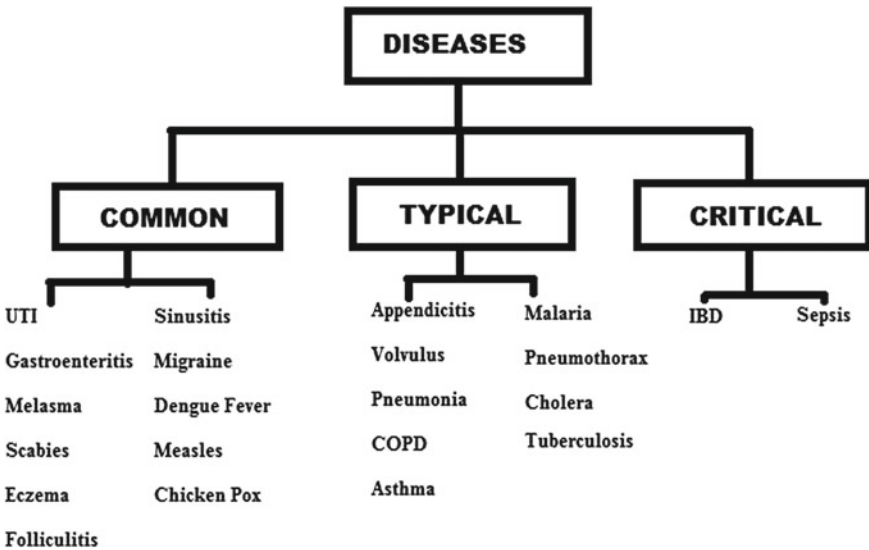
Fig. 2 Tree diagram for the classification of diseases

3.1 Entity Relationship Diagram (ERD)

The primary object that the ERD represents is the entity which has an independent existence. It could either be a physical object (e.g. a person, car or a house) or an abstract object (e.g. a job, a school course or a company). The specific properties that define an entity are called attributes. The 'entity type' describes a particular set of entities with similar attributes. This describes the schema or intentions for a group of entities which utilize the same structure. The entity types in the database are addressed by their name and attributes.

The entity set is the group of all entries of a specific entity type in the database at any given instant. Figure 3 shows the E-R model for the telemedicine network. Our first job is to identify the entities for the model. The patient(s), symptom(s) and disease(s) are chosen as the three entities and the connection among them is determined.

For each of the three entities, name and identity are taken as constant characteristics as they are unique to themselves. The relationship among them can be easily understood from the E-R model. The diagnosis of disease for a patient may be done by considering multiple symptoms.

- Many symptoms can lead to a single disease.
- Many diseases can have some common symptoms.
- A patient has a disease due to which he may show one of many symptoms.
- Not all symptoms will be present in one patient.
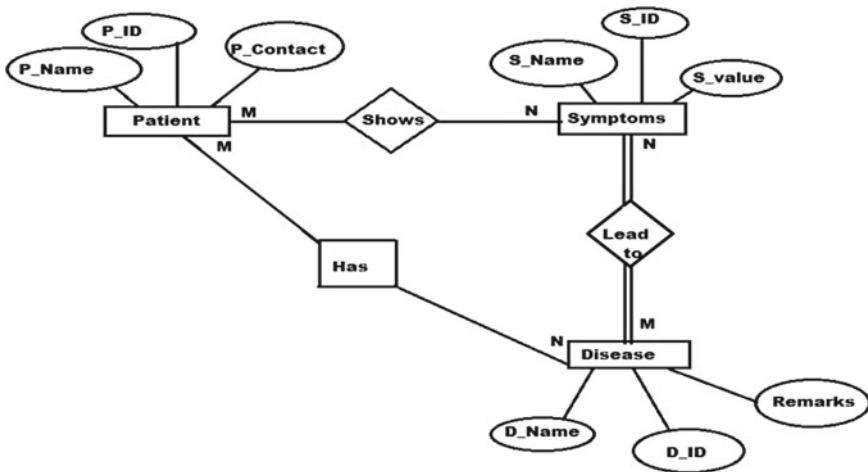
Fig. 3 ERD of the diagnosis software

3.2 Data Flow Diagram (DFD)

The primary objective of DFD is to graphically represent the flow of data through an information system and modelling its process features. Input and output data of the system and where the data will be stored are precisely shown in DFD. Hence, visualization of data processing is made easy by DFD. However, information regarding the timing and types (sequential or parallel) of processes are not depicted in DFD (Fig. 4).

The 0-level DFD, or the context-level DFD, shows the 'Telemedicine system' bubble interacting with the user in terms of the user query and system response. The internal organization of the system is not shown in context-level DFD and the entire system is shown as a single process.

This level 0 DFD is then 'exploded' to produce the level 1 DFD which depicts some of the particulars of the system being modelled. The first 'process' is the task of reading the symptoms from a patient and storing them for the purpose of comparing with either of the databases. The user is given a choice as to which database he wishes to refer to. Depending on the choice made by the user, he is provided access to the local, regional or the national database enlisting various diseases, their symptoms, tests required for diagnosing them and a differential diagnosis. This procedure is followed for the diagnosis of the disease by the system and providing the results back to the user.
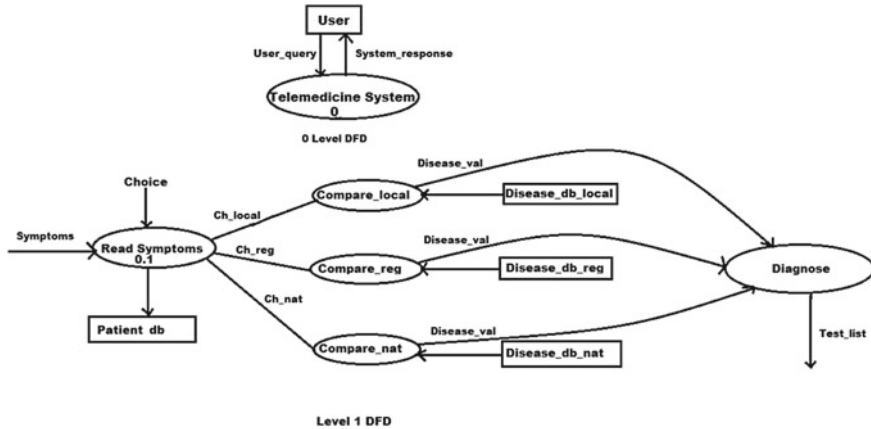
Fig. 4 DFD of the diagnosis software

3.3 Diagnosis Algorithm

The symptoms of the patient are coded in the form of zeroes and ones, depending on whether they are present or absent, and stored in the form of a matrix.

The databases are created in the same fashion coding the presence and absence of symptoms in the form of zeroes and ones and stored in the form of another matrix.

The classification takes place through k-nearest neighbours' algorithm.

The k-nearest neighbours' (k-NN) algorithm is an effective non-parametric technique for classification and regression. The k-NN can predict the values or class memberships of objects depending on the k neighbouring training examples in the feature space. The k-NN is an instance-based learning (lazy learning), here only the local approximation of the function is carried out, and hence all computations are delayed until the classification completes. The k-NN is one of the primary machine learning algorithms: assuming k as a small positive integer, an object is categorized by a majority choice of its neighbours, with the object being allotted to the class most likely amongst its k closest neighbours. Unity value of 'k' signifies that the object has been assigned to the class of the single nearest neighbour.

The algorithm can be potted as

1. k is a predefined small positive integer with a new sample.
2. k entries are selected in the given database those are closest to the new sample.
3. The utmost general classification of these entries is identified.
4. The new sample is assigned this classification.

Figure 5 demonstrates an example of k-NN-based classification. Let us consider that the green pentagon is the test sample which needs to be classified either to the class of blue triangles (class-I) or to the class of red stars (class-II). If 'k' value is 3 (circle with solid line), then it is assigned to class-II since there are two stars and one triangle inside the first circle. On the contrary, 'k' value is 5 (circle with dashed line), then it is assigned to class-I (since there are three triangles and two stars inside the second circle).

Fig. 5 Example of k-NN classification
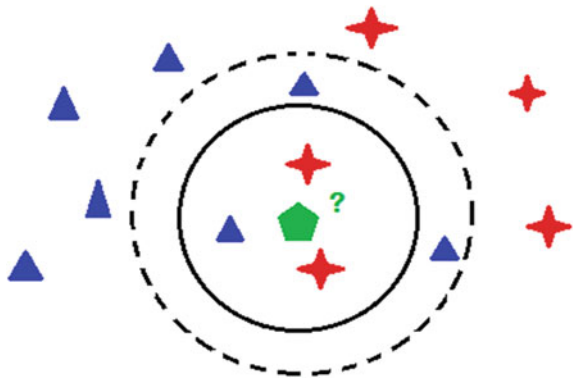
4   Validation of Telemedicine Software

We can comprehend the working of the cloud-based telemedicine system by considering the case of a certain patient in a rural dispensary or hospital who must be diagnosed (shown in Fig. 1).

Let us assume that the patient comes to us with the following symptoms:

• Vomiting	• Rashes
• Diarrhoea	• Inflammation of eye
• Rectal bleeding	• Tiredness
• Abdominal pain	• Lack of concentration
• Weight loss	• Dehydration

We enter the details in the diagnosis software and find out the diagnosis in case of each of the following:

Layer-I: The local hospital or rural dispensary.
Layer-II: The regional/district hospital.
Layer-III: The national/metro hospital

Figure 6 shows a window of the diagnosis software which enlists various parameters, or rather, symptoms based on which the diagnosis is made. This includes parameters such as breath rate, body temperature, etc.

Fig. 6 Window for entering the symptoms
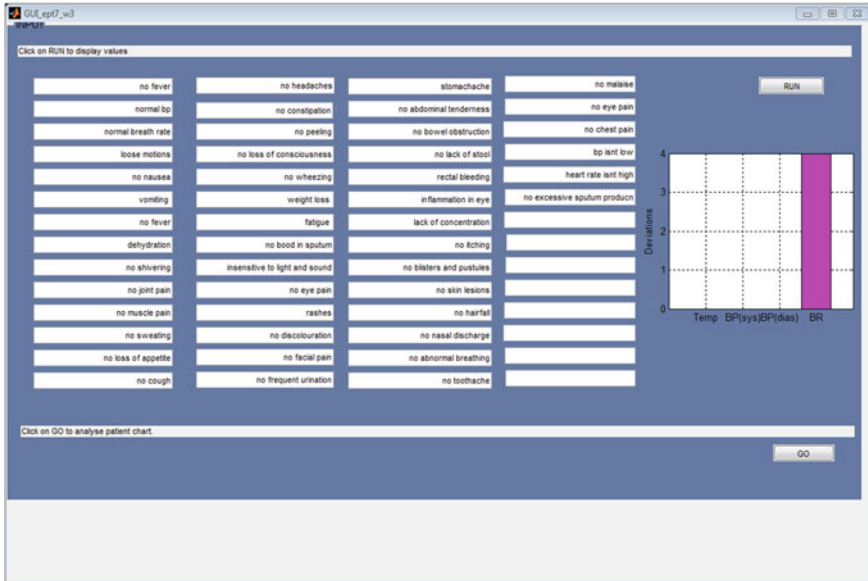
Fig. 7 Generation of patients chart
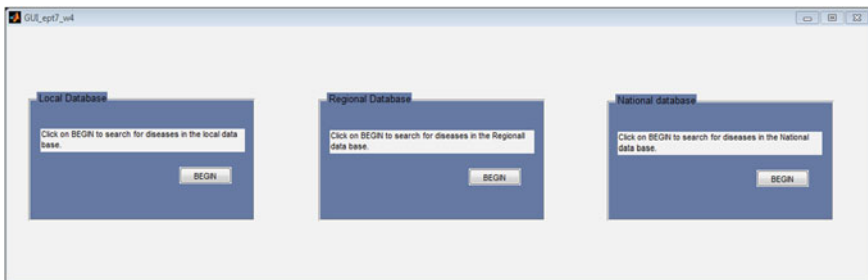
Fig. 8 Choice of database

The next window of the diagnosis software, which is shown in Fig. 7, generates the patient chart containing all of his/her symptoms along with a graph showing the deviations from normal body temperature, blood pressure and breath rate. From here, we can proceed to diagnosing the patient with the help of the patient chart.

Figure 8 shows a window which has provisions for the user to determine whether he/she wants to refer to the local hospital or dispensary, regional hospital or a national hospital subject to the severity of the patient.

The diagnosis made may be same or different in case of each of the levels of the databases. In order to make an appropriate diagnosis, the relevant tests should be carried out based on which the patient shall be treated.

5 Discussion and Conclusion

Currently, there exist a few software models, according to our literature survey, that can be used for transmission of data from telemedicine perspective. The traditional methods involve only data transfer from one station to another. In our work, we have designed and validated the multilayer cloud-based telemedicine network architecture by developing a user interface using MATLAB for both implementing data transfer as well as to diagnosing the diseases using k-nearest neighbour algorithm. Furthermore, with the consultation of a doctor, we have validated the software with a case study. The system has been designed assuming a limited number of symptoms and diseases. With more research and consultation with a medical team, larger databases containing more number of diseases, symptoms and tests can be incorporated.

References

1. Distefano S et al (2017) Hospitalized patient monitoring and early treatment using IoT and cloud. Bio Nano Sci 7(2):382–385 (Springer). https://doi.org/10.1007/s12668-016-0335-5
2. Jeong YS, Shin SS (2016) An IoT healthcare service model of a vehicle using implantable devices. Clust Comput, 1–10 (Online First). https://doi.org/10.1007/s10586-016-0689-z
3. Currell R, Urquhart C, Wainwright P, Lewis R (2000) Telemedicine versus face to face patient care: effects on professional practice and health care outcomes. Cochrane Database Syst Rev, no 2. Art. No.: CD002098. https://doi.org/10.1002/14651858.cd002098
4. White Paper: ISRO (2005) Telemedicine healing touch through space: enabling specialty health care to the rural and remote population of India
5. Sîrbu N, Andrei V, Cleju II (2008) A Zigbee solution for telemedicine applications. ACTA Tech Napoc 49(3)
6. Carmen C, Poon Y, Zhang Y (2006) A novel biometrics method to secure wireless body area sensor networks for telemedicine and M-health. IEEE Commun Mag
7. Dhar S, Ray A, Bera R (2013) Cognitive vertical handover engine for vehicular communication. Peer-to-Peer Netw Appl 6:305 (Springer). https://doi.org/10.1007/s12083-012-0171-5
8. Mishra S, Bhutia SD, Dhar S (2014) Design and simulation of tele health care architecture. In: National conference on emerging trends in communication & biomedical engineering (NCECB-2014), SSIT, Tumkur, Karnataka, India-572105, 9–10 May 2014

Smart Water Pump Controller

Urmika Sengupta, Priyanka Kumari, Tuhin Paul and Md Ruhul Islam

Abstract This project aims toward developing a technology for controlling water pumps that can be made with smartphones. A representation of an Android-based application is made which serves as a model through which the operations of the water pump can be controlled with the help of a microcontroller like NodeMCU and a Wi-Fi router. The minimum operating system version required for the application is Android Lollipop. The project depicts the connections between the required equipment in brief. The proposal can be helpful to prevent wastage of water as well as electricity on implementation. Also, it makes us aware of how much electricity is consumed normally which is not available in other models. Finding the amount of water present in any tank overhead or underground is a difficult task. Usually, anyone will end up climbing up the stairs to the tank, for checking the water level manually or will be hearing the sound of water overflowing from the overflow outlet. But nowadays many such electronic indicators are available in the market which are able to fix some problem, but they are often costly enough to empty your pocket to quite an extent in your pocket and are usually difficult to install. Most of the systems which are available in market currently, generally use dipped electrodes or float switches, which can be a headache in the long run. We present a different way of finding the water level using an ultrasonic module with the help of Arduino. Since the method is contactless, so various issues related to the electrodes, for example, corrosion would not affect this system, Furthermore, this Arduino water level indicator is much easier to install than regular systems.

U. Sengupta · P. Kumari · T. Paul · M. R. Islam (✉)
Department of Computer Science and Engineering,
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar,
Rangpo 737136, Sikkim, India
e-mail: ruhulislam786@gmail.com

U. Sengupta
e-mail: urmikasengupta@gmail.com

P. Kumari
e-mail: psdpriya23@gmail.com

T. Paul
e-mail: mynameistuhin.tp@gmail.com

563

1 Introduction

Every day thousands of pumps are used to fill up water tanks. The process of switching ON the pump and switching it OFF after the tank is filled is a manual process, i.e., tank has to be manually checked continuously if enough water has been filled, if not so, then the pump has to be switched ON manually to fill it up, then we also need to keep a check that if the tank is filled completely or if water is overflowing and turn the pump OFF accordingly. After a tank is filled up, lots of water is wasted before turning the pump off. According to a case study, the amount of water that an average household wastes per day through leakages or overflowing adds up to 10,000 gallons of water each year [from web]. The present rate of wastage during day-to-day life in present day is a major concern at the time when water level has gone down below the critical level making it a huge problem to get water which is fit for drinking purpose at number of locations all over the world [1]. It is also seen that even though the tank is filled, there is no one present at that moment to switch the pump off. The result of it is wastage of water. This project proposes a technology to put an end to the ancient or age-old practices. Smartphones nowadays are widely used technology in today's daily life [2]. Apps have made daily life more convenient. Everything nowadays is one touch away from sending money to launching a nuclear missile. Smartphones, nowadays, are one of the basic needs of an average human and as a step toward the achievement of Digital India, this project shows a way to prevent the wastage and reducing water wastage by bringing it down to 0% using Application for Android-based platform and hardware [2]. This application can be helpful to check the water level in a tank and then to turn the pump ON and OFF from his house accordingly.

2 Problem Definition

The main problem that arises with manual water pumps available in the market is that you have to be physically present to switch the pump ON and OFF. Another matter on concern is that there is no information regarding the water level so people have almost no idea how much time they should keep the motor ON so that the tank fills up without overflowing. We, here, are trying to overcome the problem of overflowing which is a wastage of water as well as extra consumption of electricity. Also, normally pumps do not give out the amount of water consumed or amount of electricity consumed during the duration of the pump running. This smart pump controller is intended to overcome these problems and help to ease the routine activity of almost every individual with a simple and effective solution.

3   Analysis of Problem

We will use Arduino for connecting various sensors and modules. Sensors will be used for the purpose of collecting various information like amount of water flowing into the tank, level of water in the tank presence of water in the underground reservoir tank. An Android application is built for the purpose of collecting and providing information to the user. Android application can be used to fix time of switching ON and switching OFF the pump. The Android application will have a facility of record-keeping of the daily usage of water as well as amount of electricity consumed. For calculation of the electricity consumed, we will be using formula power consumed = watt × time used. And, unit = power consumed/1000. And, the level of water in the tank is determined depending on the reading of the ultrasonic sensor.

The modules that will be used are

   (i)   An Arduino,
  (ii)   Moisture sensor,
 (iii)   Ultrasound sensor,
 (iv)   Breadboard,
   (v)   Jumper wires,
 (vi)   Relay,
 (vii)   Node MCU, and
(viii)   Android app.

4   Proposed Solution Strategy

Figure 1 shows us a basic layout of the system that will be working using the Android application connection via **ES8266** to the hardware and using Arduino UNO by replacing Amtega128.

Fig. 1 Block diagram [1]
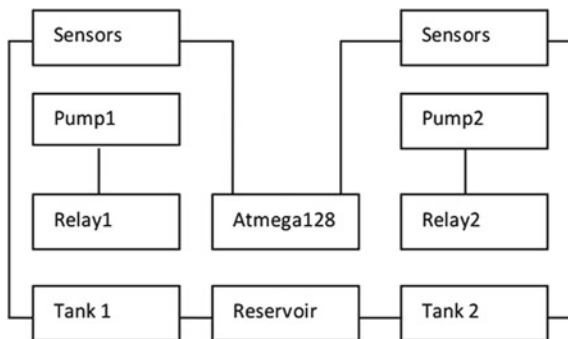
5   Design Strategy for the Solution

Figure 2 shown above gives complete working modes including hardware and software (i.e., Android application) and their internal decision-making schemes and parameters.
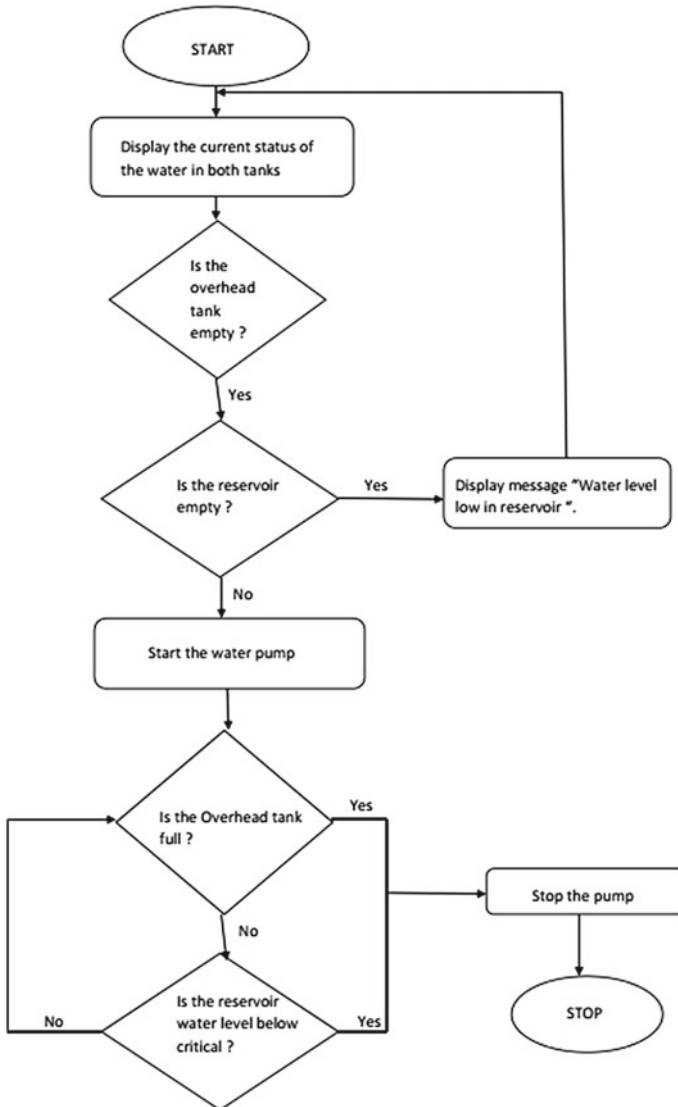
Fig. 2 Working model flow diagram
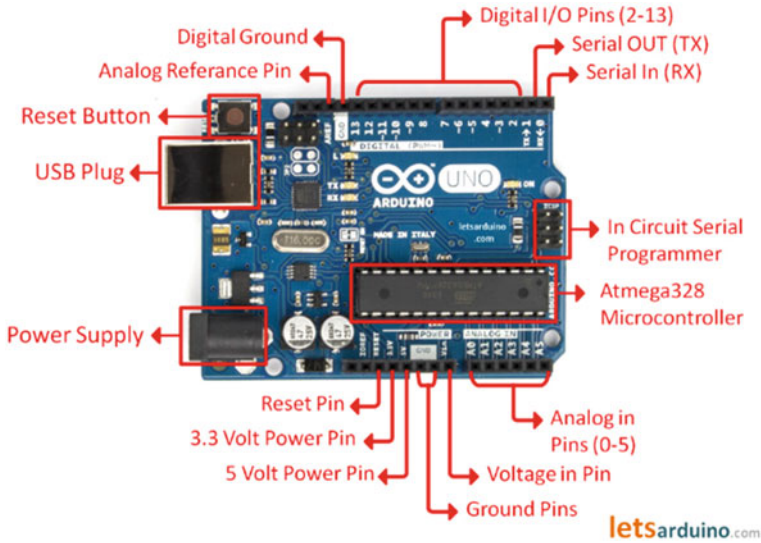
Fig. 3 An Arduino UNO [2]

6 Working Principle

6.1 Arduino UNO

An Arduino is originally a microcontroller-based kit. It can be purchased and used or can be assembled and make the same at home using the components, owing to its open-source hardware feature. The communication and controlling or operating many devices is possible with this microcontroller-based kit [3].

In Arduino Uno, 14 digital pins are present that can be used both as input or output (of which 6 can be used as PWM outputs), 6 analog inputs, a 16 MHz crystal oscillator, an USB connection, a power jack, an ICSP header, and a reset button [4] (Fig. 3).

6.2 Raindrop Sensor

Raindrop sensor comprises a board on which nickel is coated in the form of straight lines which are parallel to each other. The principle that plays a role is resistance principle. When there is no drop of water on board the resistance becomes high, and therefore we get high voltage according to the formula, $V = IR$, where V, I, and R represent voltage, current, and the resistance, respectively. In presence of raindrop, the resistance reduces because water is a conductor of electricity and presence of

Fig. 4 A raindrop sensor
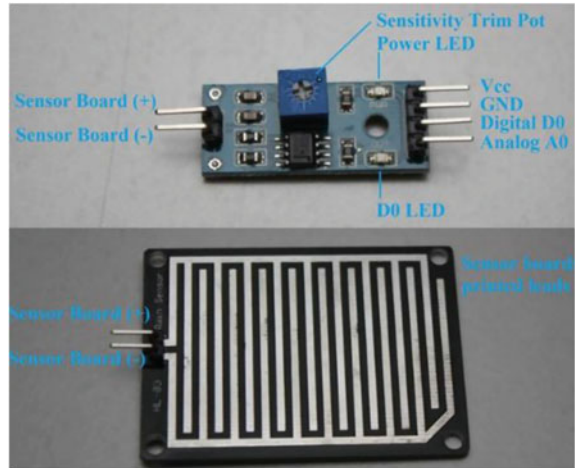
water connects lines made out of nickel on the board in parallel, so the decrease in resistance results in reduced voltage across the circuit board [5] (Fig. 4).

6.3   Ultrasound Sensor

This module emits ultrasound waves having frequency 40,000 Hz. The mode through ultrasound wave's travel is air and if there is any object or obstacle on its path, the wave bounces back to the module. Considering travel time of wave to return and the speed of the sound, the distance of object from the module can be calculated.

In HC-SR04 ultrasonic module, four pins, namely, Ground, VCC, Trig, and an Echo pin are present. The pins present in module which is marked as Ground and VCC need to be connected to the Ground and the 5 V pins present on the Arduino board, respectively. Whereas the Trigger and the Echo pins can be connected to any of the digital input/output pins present on the Arduino board [4] (Fig. 5).

6.4   Breadboard

Breadboard is the most basic and important tool to design and test any circuit. Soldering wires and components to make a circuit is not required if a breadboard is used. It is easier to connect components and reuse them this way. Since components are not fixed using any extra component like glue/soldering, we can change our circuit design at any point. With soldering the whole setup of wires and components performing any change becomes a tedious job. A breadboard is a collection of conductive metal connection ports put in a box made of plastic, where each clip is insulated with

Fig. 5 Ultrasonic sensor [4]

another clip [from web]. Number of holes is made on the plastic box, arranged in a specific pattern. A breadboard consists of two regions also known as strips, these strips are, namely, bus strips and socket strips. To distribute power supply to the circuit a type of strip known as bus strips is generally used. Bus strips consist of two columns, out of those two, one serves as power voltage and other as ground. Most of the components in a circuit are held by socket strips. A socket strip consists of two sections each with 5 rows and 64 columns. Every column is electrically connected from inside [4] (Fig. 6).
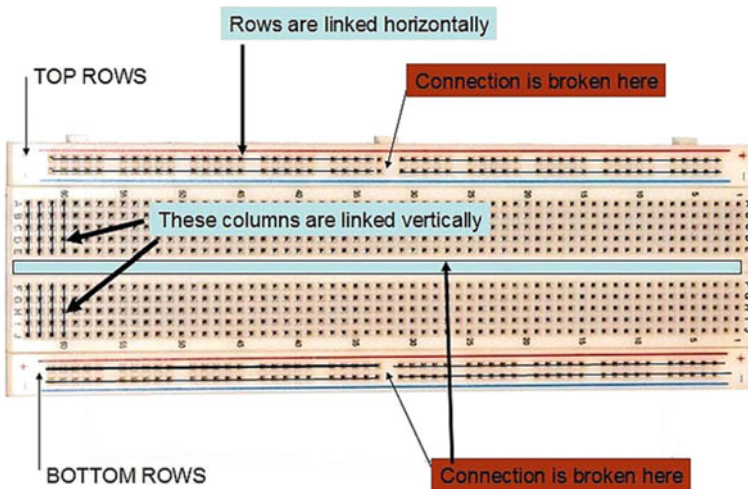
Fig. 6 A breadboard [10]
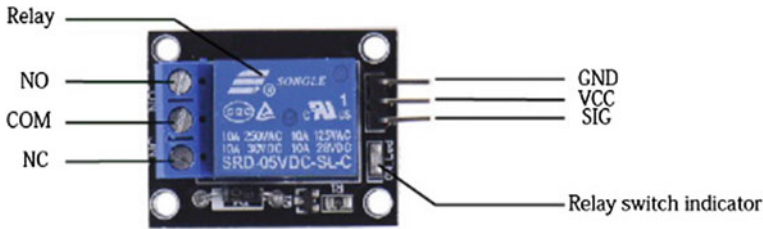
Fig. 7 A relay module [7]

6.5 Relay

Relays are electrical switches that are controlled by another switch remotely [6], e.g., a horn switch. A small amount of current flowing through the relay circuit allows the controlling of a higher current circuit. The relay that is used here for controlling power is a Single Pole Double Throw (SPDT) which has magnetizing coil terminals and they operate on 6 V DC supply. Terminals of relay used are described as follows [7] (Fig. 7).

6.6 NodeMCU

A NodeMCU includes a particular type of firmware that runs on the ESP8266 Wi-Fi system on a chip which is made by Espressif systems, a Chinese manufacturer, and hardware is based on the ESP-12 module. The term "NodeMCU" usually refers to the firmware rather than the development kit [8]. The firmware uses the scripting language Lua. NodeMCU is nothing [6] but basically a combination of Esp8266 Wi-Fi module along with USB jack and some input/output pins. So that the microcontroller can directly be powered using a mobile phone. To know its working, we can consider the datasheet of Esp8266 Wi-Fi chip [9] (Fig. 8).

7 Implementation and Result

See Table 1.

Fig. 8 An ES8266 [8]
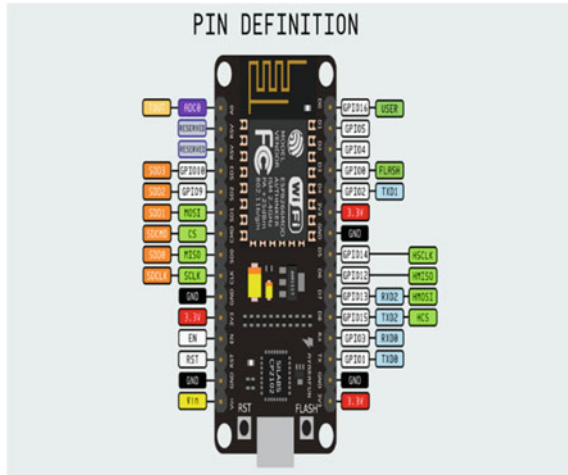
Table 1 Implementation and results

Sl. no.	Test case	Response
1	Testing Arduino by blinking in-built LED	LED blinking Arduino working message
2	Water is applied to the raindrop sensor plate	Analog reading of the sensor displayed on serial monitor
3	Finding obstacle/water-level distance using ultrasonic sensor	Depth of water displayed
4	Turning ON lamp using low-level trigger	Bulb glowing
5	User sending ON through App	In-built LED of Arduino turns ON
6	User sending OFF through App	In-built LED of Arduino turns OFF
7	User clicks on switch button	Tank is full
8	Click the status button on menu	Display status of the pump
9	User enter wrong time, i.e., past time	Timer will not set
10	User clicks OFF switch when time is not completed	Display alert message

8 Conclusion

We have successfully automated the process of filling up the tank from critical level to the desired level. The connection between the application and Arduino is working seamlessly. The users have the option to choose between the pump to be switched ON or switched OFF. There is an additional provision in which the pump gets switched ON itself if the level of the water reaches below the critical level and switches OFF itself when the water level reaches the threshold level. The user can also monitor the

amount of electricity that has been used by motor during its runtime. The connection between Arduino and relay is wired and not wireless. The connection of Arduino and application is connected to the same router so that it can have a fixed range and cannot be accessed from anywhere. The working as well as the exchange of information happens locally and it does not involve the connection of the Internet for updating values.

References

1. Patil Y, Singh R (2014) Smart water tank management system for residential purpose. Int J Sci Eng Res 5, 6 June 2014
2. Gaikwad1 PV, Kalshetty YR (2015) Bluetooth based smart automation system using android. Int J Sci Res (IJSR). ISSN (Online):2319–7064
3. Roy S, Wangchuk TR, Bhatt R (2016) Arduino based Bluetooth controlled robot. Int J Eng Trends Technol 32
4. Latha NA, Murthy BR, Kumar KB (2013) Distance sensing with ultrasonic sensor and Arduino. Int J Adv Res Ideas Innov Technol 2
5. https://www.openhacks.com
6. Kumar V, Suryavanshi1 RS, Khivensara K, Hussain G, Bansal N (2014) Home automation system using android and WiFi. Int J Eng Comput Sci 3(10):8792–8794. ISSN:2319-7242
7. Nasution TH, Muchtar MA, Siregar I, Andayani U, Christian E (2016) Electrical appliances control prototype by using GSM module and Arduino
8. https://arduinobasics.blogspot.in/2014/09/relay-module.html
9. https://en.wikipedia.org/wiki/NodeMCU
10. http://wiring.org.co/learning/tutorials/breadboard/

IoT-Based Smart Home

Khara Nanda Sharma, Nirmal Rai, Manish Tamang, Jabrang Basumataray and Swastika Chakraborty

Abstract In recent years, there has been a huge growth in the field of IOT-based smart home, where different home appliances are connected to the cloud such that a user can control and monitor it remotely. This paper presents a concept for controlling and determining the actual state of the appliances through the mechanism of automatic feedback to the user. A circuit consisting of current sensor at the device side is used to determine whether the device is actually active or not. Smart monitoring of electricity and water usage using sensors and IoT is also proposed and implemented in this paper.

Keywords IoT · Smart home · Smart switch · Water flow sensor · Raspberry pi

1 Introduction

In IoT-based smart home various things like lightning, home appliances, security cameras, etc are connected to the Internet. In addition to this, various information collected through electronic sensors are sent to the user through the Internet. Such a system enables a user to control and monitor their home environment remotely through the Internet.

Considerable amount of work has been done in the field of smart home. Various research groups are working on this domain, such as MIT, Siemens, Cisco, IBM, Xerox, Microsoft, etc [1]. Bluetooth- and Ethernet-based smart home have also been implemented [2]. Patru et al. [3] has presented a solution for connecting more types of home appliances in one system using gateways. Malche and Maheshwary [4]

K. N. Sharma (✉) · M. Tamang · J. Basumataray
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India
e-mail: kharananda0000@gmail.com; Khaling_1@yahoo.com

N. Rai · S. Chakraborty
EC Department, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India
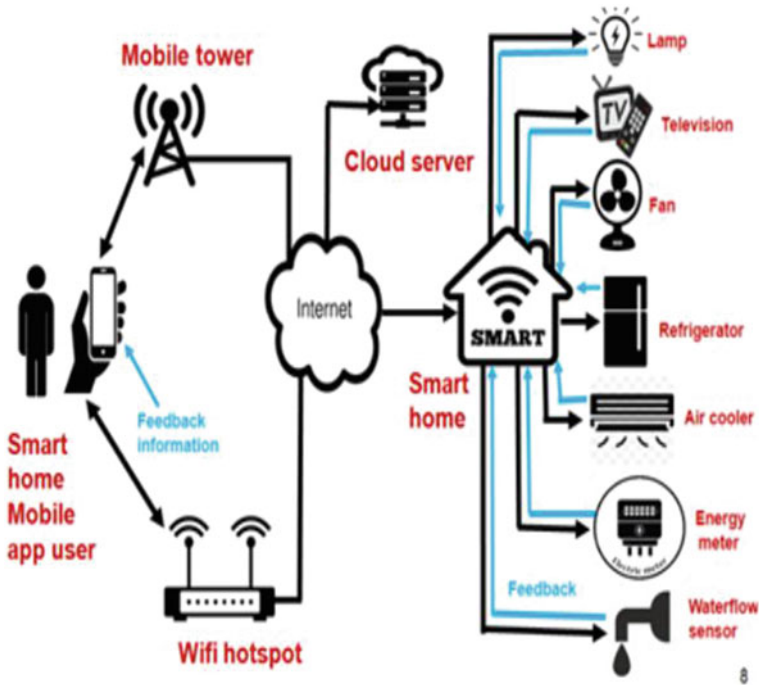
Fig. 1 Design concept of the IoT-based smart home system

has discussed the functions of smart home and its applications and introduces FLIP architecture. FLIP developed by Frugal Labs Bangalore, India is an open-source IoT platform.

All the smart-home-related works are focused on controlling and monitoring operations and the available literature shows that no work has been done for providing feedback from the devices to the users.

This paper presents a smart home concept which enables remote controlling of appliance, monitoring of electricity, and water consumption as well as provides feedback to the user about the real status of the appliances (Fig. 1).

2 System Design

The entire system design can be divided into three parts. They are user interface, controller with feedback, and water and electricity monitoring unit.

The user interface part consists of an Android app (Fig. 2). It enables the user to control the appliances remotely through the Internet as well as monitor electricity and water consumption in their homes.

Fig. 2 Android app for user interface

The controller controls the home appliances as well as connects them to the Internet. The Raspberry Pi acts as the controller. The current sensor detects the real state of the appliances. Output signal from the current sensor is processed and forwarded to the user, enabling him to know the real state of the device.

The water flow sensor is used to monitor the water consumption. The electricity consumption monitoring is done by processing the output from the household digital electric meter (Fig. 3).

3 Working of the System

3.1 Appliances Feedback System

As the user at remote location does not have any information about the status of device, i.e., device is active or not. To overcome this problem the concept of feedback is implemented. For feedback, the current sensor senses the flow of current and its output is fed to Arduino for processing and finally its output is fed to Raspberry Pi for sending information to the user (Fig. 4).
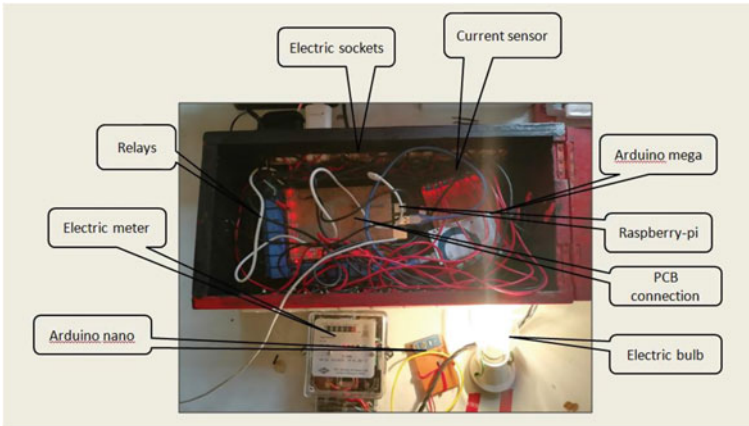
Fig. 3 Device setup
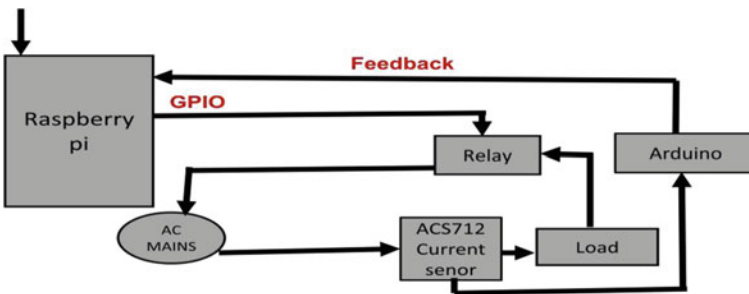
Fig. 4 Appliances feedback system

Current sensor (ACS712) is used to sense the current flowing through the appliance. It works on the principle of Hall effect and produces output voltage proportional to the current flowing in the circuit.

3.2 Water and Electricity Monitoring Unit

Flow sensor measures the speed of fluid passing through a pipe to measure the volumetric flow using laws of electromagnetic induction. As the fluid flows through the pipe, the rotor rotates and electric pulse is generated at the output of sensor. By processing the number of pulses, we can determine the water consumption of the house.

Consumption of energy is measured by counting the pulses from the LED pin of the digital energy meter. Each pulse indicates 0.0,001,562 KWH of power consumption.

DATE	CURRENT TIME	USER	DEVICE	TRIGGER STATUS	DEVICE STATUS	ENERGY CONSUMED	CHARGE
24:4:2018	13:23: -0.06	jackandjohn	lamp_1	on	on	0.55 watt hour	Rs.0.0028
24:4:2018	13:23: -0.06	john	lamp_2	off	Disconnected	0 watt hour	Rs.0

Fig. 5 Output of Android app

The Arduino unit will count the number of pulses and forwards the count value to Raspberry Pi by serial communication. The energy consumption and cost are calculated and saved in Raspberry Pi database. The user can access the information through the Android app.

4   Testing and Results

For testing the operation of feedback system, we checked feedback information in the Android app after removing the electric bulb from its holder. When the bulb was removed the "DEVICE STATUS" changed from "on" to "disconnected." After this, when we connect a damaged bulb, the "DEVICE STATUS" still showed "disconnected" status as shown in Fig. 5. The "DEVICE STATUS" changed to "on" when we connect a working bulb.

For testing the electricity monitoring unit, we made a setup as shown in Figs. 6 and 7, consisting of a load of 60 W bulb and a digital electric meter. After switching on the bulb, we determined the time interval between two pulses or LED blink. The value obtained by multiplying the power of the bulb and time interval was approximately equal to the value specified for the meter. Testing of water consumption through flow sensor was done manually by passing a fixed quantity of water through the sensor.

5   Conclusion

The above-proposed system can be used to design a smart home based on IOT. The system enables a user to remotely control their home appliances as well as monitor the electricity and water consumption of their homes. The concept of appliances feedback implemented in the system enables the user to know the real status of the appliances.

Fig. 6 Testing of electricity consumption monitoring unit

Fig. 7 System setup for electricity monitoring

References

1. Jiang L, Llu D-Y, Yanghakor B (2004) Smart home research. In: Proceedings of the third international conference on machine learning and cybernetics, Shanghai, 26–29 Aug 2004
2. Mandula K, Parupalli R, Murty CHAS, Magesh E, Lunagariya R (2015) Mobile based Horne automation using internet of things (IoT). In: 2015 international conference on control, instrumentation, communication and computational technologies (lCCICCT)
3. Patru II, Carebas M, Barbulescu M (2016) Smart home IoT system. In: IEEE conference, IEEE Xplore
4. Malche T, Maheshwary P (2017) Internet of things (IoT) for building smart home system. In: International conference on I-SMAC, IoT in social, mobile, analytics and cloud, I-SMAC 2017

Implementation of Home Surveillance System Using Local Area Network and Internet of Things

Amit Agarwal, Rigzing Norbu Sherpa and Ankit Kumar Singh

Abstract As the home surveillance and security via IP address are getting more popular since they are wirelessly monitored and need not to physically monitor it. This is mostly related to physically handicapped people and old people at home. TCP/IP network is a web of network where different physical devices like electronic devices and home appliance can be connected and allows the people to access the data from anywhere. This paper focuses on key points like establishing connection between controllers through static IP address, transfer of data from different sensor nodes to central hub and then to the web server. This data will be accessed from local area network and from any part of the world using the Internet.

Keywords Internet of things · ESP8266 Wi-Fi module · Networks · Security · Sensors

1 Introduction

Internet of things (IoT) [1] is interconnection of different devices to a network for exchanging data and to control devices as per the application. Modern IoT systems have used standards of TCP/IP protocol suite [2] of which most of the IoT applications are of low-power end devices and because of this IoT employs low-energy-layer-2 technologies, such as IEEE 802.15.4 [3], Bluetooth, low-power Wi-Fi module. Most of the IoT applications are performed by sensors and actuators for various controlling and monitoring parts [4].

Kanase et al. [5] have worked on smart hospitals using IOT which comprises only one client, which sends data to the Internet. Razzaq et al. [6] focus on the fact that IOT has security issues but it did not mention how security issues can be solved. Chari et al. discussed on monitoring of system on a cloud through wired system

A. Agarwal (✉) · R. N. Sherpa · A. K. Singh
Department of Electronics and Communication, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, Rangpo 737136, Sikkim, India
e-mail: amiteng2007@gmail.com

579

network [7], which itself is the limitation for network-based IoT system. Sfikas et al. use the MQTT protocol [8, 9]. An MQTT client is running at each of the available sensors, connected but due to the fact MQTT protocol uses few bytes to describe the content of messages. So there is a limitation in the amount of data being sent to the web server. In this paper, we have resolved the above issues using a wireless network system. This system is more flexible, as it does not always need the Internet to receive and send data to the web server. We have also resolved the security issues to an extent using local area network ssid and password. We have used HTTP protocol [10] as it will be able to send larger bytes of data.

2 System Design

Figure 1 shows the hardware model of our system design. It comprises three clients (client here is the sensor module in the room as shown in Figs. 2 and 3), network interface between client and website. Each client (client 1 as shown in Fig. 4, client 2 as shown in Fig. 5, and client 3 as shown if Fig. 6) comprises following hardware module:

(i) Sensor device from where all the input data is collected from the surroundings and are converted into proper digital signal for processing part in the controller. As per the device configuration.
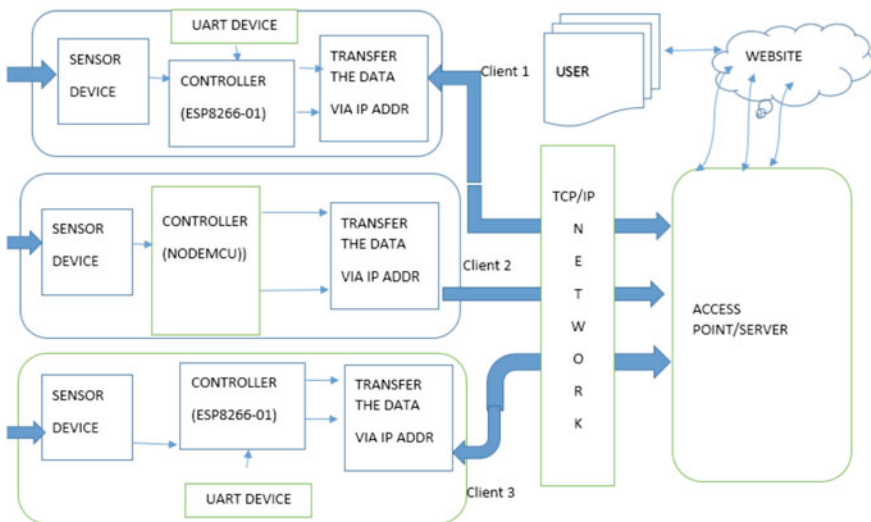
Fig. 1 Hardware system block diagram

Fig. 2 Dedicated Ip addresses of each client
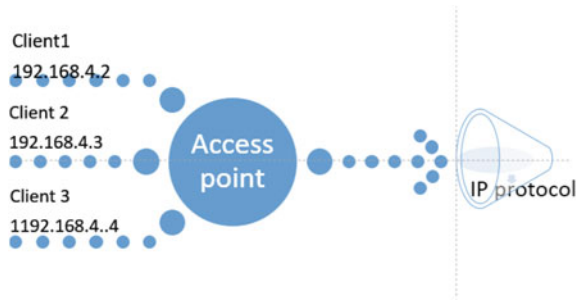
Fig. 3 Different clients in different areas of the house
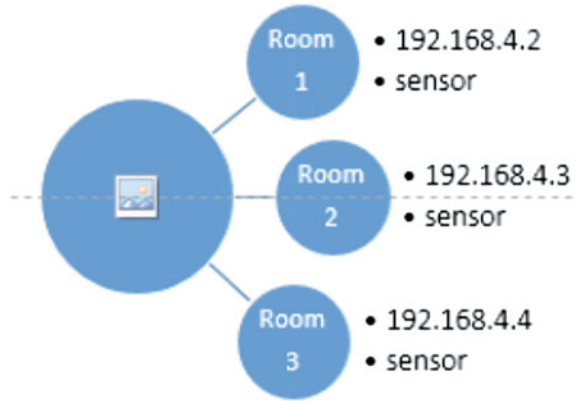
Fig. 4 Client 1 with IR sensor and Wi-Fi module in Room 1
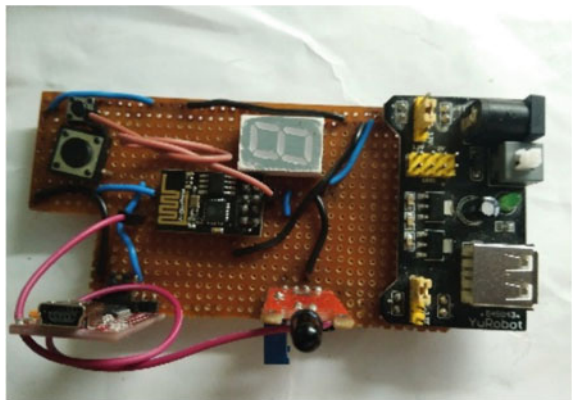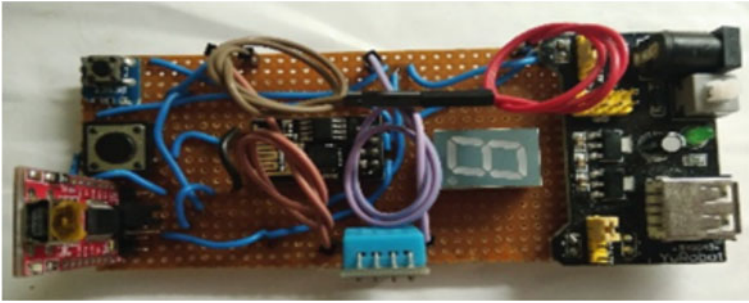
Fig. 5 Client 2 with DHT sensor and Wi-Fi module in Room 2
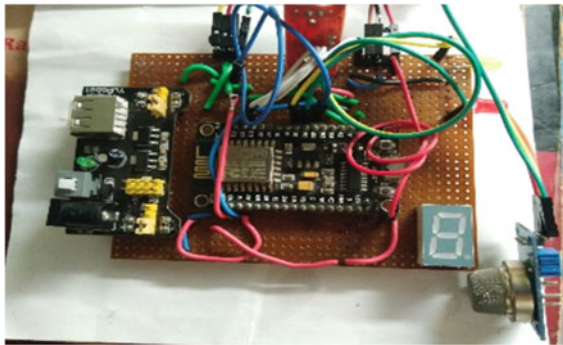
Fig. 6 Client 3 with gas sensor and Wi-Fi module in Room 3

 (ii) Controller portion, which plays the key role in the subsystem of client for processing the collected data for further transmission via transfer block. The controller that is used in it is ESP8266-01.
(iii) The transfer block is used for transmitting the processed data to network through IP address. Its transmission range is max 100 m.
(iv) UART device is used for uploading the code onto the controller.

2.1 Software System Design

The system model consists of three major module, i.e., network block, server block, and website. These network blocks allow all the three clients to get connected to the user at a time through single IP address. Server block is the key part of the system as it collects all the data which are collected by the individual client/subsystem and also it itself creates a protected network onto which other subsystem can get attached to it through IP address. Also, the HTML code that is used in creating webpage in the website is coded into it. It is also named as an access point since server gives the

Fig. 7 Different client data in the website
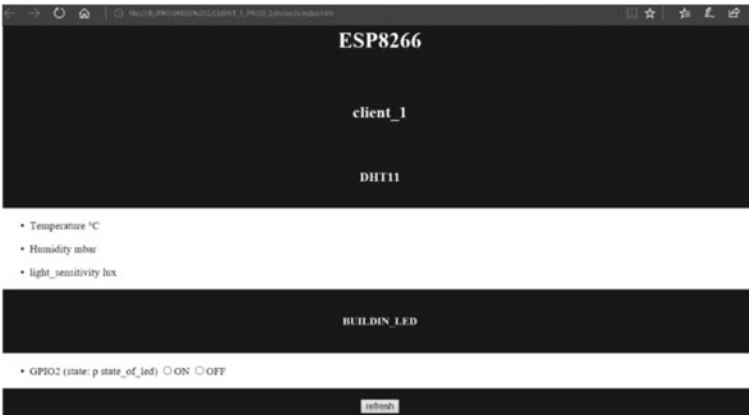
Fig. 8 Client 2 data in the website

access point for the website to the clients. Finally, in the website, all collected data gets displayed and can be accessed by simply connecting the server Wi-Fi network and then writing the server IP address onto the device browser, which is used by the user for surveying the collected data. For creating website basic HTML is used with CSS code for designing the website.

Figure 7 shows different client data available in the website. After security verification each client data will be displayed. Figures 8 and 9 show the data for client 2 (humidity, temperature, and light illumination) and client 3 in the website.

Fig. 9 Client 3 data in the website

3 Conclusion

We have implemented home surveillance system which can be accessed locally as well as through the Internet from any part of the world. We designed a standalone system which consists of sensors along with wireless connectivity. This standalone system was named as clients and put in different parts of the house (room1, room2, etc. along with dedicated IP addresses). We collected all the data from different clients (temperature, humidity, light illumination, and gas sensor) and uploaded to the website with proper user credentials.

References

1. Rose K, Eldridge S, Chapin L (2015) Internet of things: an overview. The internet society (ISOC)
2. Benammar M, Abdaoui A, Ahmad SHM, Touati F, Kadri A (2018) A modular IoT platform for real-time indoor air quality monitoring. Accessed 14 Feb 2018
3. Luo Q, Zhao J, Chen M (2011) Research of TCP/IP protocol stack based on embedded system. Computer research and development (ICCRD). Accessed 05 May 2011
4. Wong C-M, Hsu W-P (2014) Study on cognitive radio in IEEE 802.15.4 wireless sensor networks. Internet of things (WF-IoT), IEEE World Forum. Accessed 24 Apr 2014
5. Kanase P, Gaikwad, S (2016) Smart hospitals using internet of things (IoT). Int Res J Eng Technol 3(3) March
6. Razzaq MA, Gill SH, Qureshi MA, Ullah S (2017) Security issues in the internet of things (IOT): a comprehensive study. Int J Adv Comput Sci Appl 8(6)
7. Manjunatha chari S Dr., Sivakumar B (2017) Development of smart network using WSN and IoT for precision agriculture monitoring system on cloud. Int Res J Eng Technol (IRJET) 4(5) May
8. Sfikas G, Akasiadis C, Spyrou E (2016) Creating a smart room using an IoT approach. In: Conference: AI-IoT workshop, in conjunction with SETN. Accessed 09 June 2016
9. Hunkeler U, Truong HL, Stanford-Clark A (2008) MQTT protocol for wireless sensor networks. In: 3rd international conference on communication systems software and middleware and workshops, COMSWARE 2008. Accessed 27 June 2008
10. Chen J, Cheng W (2016) Analysis of web traffic based on HTTP protocol. In: 2016 24th International Conference on software, telecommunications and computer networks (SoftCOM). Accessed 08 Dec 2016

Author Index