Lecture Notes in Electrical Engineering 692

Vijay Nath
J. K. Mandal *Editors*

Nanoelectronics, Circuits and Communication Systems

Proceeding of NCCS 2019

Springer

Lecture Notes in Electrical Engineering

Volume 692

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China

Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India, Japan, Rest of Asia

Swati Meherishi, Executive Editor (Swati.Meherishi@springer.com)

Southeast Asia, Australia, New Zealand

Ramesh Nath Premnath, Editor (ramesh.premnath@springernature.com)

USA, Canada:

Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:

Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)

** **Indexing: Indexed by Scopus.** **

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

Vijay Nath · J. K. Mandal
Editors

Nanoelectronics, Circuits and Communication Systems

Proceeding of NCCS 2019

Springer

Editors
Vijay Nath
VLSI Design Group, Department of
Electronics and Communication Engineering
Birla Institute of Technology
Ranchi, India

J. K. Mandal
Department of Computer Science
and Engineering
Kalyani University
Kolkata, India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-15-7485-6 ISBN 978-981-15-7486-3 (eBook)
https://doi.org/10.1007/978-981-15-7486-3

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This volume contains the papers that were presented at the 5th International Conference on **Nanoelectronics Circuits & Communication System (NCCS-2019)** organized by Indian Society for VLSI Education (ISVE) & IETE Ranchi Centre with IETE Zonal Seminar & ISF Congress with support of BSNL at ARTTC BSNL Ranchi during 9–10th Nov. 2019. It provided a great platform for researchers from across the world to report, deliberate, and review the latest progress in the cutting-edge research pertaining to smart computing and its applications to various engineering fields. The response to NCCS-2019 was overwhelming with a good number of submissions from different areas relating to, artificial intelligence, machine learning, internet of things, VLSI design, micro & nanoelectronics circuits &devices, green energy, signal processing, smart computing, computational intelligence, and its applications in main tracks. After a rigorous peer-review process with the help of program committee members and external reviewers, only quality papers were accepted for publication in this volume of LNEE series of Springer. Several special sessions were offered by eminent professors in many cutting-edge technologies. Several eminent researchers and academicians delivered talks addressing the participants in their respective field of proficiency. We thank to Prof. P. S. Neelakanta, Florida Atlantic University, USA; Prof. Gopal Pathak, Vice Chancellor, Jharkhand Technical University, Ranchi; Prof. Nand Kunar Yadav 'Indu', Vice Chancellor, Central University of Jharkhand, Ranchi; Sh. Anurag Dubey, Xilinx, USA; Prof. Abhijit Biwas, University of Kolkata, Kolkata; Prof. A. A. Khan, Former VC, Ranchi University, Ranchi; Sh. K.K. Thakur, CGMT, BSNL, Ranchi; Dr. Raj Kumar Singh, Ranchi University, Ranchi; Sh. Ajay Kumar, Governing Council Member, IETE, New Delhi; Sh. Viney Kakkar, Treasurer, IETE, New Delhi; Dr. P.R. Thakura, BIT Mesra, Ranchi; for their valuable talks for the benefits of the participants. We would like to express our appreciation to the members of the program committee for their support and cooperation in this publication. We are also thankful to the team from Springer for providing a meticulous service for the timely production of this volume.

Our heartfelt thanks to our loving founder Dr. J.W. Bakal, President, IETE, New Delhi, and Sh. K.K. Thakur, Chairman, IETE Ranchi Centre, Ranchi; Prof. P.R. Thakura, President, ISVE, Ranchi and Executive Committee of IETE & ISVE Ranchi for extending excellent support to host this in ARTTC, BSNL Campus. Professor P. S. Neelakanta, Florida Atlantic University, USA; deserves a big round of applause from all of us for his continuous guidance and support from the beginning of the conference. Without his support, we could never have executed such a mega event.

Special thanks to all guests who have honoured us in their presence in the inaugural-day of the conference. We thank to all special session chairs, track managers, and reviewers for their excellent support. Last but not least, our special thanks go to all the authors who submitted papers and all the attendees for their contributions and fruitful discussions that made this conference a great success.

Ranchi (JH), India Vijay Nath

Contents

Layout: T1_Standard Book ID: 499559_1_En Book ISBN: 978-981-15-7485-6
Chapter No.: FM 1 Date: 9-6-2021 Time: 3:59 pm Page: 11/13

About the Editors

Dr. Vijay Nath received his Bachelor's degree in Physics and Master's degree in Electronics from DDU Gorakhpur University, India, in 1998 and 2001, respectively. He received his PGDCN (GM) from MMMUT Gorakhpur in 1999 and Ph.D. degree in VLSI Design & Technology from Dr. RML Avadh University Ayodhya in association with CEERI- Pilani in 2008. From 2000 to 2001 he was project trainee in the IC Design Group at CEERI-Pilani. He is Associate Professor in the Dept. of ECE, BIT Mesra Ranchi (JH), India. He has joined the Institute in 2006. His research interests include micro & nano-electronics, analog & digital VLSI design, ASICs, embedded system design, Internet of things, & machine learning. He has published 175 papers in respected Scopus & SCI journals and at conferences. He has successfully completed two R&D projects funded by DST New Delhi with DRDL Hyderabad and MHRD New Delhi, and a third project, funded by RESPOND SAC ISRO Ahmadabad, is ongoing. He is the editor of Nanoelectronics, Circuits &Communication Systems, Proceeding of NCCS-2015 & NCCS-2017, and of Microelectronics, Computing & Communication Systems, Proceedings of MCCS-2015, MCCS-2017 & MCCS-2018 in a Scopus book series: LNEE, Springer. He is a member of several professional societies and academic bodies, including IETE, ISTE, ISVE and IEEE.

Prof. J. K. Mandal received his M.Sc. in Physics from Jadavpur University in 1986 and M.Tech. in Computer Science from the University of Calcutta, and was awarded a Ph.D. in Computer Science and Engineering by Jadavpur University in 2000. Currently, he is a Professor of Computer Science and Engineering and was Dean of the Faculty of Engineering, Technology and Management, Kalyani University, West Bengal for two consecutive terms, having started his career as a lecturer at NERIST, Arunachal Pradesh in 1988. His areas of research include coding theory, data and network security, remote sensing and GIS based applications, data compression, error correction, visual cryptography, steganography, security in MANET, wireless networks and unify computing. He has published than 30 proceedings volumes, 6 books, 160 journal articles and 300 other articles. He is one of the editors of the Springer AISC and CCIS Series.

Active Security by Implementing Intrusion Detection and Facial Recognition

Ravi Prakash and Premkumar Chithaluru

Abstract The current situation of security cameras has shown that there is a large scope for improvement in the way they operate from a fundamental level. Security cameras have always been used as a monitoring system but not as an intrusion detecting and notifying system. CCTVs are ineffective when they are used just to monitor a room, but if we can implement facial recognition using the existing system, we can add a level of extra security to our home or office spaces. This paper will analyse the use of intruder detection in IoT, mainly CCTV or closed-circuit television using face detection while going in depth into how this can be achieved and what measures are taken to make this possible. This research will provide a deeper look into the problem of CCTVs being a passive form of security and how to make them a more active form of security by implementing intruder detection. Face detection using OpenCV is used to achieve the functionality of intruder detection. This information will impact the way CCTV cameras are made in the future and will demonstrate that the extra security layer of intrusion detection is a feasible and ready-to-use product.

Keywords Face · Detection · OpenCV · Intruder · CCTV · Security

1 Introduction

As we are getting busier in our daily lives, we often leave our homes and offices prone to robbery without implementing any protection [1]. Today, almost all technologies have seen advancements except physical security. Most people still use the old methods of securing one's property by employing security guards or using monitoring tools like closed-circuit televisions (CCTVs) cameras which can only help monitoring [2, 3]. Most physical security methods that are implemented today are

R. Prakash (✉) · P. Chithaluru
School of Computer Science, UPES Bidholi, Dehradun, Uttarakhand, India
e-mail: ravishahi71@gmail.com

P. Chithaluru
e-mail: chpremkumar@ddn.upes.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_1

very old and ineffective. While every device is getting smarter, CCTVs that are used to monitor the activity in our work space or home have not gotten any smarter, and one still has to go through the whole footage to find important data [4]. They do not have any systematic way of securing the space as they just monitor and do not interact with the authorized user in the case of certain crisis. They do not give any notifications to the authorized user at the time of the incident [5].

As far as human beings are concerned, face recognition is an easy task, but for computers, it is quite difficult. We give them commands that they execute. So, in order for computers to detect faces, we need to give them data, allow them to learn different faces. This task can be accomplished using concepts like artificial intelligence—to recognize facial patterns and deep learning—allowing them to learn different faces. The problem arises in choosing which features to base the algorithm on, the inner features like the nose, mouth and eyes or the outer features like the head, hairline and shape [6]. To solve this problem, we need to extract different features of relevance from an image and put them into a suitable representation and classify them.

The computer vision and image processing software namely OpenCV is a commonly used library, with many different algorithms for image and video analysis. In today's era of information explosion, there is a need for such computational methods. Popular sites such as Facebook and Google also deploy it to identify individual persons in the data set of images uploaded by users so that you do not have to code each person multiple times in every single photograph. Willow Garage in 2008 designed the software OpenCV 2.3.1 which has coding interface to languages namely C, C++, Python and Android. BSD licence was adopted to release the software OpenCV which is quite user friendly library in commercial and academic products alike [7].

Built on the concept of Internet of things (IoT) and different electronic sensors and microcomputers, a new security system is being implemented. Internet of things (IoT) is an advanced automation and analytics system that allows things (devices, appliances, machines) to connect and share data while running relevant applications towards a normal user. The basic premise and goal of IoT are to "connect the unconnected". In this research work, we are designing an automation system which produces an alert when an illegal human intervention occurs [8–11].

2 Literature Review

(1) Steps for face recognition

i. Look at the image and then identify all faces in it.
ii. Analyse each face to comprehend if it is turned in a peculiar direction or in a bad lighting as it can be a known person.
iii. Highlight the distinctive features of that face which can be used to tell that face apart from others.

iv. Compare these distinctive features of the face to the ones you already recognize.

We will be using a procedure conceived in 2005 named Histogram of Oriented Gradients (HOG) to detect the facial patterns in an image [12]. We make our image black and white as colour data is not needed to identify faces. Then, we will look at a pixel and its immediate neighbours. Then, we draw arrows towards the darkest pixel showing where the image gets darker. This makes really bright and really dark images with exact same representation. Doing this for every pixel will require excess computation and makes the detection slower, so we will break the image into squares of 16 × 16 pixels each. We will then compute the gradients pointing in each major path and replace the square with the arrow where the paths were the strongest. This is called an HOG pattern [13]. Then, we just compare the HOG pattern to the ones extracted from other training faces. If the faces are turned into different directions, then the computer will think that they are two different faces even if they are the same person. To prevent this from happening, we will try to rotate each picture so that the eyes and lips are always in the same place in the graphic object [14]. We will be using face landmark estimation algorithm to do this task. We will arrive at some points called landmarks which are present in every face which will allow us to tell where the mouth and eyes are. Then, we scale, rotate and shear the image so that the eyes and mouths are centred as finest as possible. We cannot do any 3D transformations as that would present alterations [15].

(2) The program needs to be trained using a database of faces with multiple images for each individual for it to be effective.
(3) Since faces are so complex, there can be no single test that will allow us to identify a face. Instead, there are thousands of small patterns and features that need to be matched to find a face. OpenCV uses various machine learning algorithms to make it easier to recognize faces [16]. If our computer starts from the top left to the bottom, then at each block, our system will need to perform huge computations to decide if it is a face which will require huge computation power if we need it to be fast, so OpenCV uses cascades which simply means a series of waterfall [17–21]. OpenCV breaks the face detection process into multiple stages and performs a rough and quick test for each block. If a block passes a stage 1 test, a slightly detailed stage 2 test is performed [22]. This continues for all the stages as a block keeps passing stages. If a block passes all the stages, only then a face is recognized. The benefit of this method is most blocks get eliminated during the first few stages saving us a lot of computation power [23].

3 System Model

The proposed system describes a simple and easy hardware implementation of Raspberry Pi, PI Cam and PIR sensor. This system will use face detection and face recognition algorithms to compare and recognize individuals. Raspberry Pi is a powerful credit card-sized minicomputer. Here, PIR sensor is used for motion detection. When any individual will come near the implemented system, this sensor will activate PI Cam, and this camera will take a picture of that individual. There will be Python programming for face detection, generating trained data set for making database for home members and face recognition for comparing and recognizing the face. If there will be any mismatch, an email will be sent to the email id of the house members (Fig. 1).

4 Problem Definition

Cyberwar is one of the most popular and oldest attacks that deactivate financial and organizational systems by thieving or modifying classified data to undermine networks, Websites and services. Since the introduction of the first lock and key and thereafter RFID tags and biometric identification, anti-theft systems have progressed to cater to the needs of the society to resolve these information security attacks. A cheaper and advanced system is needed.

Algorithm

1. Capture image on motion detection.
2. Find all the faces in the image.
   - 2.1 Check and rectify any bad angles.
   - 2.2 Pick out the unique facial features which can be used to distinguish it from other faces.
   - 2.3 Compare the facial features to the ones already known using Eigen Faces.
     - 2.3.1 If intruder is detected
       - 2.3.1.1 Alert administrator.
     - 2.3.2 End if
3. Process starts again

5 Simulation Tools

Java Swings used to make graphical user interface (GUI) for planning Web applications and applets. Learning Swing with the NetBeans IDE is substantially speedier and less demanding approach to start working with Swing. NetBeans IDE's GUI developer, a capable component that to build GUI.

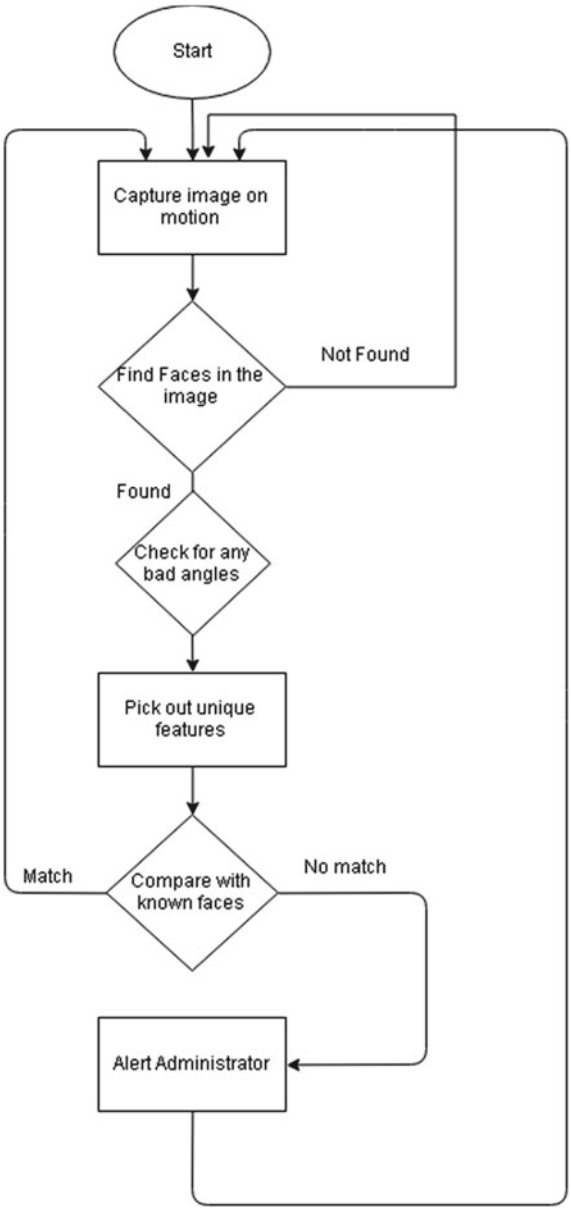


Fig. 1 Process for intruder detection

6 Conclusion

This article mainly focuses on detection and recognition of face from camera. The implementation of application, computing and execution is very simple. Detection

of face is always a challenge, especially when recognizing objects in far. So, we introduced this methodology to detection and as well as comparing things with close objects of the face from camera shoot. Moreover, collected faces comparing with our existed database under legitimately circumstances, position and depth of the picture, the system is additionally prone the errors. However, the area under surveillance basically has suitable light to the objects, and camera can be positioned anywhere. Our intelligent surveillance system can be adopted the area, professional bodies and delicate objects. To challenge this issue with far facial objects and depths of the picture, consider the more the one video camera to fix the all different places and occupied regions. Our system can be used to assist the all camera clips to be merged, for making the more effective system, robust and less error rate in picture quality. In addition, this idea can be merged with many smart algorithms to make it a more effective and error-free surveillance system.

7 Future Scope

The system can be extended and implemented to a closed-circuit television system for intruder detection and can also be extended to automated response in case of a intruder detection. The name of known person will be announced. Can also be used for automated responses of known, system can be utilized for car recognition and fugitive recognition.

References

1. Lwin HH, Khaing AS, Tun HM (2015) Automatic door access system using face recognition. Int J Sci Technol Res 4(06)
2. Udaya Kumar B, Murty DS, Phani Kumar CR (2013) Implementation of low cost ethernet based home security using wireless sensor network. J Algorithms Res
3. Karri V, Daniel Lim JS (2005) Method and device to communicate via SMS after a security intrusion. In: 1st international conference on sensing technology, Palmerston North, New Zealand, 21–23 Nov 2005
4. Priyadarshi R, Soni SK, Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-13-0776-8_27
5. Mahalakshmi R, Manjula M, Saranya S, Vaishnavi P, Shalini J, Arasa KR (2013) Intrusion detection by facial recognition using CCTV cameras with video management system. Int J Adv Electr Electron Eng 2(2):59–62
6. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
7. Brodsky T, Cohen R, Cohen-Solal E, Gutta S, Lyons D, Philomin V, Trajkovic M (2001) Visual Surveillance in retail stores and in the home. In: Advanced video-based surveillance systems, Chap 4. Kluwer Academic Publishers, Boston, pp 50–61. Accessed 2 May 13
8. Szeliski R (2011) Computer vision: algorithms and applications (texts in computer science), 1st edn. Springer, Berlin

9. Yuan X, Sun Z, Varol Y, Bebis G (2003). A distributed visual surveillance system. IEEE international conference on advanced video and signal based surveillance. Miami, July 2003
10. Lee DJ, Zhan P, Thomas A, Schoenberger R (2004). Shape-based Human Intrusion Detection, SPIE International Symposium on Defense and Security. Visual information processing XIII, vol 5438. Orlando, Florida, USA, pp 81–91, 12–16 Apr 2004
11. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
12. Shankar Kartik J, Ram Kumar K, Srimadhavan VS (2013) SMS alert and embedded network video monitoring terminal. Int J Secur Privacy Trust Manage IJSPTM 2
13. Rawat P, Chauhan S, Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circ Syst Comput. https://doi.org/10.1142/S0218126620502047
14. Khan SR, Mansur AA, Kabir A, Jaman S, Chowdhury N (2012) Design and implementation of low cost home security system using GSM network. Int J Sci Eng Res 3(3)
15. Bangali J, Shaligram A (2013) Design and implementation of security systems for smart home based on GSM technology. Int J Smart Home 7(6)
16. Mae Y, Sasao N, INNoue K, Arai T Person detection by mobile manipulator for monitoring. In: SICE 2003 annual conference, pp 2801–2806
17. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
18. Nebel JC, Lewandowski M, Thevenon J, Martinez F, Velastin SA (2011) Are current monocular computer vision systems for human action recognition suitable for visual surveillance applications? In: 8th international conference on information visualisation, LNCS 6938. Springer, Las Vegas, USA, 26–28 Sept, pp 290–299
19. Kosba AE, Saeed A, Youssef M, RASID (2012) A Robust WLAN device-free passive motion detection system. In: The tenth annual IEEE international conference on pervasive computing and communication (Per-Com 12), Mar 2012
20. Randheer Soni SK, Kumar S, Priyadarshi R (2020). Energy-aware clustering in wireless sensor networks. In Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_39
21. Kumar S, Soni SK, Randheer, Priyadarshi R (2020). Performance analysis of novel energy aware routing in wireless sensor network. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_44
22. Tuscano C, Lopes B, Machado S (2013) Smart web cam motion detection surveillance system. Int J Mod Eng Res 3(2):1169–1171
23. Rani A, Hafiizh MM (2007) Active infrared motion detector for house security system. Eng D thesis, University Malaysia Pahang

Computing Density of States for Pöschl–Teller Potential in Double Quantum Well Structure

Arpan Deyasi, Suporna Bhowmick, and Pampa Debnath

Abstract Density of states for Pöschl–Teller potential in double quantum well structure is analytically computed in the presence of external electric bias applied along quantized direction. Realistic band diagram is considered by assuming nonparabolic band along with variation of well-designed parameters. Simulated findings are compared with ideal rectangular profile, which speaks for lower electron concentration for the present work, thus suitable for better current control. Dimensional parameters are tuned to observe the variation for lowest three eigenstates, along with material parameters. Simulated data will help to calculate the current if proposed for quantum transistor design.

Keywords Density of states · Pöschl–Teller potential · Depth parameter · Width parameter · Eigenstates · Dimensional variation

1 Introduction

Electronic application of quantum well structure is a well-known subject for the last thirty years, which is revolutionized only when complex models are presented for specific applications [1–3]. Several new novel devices are theoretically proposed [4, 5] and even experimentally realized [6, 7] also based on the arbitrary quantum well, wire or dot, where key focus is made to tune the eigenstates as per the application requirement. Shape of the potential is designed with respect to that condition and is realized by fabricating with heterostructure layers, and dimension of each layer

A. Deyasi (✉) · S. Bhowmick · P. Debnath
Department of Electronics and Communication Engineering, RCC Institute of Information Technology, Kolkata, India
e-mail: deyasi_arpan@yahoo.co.in

S. Bhowmick
e-mail: bhowmicksuporna@gmail.com

P. Debnath
e-mail: poonam.4feb@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_2

becomes critical in order to develop quantum confinement effect. Accurate position of eigenstates can be obtained either from transmission coefficient calculation or from density of states profile. Among them, later is more acceptable to theoretical researchers as it is required for calculation of current.

Different geometries of quantum well [8, 9] is proposed where rectangular is the easiest to compute owing to its ideal nature. Though specific geometries like triangular, parabolic, Gaussian are well-established; but none of them are suitable to work at lower temperature. In this context, Pöschl–Teller potential is the most desirable one, as it successfully exhibits the Bose–Einstein condensation of K atoms [10]. Due to this feature, the structure is now considered as the subject of interest among other potential patterns and thus taken as the choice for the present work. In this paper, authors simulated the potential inside a double quantum well triple barrier (DQWTB) structure, and density of states is calculated under different biasing conditions. Results are plotted with energy, and peak of the profile at different energy states can be utilized for current calculation.

2 Mathematical Formulation

Motion of any electron inside a quantum-confined structure is calculated by using Schrödinger's time-independent equation in the absence of external excitation:

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left[\frac{1}{m^*(z)}\frac{\partial}{\partial z}\psi(z)\right]+V(z)\psi(z)=E(z)\psi(z) \tag{1}$$

where, we have incorporated the concept of effective mass mismatch.
In presence field, Eq. (1) is modified as

$$-\frac{\hbar^2}{2}\frac{\partial}{\partial z}\left[\frac{1}{m^*(z)}\frac{\partial}{\partial z}\psi(z)\right]+V(z)\psi(z)-q\xi(z)\psi(z)=E(z)\psi(z) \tag{2}$$

Density of states gives the number of energy eigenstates per unit energy interval, and it depends on the band curvature of the semiconductor. For a quantum well system at any particular energy E_i, DOS can be computed by summing over all the subbands below the energy level.

$$\rho(E)=g_s g_v\sum_{i=1}^{n}\frac{m*}{\pi\hbar^2}\Theta(E-E_i) \tag{3}$$

where Θ is the Heavyside step function. Here, summation is performed over n confined states, g_s and g_v are spin and valley degeneracy, respectively.

3 Results and Discussions

Using Eq. (2), density of states is computed and plotted for Pöschl–Teller potential configuration of the double quantum well system in the presence of electric field applied along the direction of dimensional confinement. Here, we consider those Eigen energies, obtained when band nonparabolicity is taken into account. The result is compared with rectangular potential well for validation of the result. The profile shown in Fig. 1 represents the comparative study of density of states considering rectangular well geometry and Pöschl–Teller potential well geometry.

Figure 2 shows density of states of Pöschl–Teller potential for three different well dimensions. It is seen that with increase of well width, more energy levels are formed. Also, the gap between two consecutive enrgy states reduces when the well

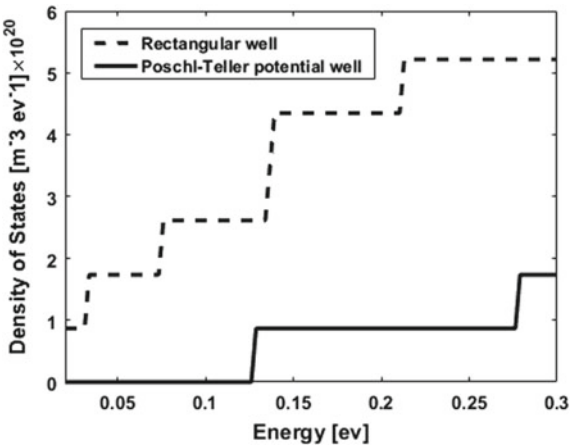


Fig. 1 Comparative study of density of states for Pöschl–Teller potential with rectangular potential

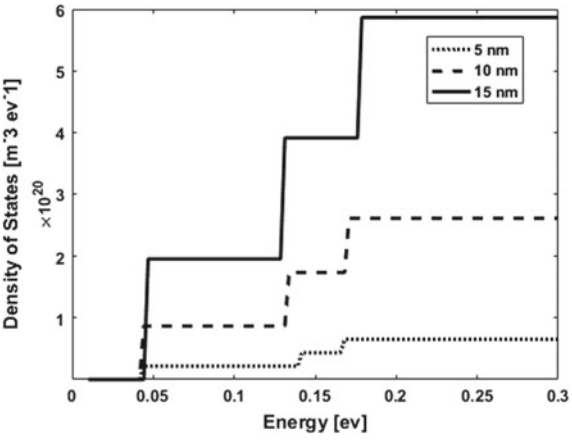


Fig. 2 Density of states of Pöschl–Teller potential for three different well dimensions

width decreases. This is due to the fact that with increase of well width, quantum confinement decreases, which makes the energy seperation higher.

Similar simulation is carried out for the contact barrier and middle barrier width, which is represented in Figs. 3 and 4 simultaneously. The energy separation between two layers does not depend on the width of the contanct barrier and middle barrier width (Fig. 5).

By varying the material composition of barrier layers, it is observed that quantum states appear in higher energy values with increase in mole fraction. The result is shown in Fig. 6. This is because with increase of x, mismatch of effective mass increases as well as the conduction band discontinuity. This enhances the quantum confinement. Hence, eigenenrgy increases. This is reflected via density of states plot. The variation in depth and width parameter is shown in Figs. 6 and 7 simultaneously.

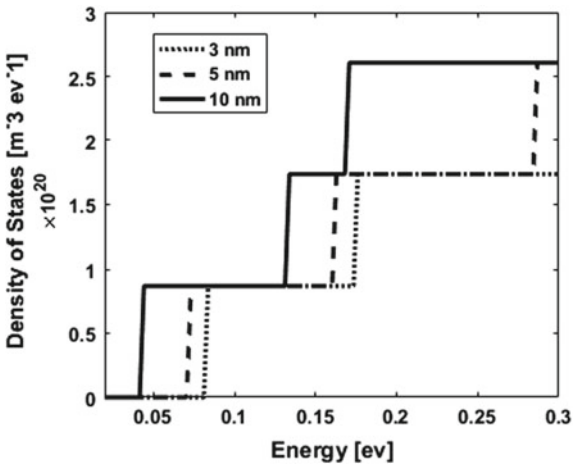


Fig. 3 Density of states of Pöschl–Teller potential for three different contact barrier widths

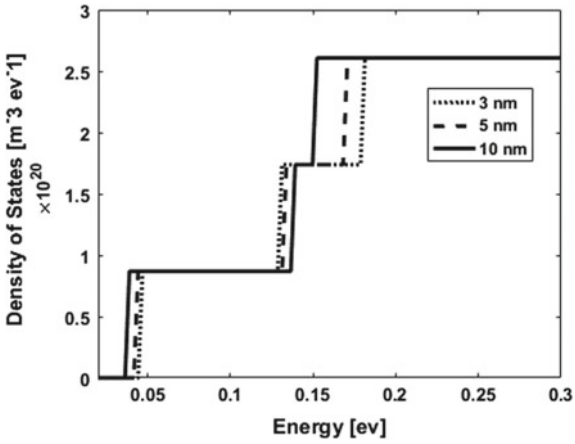


Fig. 4 Density of states of Pöschl–Teller potential for three different middle barrier widths

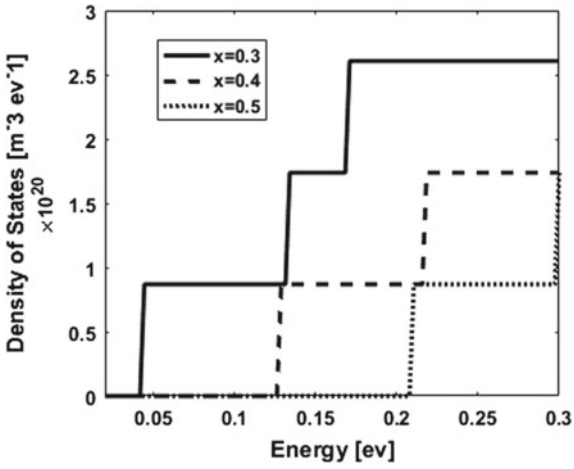


Fig. 5 Density of states of Pöschl–Teller potential for three different barrier material compositions

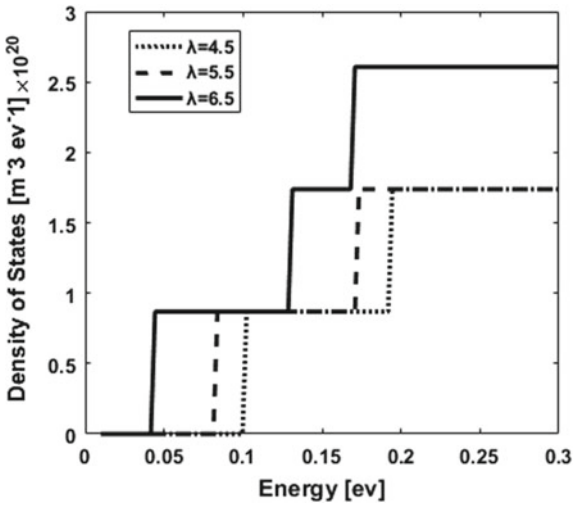


Fig. 6 Density of states of Pöschl–Teller potential for three different depth parameters

Figure 7 indicates the biasing effect on density of states. When field is applied along the direction of confinement, the structure is inclined along that direction which results lowering of quantum states. This causes more possible eigenvalues with reduced separation (Fig 8; Tables 1 and 2).

4 Conclusion

Density of states for Pöschl–Teller potential inside a double quantum well triple barrier structure is analytically simulated in the presence of low electric field. Results

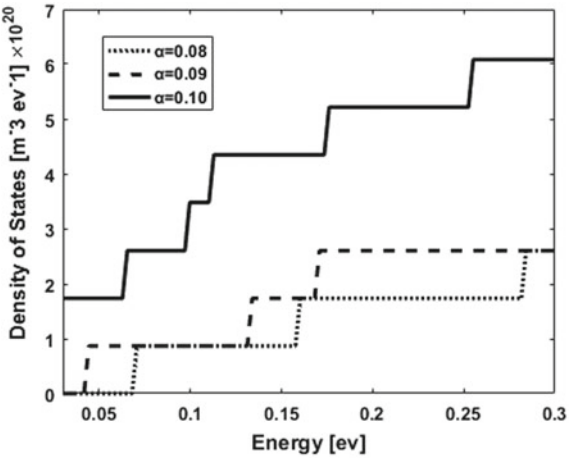


Fig. 7 Density of states of Pöschl–Teller potential for three different width parameters

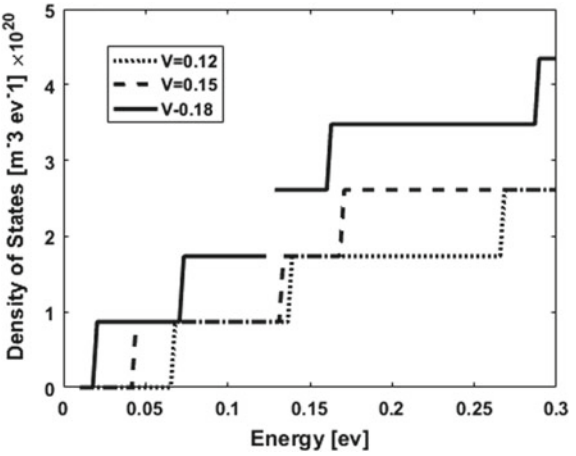


Fig. 8 Density of states of Pöschl–Teller potential for three different electric fields

Table 1 Lowermost two eigenstates for variation in well dimensions

Name of parameter	Dimension (nm)	$E1$ (eV)	$E2$ (eV)	$\Delta E21$ (eV)
Well width	5	0.04203	0.14	0.09797
	10	0.04253	0.132	0.08947
	15	0.04453	0.13	0.08547
Contact barrier width	3	0.08303	0.174	0.09097
	5	0.07153	0.162	0.09047
	10	0.04253	0.1325	0.08997
Middle barrier width	3	0.04503	0.1295	0.08447
	5	0.04253	0.1325	0.08997
	10	0.03753	0.1375	0.09997

Table 2 Lowermost two eigenstates for variation in width and depth parameters

Name of parameter	Dimension (nm)	E_1 (eV)	E_2 (eV)	ΔE_{21} (eV)
Width parameter	0.08	0.07053	0.1595	0.08897
	0.09	0.04253	0.132	0.08947
	0.1	0.007031	0.09703	0.08999
Depth parameter	4.5	0.01005	0.1905	0.18045
	5.5	0.08203	0.1705	0.08847
	6.5	0.04253	0.1325	0.08997

are compared with rectangular profile, and it is found that less number of eigen peaks with higher separation. This helps to apply the structure for short-wavelength photodetector design.

References

1. Ma Y, Liu L, Yu Z, Li Z (2000) Validity and applicability of triangular potential well approximation in modeling of MOS structure inversion and accumulation layer. IEEE Trans Electron Devices 47(9):1764–1767
2. Sahoo N, Panda AK, Sahu T (2014) Effect of intersubband interaction on multisubband electron mobility in a parabolic quantum well under applied electric field. Phys Semicond Devices: 231–234
3. Kumar S, Biswas D (2007) Effects of a Gaussian size distribution on the absorption spectra of III-V semiconductor quantum dots. J Appl Phys 102:084305
4. Wan Y, Inoue D, Jung D, Norman JC, Shang C, Gossard AC, Bowers JE (2018) Directly modulated quantum dot lasers on silicon with a milliampere threshold and high temperature stability. Photonics Res 6(8):776–781
5. Alves RA, Costa JC, Gomes M, Silva NA, Guerreiro A (2017) Quantum wires as sensors of the electric field: a model into quantum plasmonics. In: 25th optical fiber sensors conference
6. Lu F, Bhattacharya I, Sun H, Tran TTD, Ng KW, Malheiros-Silveria GN, Chang-Hasnain C (2017) Nanopillar quantum well lasers directly grown on silicon and emitting at silicon-transparent wavelengths. Optica 4(7):717–723
7. Dong M, Winful HG, Cundiff ST (2018) Quantum-well diode lasers for frequency comb generation. In: Frontiers in optics/laser science, OSA technical digest, paper LW6F.2
8. Kaatuzian H, Kojori HS, Zandi A, Kohandani R (2013) Effects of quantum well size alteration on excitonic population oscillation slow light devices properties. Opt Photon J 3:298–304
9. Sarkar D, Deyasi A (2016) Comparative analysis of absorption coefficient for parabolic and gaussian quantum wells for photodetector application. Adv Ind Eng Manage 5(2):197–201
10. Sarath R, Vinodkumar PC (2015) Bose-Einstein condensation in generalized Pöschl-Teller potential. PRAMANA J Phys 85(1):77–89

Image Processing Using Median Filtering for Identification of Leaf Disease

Smita Desai and Rajendra Kanphade

Abstract Agriculture plays a crucial role in India's financial growth. Today, India's position is second in farm output. Disease detection of plants assumes a significant job because of this reason. The productivity of the farm is affected by plant diseases. Different plant diseases are resulting to decrease in the quality and amount of plants. Presently, a day's study of plant disease means the study of patterns that we can see on the plants. It is hard to check the plant disease physically, so it requires a great measure of hard works, particularly in plant disease as well as excessive processing time. This paper shows a calculation for picture division strategy as it is the technique for change of computerized picture into a few subdivisions and representation of a picture into something for simpler examination. The novelty of this paper lies in designing the median filtering method to denoise the raw leaf disease images. The outcome of this paper shows the simulation results using MATLAB using the performance metrics such as peak-to-signal noise ratio (PSNR), mutual information, and mean square error (MSE).

Keywords Image processing · Noise removal · Median filtering · Leaf images · PSNR · MSE

1 Introduction

The Indian food business has enormous development, expanding its commitment to world food exchange each year because of its monstrous forthcoming Fort Worth expansion, especially with the food handling industry [1]. The Indian food and staple

S. Desai (✉)
DYPIET, Pimpri, Pune 411018, India
e-mail: smitardesai35@gmail.com

R. Kanphade
JSCOE, Hadpsar, Pune 411028, India

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_3

market is the world's biggest market, and it has 6th spot. Agriculture plays extraordinary reaction in India's fares, and it is around six percent complete modern speculation. Agriculture isn't just to feed regularly growing open yet its likewise significant wellspring of power. Plant maladies influence both quality and amount of plants in cultivating generation. Plant malady finding is exceptionally basic in prior stage so as to stop and control them [2]. The unaided eye perception of specialists is the fundamental methodology embraced in for location and ID of plant infections. Be that as it may, manual perception is tedious, costly and takes time. In certain nations, counseling specialists even cost high just as it is tedious as well.

There are different strategies to identify the plant malady, for example, thresholding, region growing, clustering, edge-based discovery, and so on. In plants, some basic ailments are singe, and others are contagious, viral, and bacterial infections. Picture preparing is the method which is utilized to quantify infection influenced part and to choose the uniqueness in the shade of the influenced region. Picture division is the path toward isolating or gathering a picture into various parts. There are at present a wide range of methods for performing picture division, going from the straightforward thresholding strategy to cutting edge shading picture division methods [3, 4].

Picture division is the route toward dividing a computerized picture into various fragments. The goal of division is to make less complex as well as change the depiction of a picture into something that is progressively significant and less difficult to investigate [5]. Picture division is regularly used to discover articles and points of confinement (lines, twists, etc.) in pictures. Specifically, picture division is the route toward delegating a name to each pixel in a picture to such an extent, that pixels with a comparable imprint share certain individuality [6]. In this paper, we presented the design of median filtering to remove the leaf images noise effectively with minimum computational efforts. Section 2 presents the brief study of some preprocessing methods. Section 3 presents the median filtering design, and how we apply it on input leaf images. The experimental outcomes were described in Sect. 5. Based on results, conclusion and further work were elaborated in Sect. 5.

2 Literature Survey

In this section, we discussed the special methods proposed for leaf disease detection published between 2011 and 2016.

Kulkarni [1] presented system for ahead of schedule and precisely plant ailments location, utilizing artificial neural network (ANN), and differing picture handling procedures. As the proposed methodology depends on ANN classifier for order and Gabor channel for highlight extraction, it gives better outcomes with an acknowledgment pace of up to 91%.

Singh et al. [7] presented technique for color expectation of paddy harvest leaf utilizing picture handling. They expected the strategy which might be utilized to

contrast the harvest leaf shading and LCC. They introduced this approach for information related to the plant necessity prior to receive the yield influenced. By utilizing picture handling innovation, a straightforward and hearty strategy for the shading expectation of paddy yield plant has been talked about alongside the scientific demonstrating which may give a staggering stage with notice bodies in the agribusiness field for the atomization of the harvest medical issues and arrangements.

Mondal et al. [8] introduced a productive procedure to distinguish and characterize the nearness of yellow vein mosaic virus (YVMV) sickness in okra leaf with the joint utilization of picture getting ready, K-implies, and credulous Bayesian classifier. The proposed framework was examined 79 standards wiped out and non-contaminated okra leaf pictures. The data leaf pictures were of four classes such as resistive (R), tolerable (T), moderately susceptible (MS), and highly susceptible (HS) based on the YVMV disease reality.

Muhammad Thaqif Container et al. [9] proposed illness identification framework for orchid utilizing picture preparing and fuzzy logic. The fundamental objective of researcher in this paper was to exhibit framework that can recognize an orchid disease by preparing its leaf picture. The leaf picture handling utilized strategies resemble gray scaling, limit division, and commotion expelling. Checking of well-being and ailment on plant accept a significant activity in effective development of harvests in the homestead. In early days, the observing and investigation of plant sicknesses were done physically by the ability individual in that field. This requires gigantic proportion of work and furthermore requires over the top preparing time. Some recent works reported for the image preprocessing in [9–11].

3 Median Filter

In this section, we present the design of median filtering to estimate and remove the noise from the 2D images. The strategy designed by considering the raw RGB leaf images. Median filtering is strategy which isn't straight, and it is utilized to expel clamor from pictures. It is extremely compelling at evacuating clamor while saving edges. The center channel works by going through the picture pixel by pixel, displacing each a motivation with the center advantage of neighboring pixels. The center is determined by first organizing all the pixel regards from the window into numerical solicitation, and after that overriding the pixel being considered within (center) pixel regard [6]. It is especially compelling at evacuating 'salt-and-pepper'-type commotion. The case of neighbors is known as the 'window,' which slides, pixel by pixel, over the whole picture.

In case of channel, a window size of three is utilized, with one section promptly going before and following every passage. Window for x [6] y [6].

X=3	3	9	4	52	3	8	6	2	2	9

Y [1] = median [3 3 9] = 3
Y [2] = median [3 4 9] = 4
Y [3] = median [4 9 52] = 9
Y [4] = median [3 4 52] = 4
Y [5] = median [3 8 52] = 8
Y [6] = median [3 6 8] = 6
Y [7] = median [2 6 8] = 6
Y [9] = median [2 2 9] = 2
Y [10] = median [2 9 9] = 9

y =

3	4	9	4	8	6	6	2	2	9

- For y [1] and y [9], extend the farthest leftmost or rightmost an incentive outside the purposes of confinement of the picture same as leaving leftmost or rightmost worth unaltered after 1D middle.
- In the above model, in light of the fact that there is no passage going before the main value is rehashed (just like the last worth) to get enough segments to fill the window.
- There are different methodologies that have various properties that may be favored specifically conditions:
- Avoid preparing the points of confinement, with or without editing the sign or picture limit sometime later.
- Fetching passages from better puts in the sign. With pictures for instance, territories from the far level or vertical farthest point may be picked.
- Shrinking the window close to limits, with the objective that each window is full.

As the median filtered designed in this paper is applied on 2D images, in experiment we first select the RGB leaf image and extracts its R, G, and B channels. We applied the median filtering on R, G, B channel of input image. The preprocessed channels are finally used to generate the preprocessed RGB image.

4 Simulation Results

Here, utilizing MATLAB we have changed over original picture into median filtering image. The picture of leaf as an information picture is converted into sifted pictures. We have tried around 30 examples. Figure 1 demonstrates the first pictures and sifted pictures utilizing median channel. Likewise, we have attempting to discover properties of pictures, that is, mutual information, mean square error, peak sign to noise ratio which will be utilized for choosing legitimate component extraction technique. Utilizing this technique, we will have the option to recognize the disease in the leaf (Figs. 1, 2, 3, 4, and 5).

For each image, we measured the mutual information, PSNR, and MSE as shown in Table 1. The results are very promising.

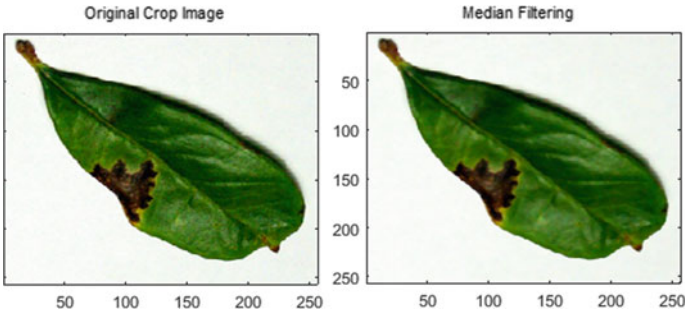


Fig. 1 Median filtering on image 1

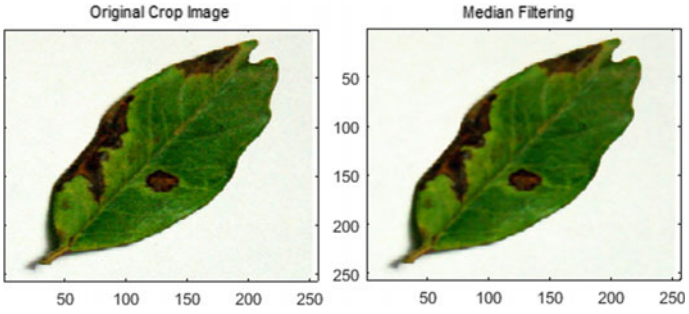


Fig. 2 Median filtering on image 2

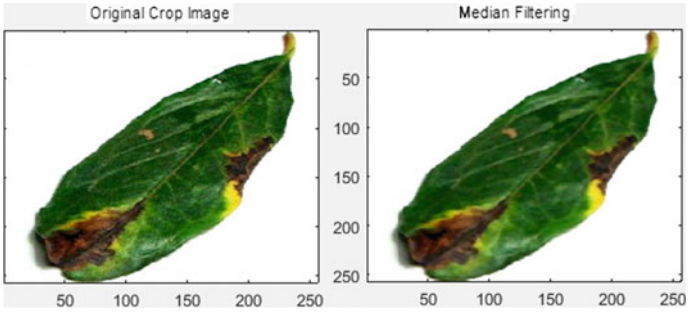


Fig. 3 Median filtering on image 3

5 Conclusion

This paper presents the design of median filtering to denoise the input leaf images effectively. The simulation results show that the proposed approach of this paper effectively improves the performance of denoising. In this paper, mutual information of various leaf images has been calculated, along with their MSE and PSNR number.

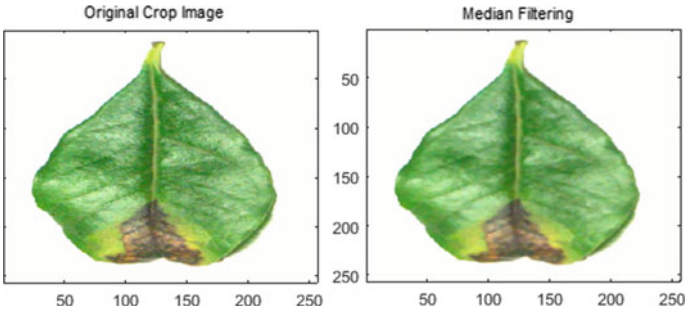


Fig. 4 Median filtering on image 4

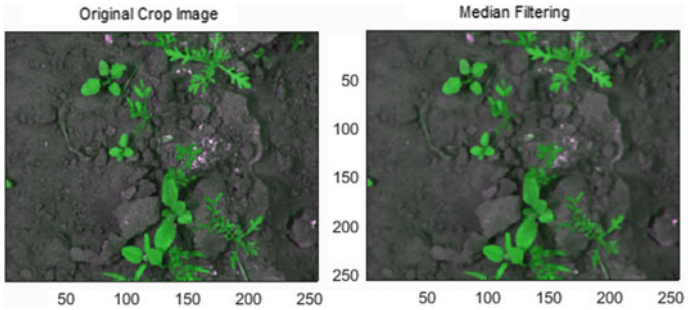


Fig. 5 Median filtering on image 5

Table 1 Denoising performance analysis

Image No.	Mutual information	MSE	PSNR
Image 1	2.1239	16.1872	108.204
Image 2	1.88945	16.6805	107.953
Image 3	1.70059	27.9607	101
Image 4	1.54687	28.7391	100.805
Image 5	1.63303	33.9494	97.8697

For further work, we suggest to work on image segmentation, features extraction, and classification phases of leaf disease detection system.

References

1. Kulkarni AH, Patil A (2012) Applying image processing technique to detect plant diseases. Int J Mod Eng Res 2(5): 3661–3664
2. Bashir S, Sharma N (2012) Remote area plant disease detection using image processing. IOSR J Electron Commun Eng **2**(6): 31–34. ISSN: 2278-2834

3. Rathod AN, Tanawal B, Shah V (2013) Image processing techniques for detection of leaf disease. Int J Adv Res Comput Sci Softw Eng 3(11)
4. Beucher S, Meyer F (1993) The morphological approach to segmentation: The watershed transform. In: Dougherty ER (ed) Mathematical morphology image processing, vol 12, pp 433–481. New York Marcel Dekke
5. Arivazhagan S, Newlin Shebiah R, Ananthi S, Vishnu Varthini S (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int CIGR J 15(1): 211–217
6. Mathew S (2016) A well organized approach for the detection of plant leaf disease using K-means clustering algorithm. IJSRD 4(10). ISSN: 2321-0613
7. Mondal D, Chakraborty A, Majumder DD, Kole DK (2015) Detection and classification technique of yellow veil amosaie virus disease in okra leaf images using leaf vein extraction and Naivy baysian classifier. Int Con Soft Comput Tech Implementations- (ICSCTI). Department of ECE, FET, MRIU, Faridabad, India, 8–10
8. Singh V, Varsha, Misra AK (2015) Detection of unhealthy region of plant leaf using image processing and genetic algorithm. Int Con Adv Comput Eng App (ICACEA) IMS Engineering College, Ghaziabad, India
9. Azmi MTBM (2013) Orchid disease detection using image processing and fuzzy logic, Int Con Electr, Electron Syst Eng
10. Pooja V, Das R, Kanchana V (2017) Identification of plant leaf diseases using image processing techniques. In: 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)
11. Alkinani MH, El-Sakka MR (Dece) Patch-based models and algorithms for image denoising: a comparative review between patch-based images denoising methods for additive noise reduction. EURASIP J Image Video Process 2017:58
12. Dhaygude SB, Kumbhar NP (2013) Agricultural plant leaf disease detection using image processing. Int J Adv Res Electr Electron Instrument Eng 2(1)

Cloud Computing: Security Issues and Challenges

Nirupam Sutradhar, Madhuwesh Kumar Sharma, and G. Sai Krishna

Abstract Cloud computing has become a vital piece of suggestions; nowadays, as a result of the advantages it offers to the reasonably connected world. Cloud computing has expedited suggestions in additional than one approach; from inserting the in-depth tools to upscaling and downscaling at no matter purpose needed, transferral concerning in-depth assets of the suggestions. Cloud computing market has become exponentially throughout the years and is trusty upon to develop at a lot of faster pace. One different purpose for the accomplishment of cloud computing is the flexibility it proposals like infrastructure as associate degree support.

Keywords Cloud computing · Multitenant architecture · Enterprise adopters · Exponentially · Infrastructure

1 Introduction

Cloud computing permits on-demand access to perform operations and knowledge storage means which will be designed to fulfill distinctive needs, constraints of the purchasers with borderline organization overhead [1–5]. Currently, the rise within the accessibility of cloud services makes them enticing and economically smart for purchasers with restricted computing or storage resources who are unwilling or unable to obtain and maintain their own computing infrastructure. Purchasers will simply source giant amounts of information and computation to remote locations, yet as run applications directly from the cloud. Cloud computing is combination of

N. Sutradhar · M. K. Sharma (✉) · G. Sai Krishna
Bharat Electronics Limited (BEL), Bangalore, India
e-mail: madhuweshkumarsharma@bel.co.in

N. Sutradhar
e-mail: nirupamsutradhar@bel.co.in

G. Sai Krishna
e-mail: krishnags@bel.co.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_4

many building blocks [6]. Hence, they inherit the protection constraints gift in every element [7].

2 Type of Cloud Computing

The following are the different types of cloud computing.

- Software as a service (SaaS)
- Infrastructure as a service (IaaS)
- Web services in the cloud
- Platform as a service (PaaS).

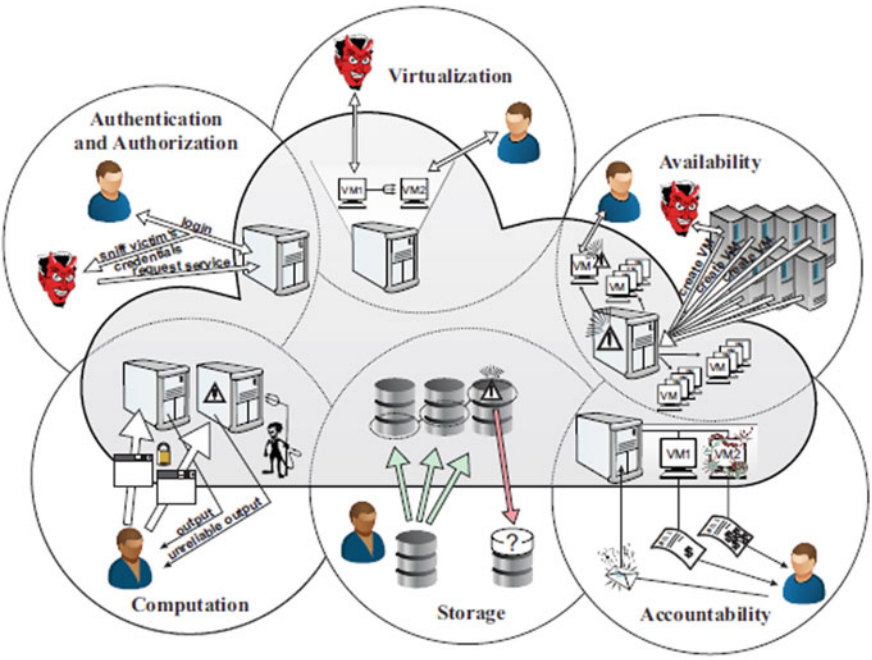
1. **SaaS (Software as a Service)**
   This style of disseminated figuring passes on a singular application through the program to an outsized assortment of buyers using a multitenant building [8–12]. On the customer angle, it suggests no frank enthusiasm for servers or programming permitting. On the supplier viewpoint, with only one application to keep awake, costs are low stood out from antiquated encouraging [13–17]. Salesforce.com offers a premier model among business segment applications, by the by SaaS is particularly essential for the unit of time applications and has even blended its far the advanced methodology of life to ERP, with players, as an occurrence, Workday. What's extra, who may have foreseen the sudden rising of SaaS "work region" applications, as a case, Google Apps and Zoho Office?
2. **Infrastructure as a Service (IaaS)**
   This type of disseminated computing is obtaining new life from Amazon, IBM, Sun and people who presently provide storage and computer-generated servers that IT will access for the asking. Primary venture adopters for the foremost half utilize utility calculation for supplemental, non-crucial needs, but in the future, they will replace quantities of the datacentre. Totally different suppliers provide arrangements that assist IT in creating virtual datacentres from product servers. For example, AppLogic, Elastic Server on Demand, LiquidQ.All these are empowering IT to affix along with memory, I/O, storage, and procedure limit as a virtualized plus, pool manageable over the system [18].
3. **Web services in the cloud**
   It is firmly known with SaaS, net specialist co-ops provide arthropod genus that empowers designers to misuse utility over the net, as hostile conveyancing all out applications. They extend from suppliers providing separate business administrations to the complete scope of arthropod genus. For example, Strike Iron, Xignite, Google Maps, ADP finance handling, the US communicating, Bloomberg, and even customary revolving credit handling administrations [19].
4. **Platform as a service (PaaS)**
   This kind of appropriated registering passes on headway things as an encourage. We can build our own applications that hazardous spike sought after for the

provider's establishment and are sent to our buyers by implies that of the net from the provider's servers. While availing PaaS facility, service providing organization will take maximum control and limits over platform that's why customer cannot exercise full functionalities to achieve consistency as well as conditioned based workplaces. Prime models consolidate Salesforce. com's Force.com, Coghead, and consequently the new Google App Engine. For unfathomably lightweight flip of occasions, cloud-based mashup stages multiply, as a case, Yahoo Pipes or fashionable.net [20–22].

3 Challenges While Using Cloud Computing

- **Client Authentication and Authorization**
  Client security requirements when services provided by cloud provider are accessed by the client through web browsers.
- **Security shortcomings of hardware virtualization**
  How virtualization can be misused to get unapproved data from defenseless clients and furthermore demonstrate alleviation procedures that can be utilized. Likewise, we will likewise deliver vulnerabilities identified with the utilization and sharing of virtual machine (VM) pictures [23].
- **Flooding attacks and denial of service (DoS)**
  Since distributed computing frameworks are intended to scale as indicated by the interest for assets, an assailant may utilize this trademark to malevolently concentrate huge segments of the distributed computing's capacity and bringing down the nature of administration that the cloud gives to other simultaneous clients [24].
- **Cloud accountability**
  Cloud ought to have the capacity to catch and uncover illegitimate action. A responsible cloud framework ought to have answers for accomplishing these abilities [25].
- **Challenges and solutions for remote storage protection**
  Methods that can be utilized by cloud customers to confirm the honesty of their re-appropriated information.
- **Protection of outsourced computation**
  How to assure the privacy and integrity of outsourced computations in cloud computing?

4 Cloud Security Architecture



4.1 How to Insure that Cloud is Safe ?

- **Authentication and Authorization using XML-based authentication in a browser**
  Customer confirmation to the cloud implies that an Internet browser needs the giving of cryptographically generous protrusible terminology (XML) tokens. Since the underlying advance of the method, the customer gives his login confirmations, that upon check and endorsement, permit the cloud to pass on the insurance tokens that the customer can later provide for the cloud server as methodologies for approval. This system, in any case, needs the help of an accepted pariah, as projects are not prepared for giving XML-put along with security tokens with pertinence their own [26].
- **Binding Issues**
  So on exactly maintain and begin-up VMs (on account of IaaS) or explicit sections (on account of PaaS), the cloud supplier must stock data depictions connected

with individual activity. A shopper could advert to those data portrayals therefore on decide the definite utility of a selected help once having a specific want. Caricaturing assault focuses on rebuilding these data.

- **Hardware Virtualization-Related Issues**
  So on fulfill the interest for calculation from various synchronic purchasers, cloud specialist co-ops often got to multifaceted the employment of their affected accessible instrumentality. Microsoft Azure and Amazon EC2 as an example, agree purchasers to increasingly take off VMs which might be modified to their explicit desires and handily oversaw, whereas in a row on the best purpose of a mutual physical framework. Vulnerabilities emerge from the approach that in exceptionally multiplexed cloud conditions, suppliers oft don't have any influence over what forms of VMs are being sent by their customers. Then again, purchasers in addition have a restricted info on the essential physical infrastructure. Associate degree assaulter will fabricate and share footage that is defiled with malware and totally different styles of dangers [27].
- **Availability of Resources at Cloud**
  Most alluring highlights of distributed computing are the way that computational force can be effectively provided on request. On the off chance that a specific help running from the cloud ensures an unexpected increment in the remaining burden, extra equipment backing can be provisioned on the fly. Notwithstanding its allure, this "flexible" normal for mists can be misused by assailants. How flooding assaults are a genuine danger to mists. This issue ascends in such conditions when, for example, all solicitations to a specific assistance should be separately checked for legitimacy, accordingly causing administration overburdening [28].
  Outcomes of administration overburdening are refusal of administration (DoS) (which might be immediate or circuitous) and inconsistencies identified with bookkeeping and responsibility [29].
- **Accountability**
  Responsibility has for some time been viewed as a property of dependable PC frameworks. A responsible framework is fit for identifying abuse and uncovering the mindful element. It should have the option to produce obvious proof of the substance's unjust action. Also, a blameless gathering in a responsible framework ought to be equipped for guarding itself against any fraudulent indictments. Such properties are especially important in conditions where customers are running their errands or administrations inside the foundation claimed or kept up by outsiders which is the situation for open mists and appropriated frameworks. To assemble a responsible framework, a few highlights ought to be mulled over:

  - **Identity binding**: Gesture-based communication each activity with the individual key of the activity engineer. This can allow various elements who hold such records to blessing them to relate degree evaluating authority to demonstrate their honesty or blame the activity engineer for wrongdoing.
  - **Tamper-evident logs**: To investigate occasion of noxious lead of a particular component, the checking on power is normally given the authentic scenery of

that component's past exercises as a log. One obvious property of this log is change evidence.
- **Execution verification**: At the point when a cloud specialist co-op executes an undertaking for the benefit of a customer and produces a result as the assignment's outcome, the customer may be keen on confirming rightness of the outcome.

- **Remote Storage Protection**
  While distributed computing offers partaking preferences, it in addition brings new security challenges. Specifically, the honesty of shoppers decentralized info turns into a worry, that is caused by associate degree the absence of straightforwardness and direct management within the cloud condition, whereas info classification is often secured through customary systems, as an instance, cryptography or mystery sharing and uprightness of knowledge transmission are often secured through information validation, trustiness of knowledge place away within the cloud is basically progressively exhausting to visualize [30].
- **Privacy and integrity of outsourced computation**
  As organizations, people, and even governments progressively redistribute calculation to the cloud, they intentionally give up a definitive command over their information and the calculation. On the off chance that the calculation includes exclusive, private, or in any case delicate information, cloud clients may wish to secure the protection of the information being utilized all the while, especially when the framework supplier cannot be completely trusted.

5 Conclusion

Cloud security has developed as a big purpose each within the examination network and much speaking as a result of the continued quick development in the accessibility and prominence of distributed computing and capability suppliers. It is often referred to as a high purpose behind aversion in receiving distributed computing administrations by organizations and could be a hard issue that resurfaces recent safety problems and carries novel ones. Within the endeavor to specialize in additional in-depth the issue, this work provides the earliest thorough management of the region by giving a piece of writing define of cloud security even as protected remote storage and calculation. The themes canvassed during this study include:

- Authentication and approval assaults and countermeasures
- Virtualization-related safety problems
- Denial of administration assaults
- Clarifications that address responsibility
- Organizations for reputability confirmation of remote storage
- Systems for security and uprightness assurance of re-appropriated calculation.

References

1. Popović K, Hocenski Ž (2010) Cloud computing security issues and challenges. In: The 33rd international convention mipro. IEEE, pp 344–349
2. Kuyoro SO, Ibikunle F, Awodele O (2011) Cloud computing security issues and challenges. Int J Comput Netw IJCN 3(5):247–255
3. Priyadarshi R, Soni SK, Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-13-0776-8_27
4. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
5. Yang J, Chen Z (2010) Cloud computing research and security issues. In: 2010 international conference on computational intelligence and software engineering, IEEE, pp 1–3
6. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
7. Rawat P, Chauhan S, Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circ Syst Comput. https://doi.org/10.1142/S0218126620502047
8. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
9. Dillon T, Wu C, Chang E (2010, April) Cloud computing: issues and challenges. In 2010 24th IEEE international conference on advanced information networking and applications, IEEE, pp 27–33
10. Soni SK, Kumar S, Priyadarshi R (2020) Energy-aware clustering in wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_39
11. Kumar S, Soni SK, Randheer, Priyadarshi R (2020) Performance analysis of novel energy aware routing in wireless sensor network. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_44
12. Priyadarshi R, Singh L, Randheer, Singh A (2018) A novel HEED protocol for wireless sensor networks. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474286
13. Harfoushi O, Alfawwaz B, Ghatasheh NA, Obiedat R, Mua'ad M, Faris H (2014) Data Security issues and challenges in cloud computing: a conceptual analysis and review. Commun Netw
14. Priyadarshi R, Singh L, Singh A, Thakur A (2018) SEEN: stable energy efficient network for wireless sensor network. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474228
15. Priyadarshi R, Yadav S, Bilyan D (2019) Performance and comparison analysis of MIEEP routing protocol over adapted LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung PA, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 237–245. https://doi.org/10.1007/978-981-13-6295-8_20
16. Gou Z, Yamaguchi S, Gupta BB (2017). Analysis of various security issues and challenges in cloud computing environment: a survey. In: Identity theft: breakthroughs in research and practice. IGI Global, pp 221–247
17. Priyadarshi R, Yadav S, Bilyan D (2019). Performance analysis of adapted selection based protocol over LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 247–256. https://doi.org/10.1007/978-981-13-6295-8_21
18. Priyadarshi R, Gupta B, Anurag A (2020) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. J Supercomput. https://doi.org/10.1007/s11227-020-03166-5
19. Zissis D, Lekkas D (2012) Addressing cloud computing security issues. Future Gener Comput Syst 28(3):583–592

20. Priyadarshi R, Gupta B, Anurag A (2020) Wireless sensor networks deployment: a result oriented analysis wireless. Pers Commun. https://doi.org/10.1007/s11277-020-07255-9
21. Morin JH, Aubert J, Gateau B (2012, January) Towards cloud computing SLA risk management: issues and challenges. In 2012 45th Hawaii international conference on system sciences. IEEE, pp 5509–5514
22. Priyadarshi R Gupta B (2019) Coverage area enhancement in wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-019-04674-y
23. Chen Y, Paxson V, Katz RH (2010) What's new about cloud computing security. University of California, Berkeley Report No. UCB/EECS-2010-5, pp 2010–201, 20 Jan 2010
24. Anurag A, Priyadarshi R, Goel A, Gupta B (2020) 2-D coverage optimization in WSN using a novel variant of particle swarm optimisation. In 2020 7th international conference on signal processing and integrated networks, SPIN 2020. https://doi.org/10.1109/SPIN48934.2020.9070978
25. Priyadarshi R, Tripathi H, Bhardwaj A, Thakur A (2018) Performance metric analysis of modified LEACH routing protocol in wireless sensor network. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i1.5.9146
26. Wang Y, Uehara T, Sasaki R (2015) Fog computing: Issues and challenges in security and forensics. In: 2015 IEEE 39th annual computer software and applications conference, vol 3. IEEE, pp 53–59
27. Priyadarshi R, Singh L, Sharma I, Kumar S (2018) Energy efficient leach routing in wireless sensor network. Int J Pure Appl Math 118(20):135–142
28. Priyadarshi R, Singh L, Kumar S, Sharma I (2018) A hexagonal network division approach for reducing energy hole issue in WSN. Eur J Pure Appl Math 118
29. Attab A, Saleh B, Fadewar HS (2014) Security issues and challenges in cloud computing. Int J Emerg Sci Eng 2(7): 22–26
30. Aery MK (2016) Mobile cloud computing: security issues and challenges. Int J Adv Res Comput Sci 7(3)

Application of Knowledge and Data Mining to Build Intelligent Systems

R. Kavitha

Abstract A great demand for new, powerful tools for analyzing data into useful, task oriented knowledge is emerged due to tremendous proliferation of databases nearly in every area of human endeavors. The disclosure of information in databases has inferred incredible significance and enthusiasm for ongoing years, particularly in national security regions where the entrance to the data is confined ordinarily. Specialists find it infeasible to analyze the large amount of data stored by conventional methods. The reason for generation of data is cost reduction and easiness of storage in digital media which has resulted in significant development in the measure of put away data in computerized structure. The trouble for an increasingly exact examination of these information implies they may become assembled pointless data. Then again, it is realized that in huge amount of information, a tremendous capability of extractable data can exist. A colossal test is the manner which abuses those data and information assets and transforms them into helpful information available to interested people, since the value of knowledge increases when people can share and take advantage on it. The capacity to gather, break down and discover new data from that information is incredibly helpful to the national security. Approaching helpful data makes them increasingly productive/wise in what they do. Today, they use it in strategic frameworks, reconnaissance and clinical databases just to give some examples. In this article, we are trying to bring out the use of data mining in knowledge discovery for yielding better decision based on the quondam and current stream of data so that the systems using these approaches can handle the situation in better way.

Keywords Knowledge · Data · Information · Intelligent · Data analytics · Decision support system · Representation

R. Kavitha (✉)
BEL Software Technology Center, BEL, Bengaluru, India
e-mail: kavithar@bel.co.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_5

1 What is KDD?

We live in a data age. In any case, we additionally need to understand that we live in information rich however very regularly information poor condition [1]. Or on the other hand, how about we put it in an unexpected way: There is a serious and extending hole between information age and information perception [2]. Information is being gathered and put away across wide assortment of fields at a striking pace. Along these lines, a prompt need has happened to create new computational speculations and devices to help with extricating significant data/information from the fast pace of volumes of computerized information. The data, even put away, turns out to be less and less helpful as we are confronted with troubles to recoup its utilization. More elevated level of summarization in simple intelligible organization is likewise hard to accomplish in this situation [3]. The above speculations and apparatuses have laid way to the developing field of information revelation in databases (KDD) [4]. KDD can be characterized in a theoretical level as a field, which is worried about advancing strategies and methods required in understanding information. Mapping low-level information which are excessively huge in number or amount and review effectively into different structures that may be increasingly unique (for instance, an expressive guess or model of the procedure that created the information), progressively conservative (for instance, a short rundown) or increasingly valuable (for instance, a prescient model for evaluating future business) [5–8].

2 What is Data Mining and Text Mining?

The assessment in databases and information development has begun to an approach to manage store and control this significant data to engage dynamic [9]. Data mining is the craftsmanship and investigation of discovering data, bits of information, individualization and models in data [10]. The system of extraction of supportive information and models from tremendous data will be data mining. It is moreover called as data divulgence process, data mining from data, data extraction or data/plan assessment [11].

Consistently the volume and assorted variety of logical writing are extending. Logical writing is recorded in detail by plenteous documents as a reason for future turn of events. Information mining attempts to find data disguised in logical writing, which is not agreeable by basic measurable methods [12]. Content mining strategies is a noteworthy subset of information mining, which plans to remove information from unstructured or semi-organized literary information and has far reaching applications in breaking down and preparing printed archives [13].

Content Mining is otherwise called smart text analysis or knowledge discovery in text or text information mining. Content mining can help with recurrence investigation of significant terms and their semantic connections [14, 15]. Content rundown is a procedure of gathering and delivering inclusive portrayal of unique content reports.

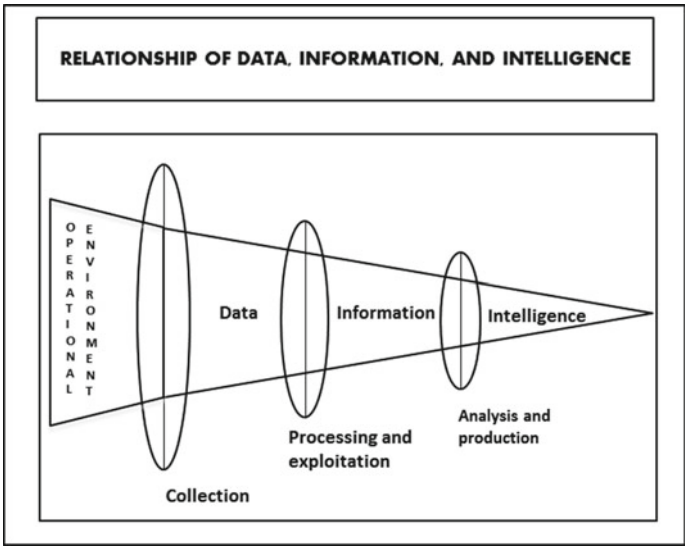


Fig. 1 Data, information and knowledge

Pre-handling and preparing activities are performed on the crude content for synopsis [16, 17]. Tokenization, stop word expulsion and stemming techniques are applied for pre-preparing. Prior, programmed content synopsis was performed based on event a specific word or expression in archive. As of late, extra techniques for content mining were presented with standard content mining procedure to improve the importance and exactness of results. Figure 1 represents the process of building unprocessed data into processed intelligence.

3 Why KDD or Data Mining is Required?

Manual analysis and interpretation are the traditional methods to rely on for converting data into knowledge. Be it any field like health care, retail, science, marketing, defense, finance or any other in case, fundamentally the classic approach for one or more analysts relies on becoming closely acquaint with the information and assists as a border between the information and products. Thus, an immediate need has occurred to generate new computational philosophies and apparatuses to support in mining significant information/knowledge from the rapid pace of capacities of digital information. For example, data generated from remotely sensed images of planets and asteroids are important to planetary geologists. In this case, it is not possible to manually analyze the data recorded; hence, a new computational assistance is required.

4 How KDD and Data Mining Work Together?

Automatic exploratory analysis and modeling of large data repositories is the core concept of Knowledge Discovery in Databases (KDD). The task of KDD is to identify novel, valid, untestable and useful patterns in an organized process in large and complex data sets. The core of the KDD process is the data mining (DM) concept, which involves in the inference of algorithms to explore the data, developing data mining model and discovery of previously unknown patterns [18, 19]. Data mining model is used for the phenomena of understanding from the data, analysis and prediction. To learn from data, quality data needs to be effectively gathered, cleaned and organized, and then efficiently mined. Technology and skill is required for consolidation and integration of data elements from many sources [20].

The process of knowledge discovery is interactive and iterative. Each step is iterative process meaning that traversing back and forth between the steps may be required. Artistic aspect of the process is defined in the sense as that one cannot present one formula and make a complete taxonomy for the correct choice for every step and application type [21, 22]. Thus, it leads us to understand the process and the various requirements and options in each step. Each KDD process starts with defining what should be the final goal, and discovered knowledge implementation is the final step.

Sequences of KDD.

4.1 Develop an Understanding of the Application Domain

There is an ever-delivering torrential slide of information accompanying higher speed, volume and assortment. Anybody needs to rapidly utilize it or lose it. Information disclosure in more astute manner requires picking where to play. One needs to settle on astute choices about what to assemble and what to disregard, in light of the goal of the information disclosure works out, as not all surges of information will be similarly wealthy in potential experiences.

4.2 Selection and Integration of Data Set on Which Knowledge Discovery Will Be Achieved

Deciding the information which will be utilized for information disclosure ought to be finished, when the objectives are characterized. This procedure incorporates distinguishing what information is accessible with us, accepting extra required information and consolidating all the information got for the information revelation into single informational index. The characterized procedure is exceptionally basic as the

information mining process learns and finds structure of the acquired information. The information quality shifts from approaching source and nature of the information.

4.3 Pre-Processing and Cleansing

At this stage, data clearing is included, such as removal of disturbance or outliers and missing values are handled to increase data dependability. Assembling and curating data take effort and time, particularly when it is semi-structured or unstructured. Unstructured data can come in many forms like images, videos, audio, databases, blogs and chats. There are streams of unstructured social media data from chats, tweets and blogs. Unstructured data might require complex data mining algorithm or statistical methods in this circumstance. For example, missing values—let us assume that a certain attribute is missing reliability, then that attribute might become destination of a supervised algorithm in data mining. A prediction model will be developed to predict the missing data for the attribute.

4.4 Data Transformation

Data transformation step is used to collect the data and place into some standard formats recognized by mining methods, for example, clustering, association or classification. Methods like anomaly detection and attribute importance of data mining are used, and data is prepared and developed for mining in this stage. This step is very important as the success of the complete data mining/KDD process is dependent on the data. Data will be usually project specific. Even if data is not correctly transformed initially, hints can be obtained from any surprising effect to us about the correct transformation needed.

4.5 Choosing the Relevant Data Mining Task

This step is very crucial, as we decide which one of among several data mining techniques will be selected as per the desired goals to be achieved. Each data mining technique and algorithm can serve different types of objectives. It is our call to decide which data mining technique to be used like classification, regression or association. The selection mainly depends on KDD goals and previous iterative steps. The two major goals of data mining: description and prediction. Prediction is referred often as supervised data mining, and the supervised data mining learning process is guided by a recently recognized ward characteristic or target. Guided information mining process endeavors to clarify the conduct of the objective as an element of a lot of free predictors or attributes. Supervised learning in general results in predictive models.

Descriptive data mining includes the unsupervised learning and visualization aspects of data mining, while the goal of unsupervised learning is pattern detection which is in contrast to controlled education. Unsupervised education is non-directed. There is no variation between autonomous and dependent features and no earlier identified outcome to monitor the development in building the model. Unsupervised education can be used for descriptive resolutions and making predictions.

4.6 Choosing the Correct Data Mining Algorithm

The previous steps will decide the strategy, now tactics to be decided in this step. This step includes selection of specific methods or technique for searching the pattern required to achieve the goal. For example, neural networks work better with considering precision values and decision trees performs better with understanding ability. Each strategy of meta learning can be achieved with several possibilities. Meta learning will be focusing on explanation of success factor about a particular problem. Each algorithm has various tactics and parameter of learning.

4.7 Employing the Data Mining Algorithm

At this step, data mining algorithm is implemented. The algorithm man requires to operate in iterative mode several time until a satisfied or desired result is obtained, for example, by alteration of definite algorithm process control factors, such as the least quantity of illustrations in a single leaf of a choice tree.

4.8 Pattern Evaluation

Identification of pattern is done in pattern evaluation step. In this step, goals are evaluated against the defined one in the first step, and it interpret the mined patterns (rules, reliability, etc.), against the goals. The effect of the data mining algorithm results is considered with respect to the pre-processing steps. Focus is mainly on the usefulness and comprehensibility of the selected model in this step. Documentation of the discovered knowledge starts at this step for further usage in the iteration. Each data mining tools or algorithms can serve different types of goals and objectives. The outcome of each data mining tools or algorithms will reflect the objective being served. The outcome of the data mining can be represented in various formats.

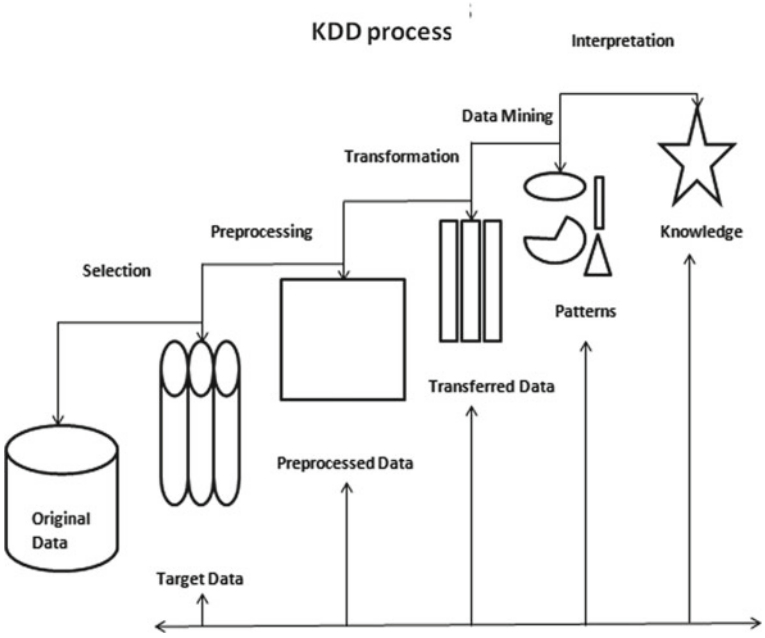


Fig. 2 KDD process

4.9 Knowledge Representation and Its Usage

This is the last stage in which the knowledge is represented using different visualization techniques. We are prepared presently to join the information into another framework for additional activity. The information gets dynamic as in we can make changes to the framework and measure the impacts. Actually the success of this step determines the effectiveness of our entire KDD process.

Figure 2 summarizes the overall process in knowledge discovery in databases.

5 What are Intelligent Systems?

An intelligent system is a machine with an introduced, Internet-related (IoT) PC that can accumulate and separate data and talk about the proportional with various frameworks. Different models for sharp frameworks can join the capacity to pick up from security, accessibility, experience, the ability to change as demonstrated by current data and the cutoff with respect to remote organization and checking [23–25].

Intelligent systems exist encompassing us in cutting edge TVs, traffic lights, retail store (POS) terminals, sharp meters, cars, electronic signage and plane controls, among a phenomenal number of various possibilities. Worked in information is a

key section of making web of things (IoT), in which about everything conceivable can be given stand-out identifiers and the capacity to typically move information over a system without envisioning that human-should human or human-to-PC joint exertion [26].

6 How Data Mining Helps Intelligent Systems?

The two significant level essential objectives of information mining by and by will in general are expectation and depiction. Information mining can give a major lift to insight gathering when there is a data storm as a result of the accompanying viewpoints related with future information:

•**Volume**: Information will develop at measure going from terabytes to petabytes.

•**Variety**: Information will be in organized, semi-organized, unstructured organizations and from various sources.

•**Veracity**: Managing the dependability and consistency of naturally uncertain information types will be testing.

•**Complexity**: Relationships, pecking orders and numerous information linkages should be associated and corresponded.

Revelation of helpful, conceivably startling examples in information, non-inconsequential extraction of certain, beforehand obscure and possibly valuable data from information via programmed or self-loader methods for examination. This prompts to find significant examples prompting information reproduction and demonstrating. As it will be the solicitation for the day as a segment of the Intelligence, Surveillance and Reconnaissance (ISR) thinking.

7 Use Case for Decision Support System

An example of an astute system helps investigators and leaders recognize focuses of premium and settle on prohibition choices. This is a far reaching choice help stage with vigorous data sharing capacities that permit accomplice organizations and partners to share data specifically and safely. It is created by coordinating information from various sources, for example, earthbound and satellite-based boat position/AIS information, and is dug for the information utilizing KDD process. This information is breaking down and introduced in an easy to use interface which bolsters a collective workplace and dependent on the examination gave a superior choice can be taken [27–30].

8 Use Case for Log Mining

Logs and log investigation are of indispensable significance in the present canny frameworks. By methods for investigation framework disappointments and oddities can be identified from occasion logs and the difficult determination procedure can be improved. Yet, as the volume of information expands supposed to gigabytes at that point, it gets unthinkable for the conventional techniques to examine such a gigantic log document and decide the substantial information. By disregarding the log information, an enormous hole of significant data will be made.

Along these lines, the answer for this issue is to utilize information digging for the log information to separate applicable information identified with any issue from the enormous information created from the activities. With this present, it is not required to peruse the entire log record information by the person. By joining the utilization of log information with the accessible calculations in the information mining procedure, it gets conceivable to pick up the important ideal presentation and exhaustive operational perceivability. Alongside the examination of log information, there is additionally a need to characterize the log document into relevant and superfluous information. With this methodology, time and execution exertion could be spared and near exact outcomes could be acquired.

9 Use Case for Security Operations

The significance of data and knowledge for the assurance of any nation is brought out from the way that it is key to arranging of any activity. Observation and reconnaissance assignments have been done by countries since days of yore before mounting any ambush on the adversary. The achievement of the mission relies on the right examination of the accessible data during the arranging stage. The assortment of data, its examination and further spread is of prime significance. Today, the techniques for gathering data have changed because of fast accessibility of data of high caliber and unwavering quality from a differed cluster of sensors and sources. An ideal ISR plan is figured by coordinating the ISR mission's dependent on abilities of different resources accessible. The data is from that point grouped and examined for taking care of in to the arranging of activities.

10 Use Case for Document Analytics

System which changes over of satellite information and specialized scrambled captures require calculation gave by information investigation. Calculations to be created to break down a huge number of open-source records produced every 60 min and contrast them and the human insight accumulated and billions of chronicled

occasions, and afterward have prescient ability to envision explicit occurrences and propose quantifies proactively. The data about a dubious individual can be questioned with enormous information devices on interpersonal organizations, shopping destinations and amusement locales, and from the network logs of the pursuit did in the net, and moves can be made.

Ideas, themes talked about in a report assortment can be considered. Content mining can be utilized here. For instance, the points that a Foreign Ministry is centered around, in light of breaking down the articles/reports on its site. The archives can be assembled in independent sections: reports talking about governmental issues, sports, games, outside undertakings and so on. Discovering patterns identified with explicit themes may be a significant method to find the shrouded bits of knowledge in a record. Later content outline can likewise be utilized by the system to create synopsis. Content rundown methods can be applied on numerous archives simultaneously.

11 Use Case for Retort Against Threat to Civilians

System to answer against danger to tasks, has immense information gathered by rambles, satellites, UAVs/specialized captures, and so forth can be naturally broke down dependent on the 10,000-foot view gave by ISR information which can be made as a component of the information mining activity. The treatment of this information needs fast advances from system engineering to developments in measurements. Enormous scope and top to bottom mining of information and representation of the mining bring about different structures like maps, diagrams and courses of events, and close to ongoing investigations of spilling information are a portion of the fundamental necessities for security issues. This will permit the insurance tasks to be done continuously.

12 Conclusion

In view of the quickly expanding measure of organized and unstructured information accessible in electronic structure, the need to help access and concentrate valuable information, reveal shrouded information and gain by information is progressively significant even in national safety. Providing security to civilians against threat/violence requires a nimbler intelligence gadget that functions more aggressively. Information mining and mechanized information investigation strategies are useful assets for knowledge and law implementation authorities battling against threat/violence. The main objective of the paper is to explain data mining and how it is used in knowledge discovery in database. Evidence-based data, superiority and fully analyzed intelligence productivities will give improved situational responsiveness in national safety field. The intelligence organizations should be able to accumulate, assemble, filter and procedure all varieties of input, from controlled to uncontrolled,

with live feed, and display it. With the use cases, it is tried to bring how quickly an intelligent system can take decision, based on the result of mined knowledge.

References

1. Cao L, Gorodetsky V, Mitkas PA (2009) Agent mining: the synergy of agents and data mining. IEEE Intell Syst 24(3):64–72
2. Cannataro M, Talia D (2004) Semantics and knowledge grids: building the next-generation grid. IEEE Intell Syst 19(1):56–63
3. Priyadarshi R Soni SK Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-13-0776-8_27
4. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
5. Ngai EW, Xiu L, Chau DC (2009) Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst Appl 36(2):2592–2602
6. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
7. Rawat P Chauhan S Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circ Syst Comput. https://doi.org/10.1142/S0218126620502047
8. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
9. Kohavi R, Provost F (2001) Applications of data mining to electronic commerce. In : Applications of data mining to electronic commerce. Springer, Boston, MA, pp 5–10
10. Randheer Soni SK, Kumar S, Priyadarshi R (2020) Energy-aware clustering in wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_39
11. Kumar S, Randheer Soni SK, Priyadarshi R (2020) Performance analysis of novel energy aware routing in wireless sensor network. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_44
12. Priyadarshi R, Singh L, Randheer, Singh A (2018) A Novel HEED protocol for wireless sensor networks. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474286
13. Kusiak A (2006) Data mining: manufacturing and service applications. Int J Prod Res 44(18–19):4175–4191
14. Priyadarshi R, Singh L, Singh A, Thakur A (2018) SEEN: stable energy efficient network for wireless sensor network. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474228
15. Priyadarshi R, Yadav S, Bilyan D (2019) Performance and comparison analysis of MIEEP routing protocol over adapted leach protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung PA, Mishra RB (eds) smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 237–245. https://doi.org/10.1007/978-981-13-6295-8_20
16. Cao L, Zhang C, Yang Q, Bell D, Vlachos M, Taneri B, Keogh E et al (2007) Domain-driven, actionable knowledge discovery. IEEE Intell Syst 22(4): 78–88
17. Priyadarshi R, Yadav S, Bilyan D (2019) Performance analysis of adapted selection based protocol over LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung PA, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 247–256. https://doi.org/10.1007/978-981-13-6295-8_21

18. Priyadarshi R, Gupta B, Anurag A (2020) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. J Supercomput. https://doi.org/10.1007/s11227-020-03166-5
19. Romero C, Ventura S, Pechenizkiy M, Baker RSJD (eds) (2010) Handbook of educational data mining. CRC Press
20. Priyadarshi R, Gupta B, Anurag A (2020) Wireless sensor networks deployment: a result oriented analysis. Wireless Pers Commun. https://doi.org/10.1007/s11277-020-07255-9
21. Morin JH, Aubert J, Gateau B (2012, January) Towards cloud computing SLA risk management: issues and challenges. In : 2012 45th Hawaii international conference on system sciences. IEEE, pp 5509–5514.
22. Priyadarshi R, Gupta B (2019) Coverage area enhancement in wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-019-04674-y
23. Peddabachigari S, Abraham A, Grosan C, Thomas J (2007) Modeling intrusion detection system using hybrid intelligent systems. J Netw Comput Appl 30(1):114–132
24. Anurag A, Priyadarshi R, Goel A, Gupta B (2020) 2-D coverage optimization in WSN using a novel variant of particle swarm optimisation. In 2020 7th international conference on signal processing and integrated networks, SPIN 2020. https://doi.org/10.1109/SPIN48934.2020.9070978
25. Priyadarshi R, Tripathi H, Bhardwaj A, Thakur A (2018) Performance metric analysis of modified LEACH routing protocol in wireless sensor network. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i1.5.9146
26. Zaïane OR, Han J, Li ZN, Hou J (1998, November) Mining multimedia data. In Proceedings of the 1998 conference of the centre for advanced studies on collaborative research. IBM Press, p 24
27. Priyadarshi R, Singh L, Sharma I, Kumar S (2018) Energy efficient leach routing in wireless sensor network. Int J Pure Appl Math 118(20):135–142
28. Priyadarshi R, Singh L, Kumar S, Sharma I (2018) A hexagonal network division approach for reducing energy hole issue in WSN. Eur J Pure Appl Math 118
29. Cios KJ, Kurgan LA (2005) Trends in data mining and knowledge discovery. In Advanced techniques in knowledge discovery and data mining. Springer, London, pp 1–26
30. Xiao F, Fan C (2014) Data mining in building automation system for improving building operational performance. Energy Build 1(75):109–118

Digital Implementation of Sigmoid Function in Artificial Neural Network Using VHDL

Manmohan Bhardwaj, Aradhana, Aditya Kumar, Prashant Kumar, and Vijay Nath

Abstract This paper deals with the top-down design methodology of an artificial neural network (ANN) whose architecture is implemented in VHDL. The simulation of this architecture is carried out on Xilinx ISE 8.1. This realization gives flexibility to user to change the description at their own will. The basic elements of artificial neurons constitute the set of input nodes that receive the corresponding input signal or pattern vector, a set of synaptic connections whose strengths are represented by a set of weights activation function Φ that relates the total synaptic input to the output (activation) of the neuron. This paper strives to achieve the linearity of sigmoidal function because it is a nonlinear function whose realization is very difficult and expensive on hardware due of infinite exponential series. To overcome this problem, the proposed paper focuses on the linearity of sigmoid function by taking piecewise linear approximation which is in the form of $y = ax + b$.

Keywords Artificial neural network · Sigmoid function · Linearity · VHDL

M. Bhardwaj · Aradhana · A. Kumar (✉) · P. Kumar · V. Nath
Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: aditya88952@gmail.com

M. Bhardwaj
e-mail: mb.bitmesra@gmail.com

Aradhana
e-mail: aradhana1997111@gmail.com

P. Kumar
e-mail: prashant.paccific@gmail.com

V. Nath
e-mail: vijaynath@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_6

1 Introduction

A human brain has about 100 billion neurons connected by 100 trillion connections. Likewise, ANN is inspired by biological neuron in our brain. Due to this massive neural network, brain is able to do functions such as voluntary movement, speech, intelligence, memory, emotion, and sensory processing. Brain processes the information mainly in parallel and distributed manner that is the information is commonly stored in myelin layers of axons and hence distributed and processed in large number of neurons parallel, and therefore, loss of few neurons does not significantly affect the information. The attractive feature of neuron is that it has ability to learn from the previously learnt information and respond accordingly [1]. Artificial neuron processes large number of applications such as voice recognition, intelligent searching, medical diagnostic, financial forecasting, and machine diagnostic [2].

An artificial neural network consists of input signals x_I. Each input signal is multiplied by corresponding weights w_I. Weight represents the strength of interconnection between neurons. Then, these weights inputs are fed to summing junction [3]. If sum exceed a certain threshold limit, signal is emitted from neuron. An activation function Φ is there which relates total synaptic input to output of neuron, which is shown in Fig. 1.

So, total input is given by Eq. (1).

$$u = \sum_{i=1}^{I} w_i x_i. \tag{1}$$

So, output is $\Phi(u)$, where $\Phi(u)$ is activation function of neuron. Commonly, we select sigmoidal activation function given by Eq. (2).

$$\Phi(u) = \frac{a}{1 + e^{-bu}} \tag{2}$$

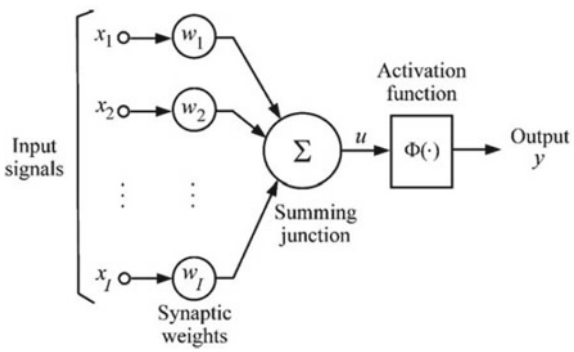
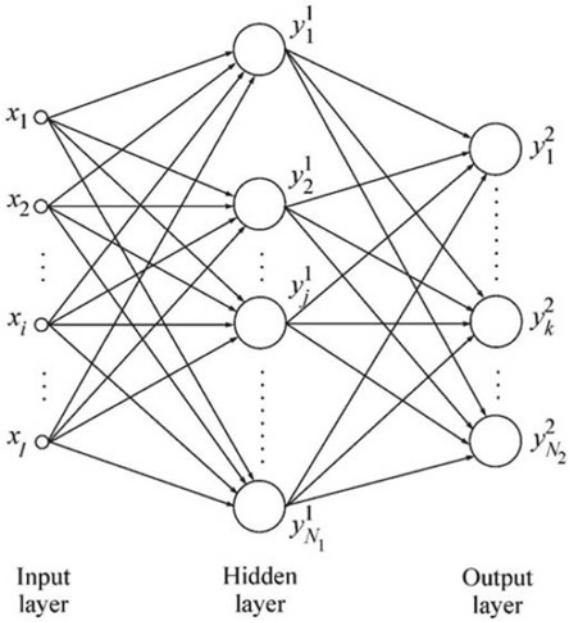


Fig. 1 Architecture of artificial neuron

Fig. 2 Multilevel perceptron



A neuron with threshold activation function is called discrete perceptron and with continuous activation function called sigmoidal function, such neuron is continuous perceptron [4]. Neural network learn either by supervised or unsupervised learning. In the first one, since the neuron is already trained with the correct data. So, for each input pattern, there is desired output pattern. In the latter part, since data is not available for training. So, for each input pattern, the network adjusts weights without knowing correct data. This is also known as delta learning.

Apart from linear perceptron, multilayer perceptron is developed which is far more efficient that linear perceptron. It is a feed-forward network that consists of input layer, two or more layers of perceptron, and output layer [5]. In between input and output layer, there are various layers present known as hidden layers as shown in Fig. 2. MLP is very useful for solving complex real-world problems. In this case, each node uses a nonlinear activation function. MLP uses back-propagation technique for its training [6].

1.1 Overview of VHDL

VHDL is a hardware description language in which there is a entity block and architecture block. In entity block, input and output are defined, whereas in architecture signals and actual behavior are defined.

2 Methodology

In proposed paper, our design suggests a method to tackle problems encountered by neural architecture activation function and artificial neural network design [7]. A neuron processes the design in three steps: the synaptic weights of input nodes are summed together and fed to sigmoidal function for the activation of neuron signals [8]. Activation function is very essential for the activation of neuron signal. Commonly used activation function is sigmoid function given by Eq. (3)

$$\Phi(u) = \frac{a}{1 + e^{-bu}} \tag{3}$$

The above function is nonlinear which makes it possible to approximate any function, but in hardware implementation, the realization of above equation is very troublesome. Therefore, for the designer, it is very difficult and expensive to implement sigmoidal function on FPGA because of infinite exponential series [9]. To overcome this problem, the proposed paper focuses on the linearity of sigmoid function by taking piecewise linear approximation which is in the form of $y = ax + b$ is given by Eq. (4).

$$\theta(v) = \begin{cases} 1 & \text{for } w \le v \\ v(\beta - \varphi) & \text{for } 0 \le v \le \omega \\ v(\beta + \varphi) & \text{for } -\omega \le v \le 0 \\ -1 & \text{for } v \le -\omega \end{cases}. \tag{4}$$

where β and φ represent the slope and the gain of the nonlinear function $\theta(v)$ between the saturation regions $-\omega$ and ω, respectively. $\theta(v)$ is implemented by one binary shift, one binary addition, and one multiplication [10]. Fig. 3 given below shows a structural diagram of the approximated sigmoid function implemented using this process [11].

The proposed scheme is implemented in Xilinx 8.1 ISE navigator.

Code

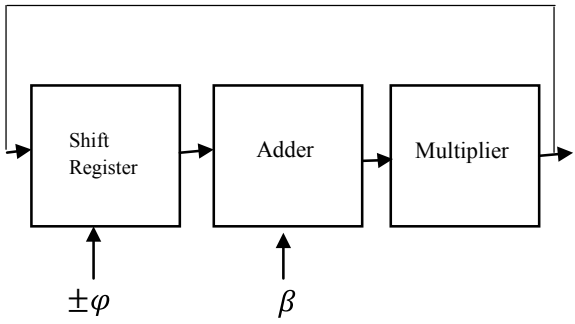


Fig. 3 Structural diagram of the approximated sigmoid function implemented

```
Library IEEE;
 Use IEEE.std_logic_1164.all;
 Use IEEE.std_logic_arith.all;
 Use IEEE.std_logic_unsigned.all;
 Entity Sigmoid is
        Port(  Z:in INTEGER range 0 to 511;
             S:out INTEGER range 0 to 255 );
 End Sigmoid;
 Architecture Sigmoidarch of Sigmoid is
Signal TEMP:integer range  0 to 255;
Signal B:integer range 0 to 511;
Signal A:integer range 0 to 511;
 Signal K:integer range 0 to 65535;
 Signal ZZ: integer range 0 to 262144;
 Constant L: integer range 0 to 255:=255;
 Constant M: integer range 0 to 1023:=512;
Begin
 A<=Z;
 B<=M-A;
ZZ<=B*A;
K<=ZZ/256;
TEMP<=K;
S<=TEMP when A<L else L;
End Sigmoidarch;
```

3 Simulations

Figure 4 shows the output wave form of the sigmoid function which shows the linearity of the function.

Figure 5 shows technological schematic view of sigmoid function. This helps us to see the technology–level schematic representation of the HDL optimized for a specific Xilinx architecture.

Figure 6 shows the RTL schematic of the sigmoid function. Register-transfer-level abstraction is used in **VHDL** to create high-level representations of a circuit, from

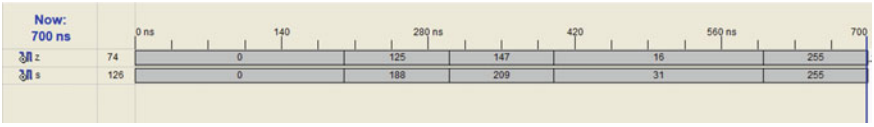


Fig. 4 Xilinx simulation output of sigmoid function

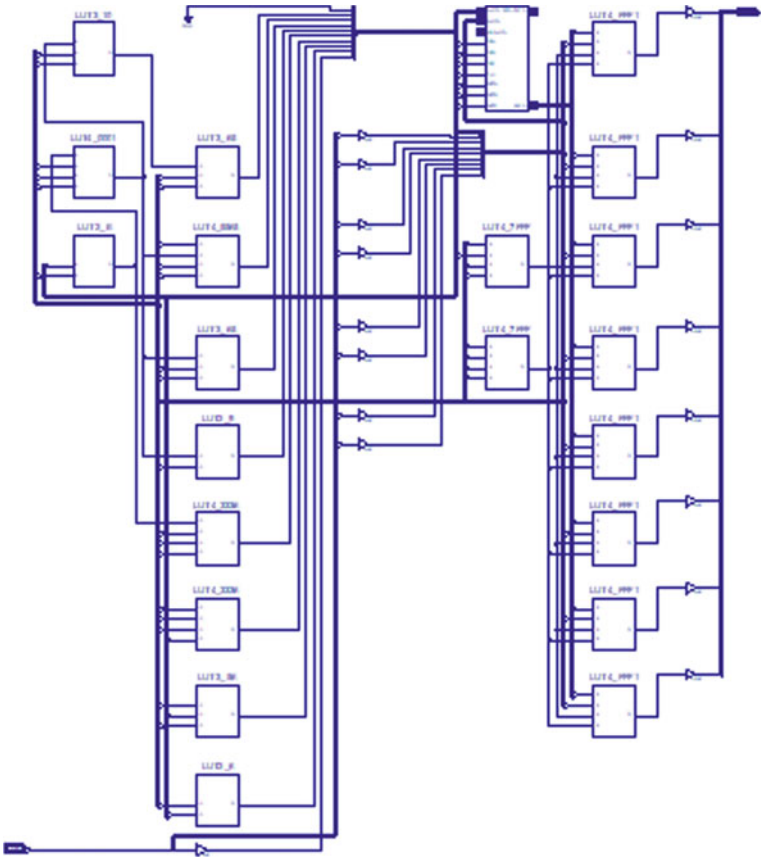


Fig. 5 Technology-level representation

which lower-level representations and ultimately actual circuitry can be derived. Design at the **RTL** level is practiced in modern digital design.

Figure 7 shows the detailed project status of device utilization summary of sigmoid function.

4 Conclusion

Through our proposed scheme, we have presented the synthesis of digital ANN in VHDL. The above graph shown in Fig. 8 describes the linearity of sigmoid function. Here, series 1 represents the sigmoid function curve, whereas the linear (series 1) is the ideal curve of linear sigmoid function. The equation of the curve is $y =$

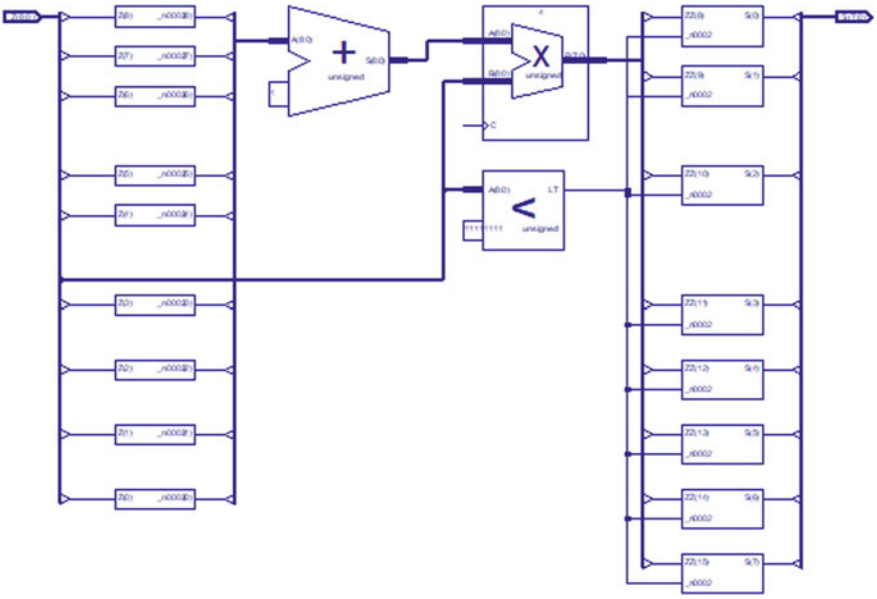


Fig. 6 RTL schematic of sigmoid function

ASD Project Status			
Project File:	asd.ise	**Current State:**	Placed and Routed
Module Name:	Sigmoid	**• Errors:**	No Errors
Target Device:	xc3s500e-5fg320	**• Warnings:**	No Warnings
Product Version:	ISE, 8.1i	**• Updated:**	Sun Jul 14 23:35:54 2019

Device Utilization Summary				
Logic Utilization	**Used**	**Available**	**Utilization**	**Note(s)**
Number of 4 input LUTs	21	9,312	1%	
Logic Distribution				
Number of occupied Slices	12	4,656	1%	
Number of Slices containing only related logic	12	12	100%	
Number of Slices containing unrelated logic	0	12	0%	
Total Number of 4 input LUTs	21	9,312	1%	
Number of bonded IOBs	17	232	7%	
Number of MULT18X18SIOs	1	20	5%	
Total equivalent gate count for design	126			
Additional JTAG gate count for IOBs	816			

Performance Summary			
Final Timing Score:	0	**Pinout Data:**	Pinout Report
Routing Results:	All Signals Completely Routed	**Clock Data:**	Clock Report

Fig. 7 Project status

$1.0465x + 22.95$ with the error of $R^2 = 0.9222$. There is 100% utilization of number of slices containing relatable logic.

The implementation of the proposed sigmoid function is easy to implement with low cost as shown in the comparison Table 1.

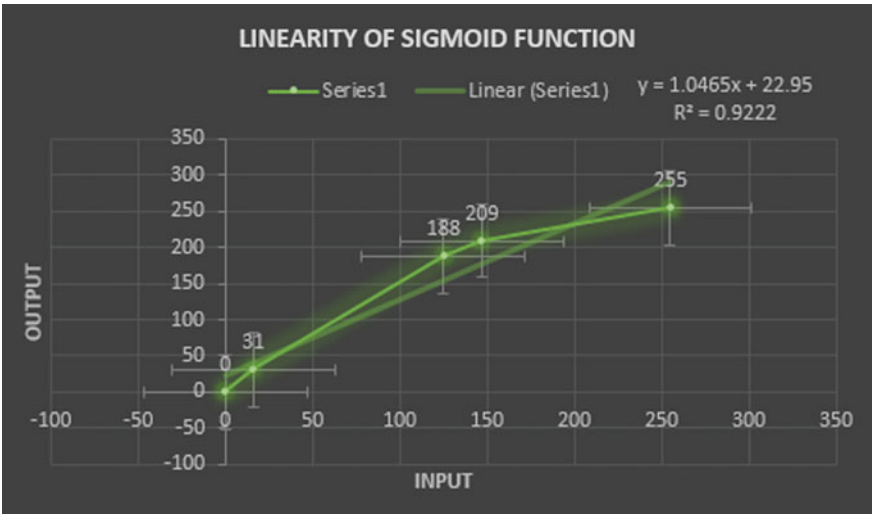


Fig. 8 Graph showing the linearity of sigmoid function

Table 1 Comparision between the proposed scheme and other related works

Our proposed scheme	Other related works
Easy to implement	Difficult to implement
Low cost	Expensive
May not approximate some function	Can approximate many function

Acknowledgement We are thankful to our Vice-chancellor Prof. S. Konar for supporting us in this project. We are also thankful to Birla Institute of Technology Mesra, Patna Campus, faculty as well as summer school on digital VLSI design 2019 for guiding us on this paper. We are grateful to our parents and our friends to support us in this project.

References

1. Singh S, Abhijit M, Pratham BS, Chirag PT, Abhinav A (2017) Hardware implementation of a neural network based robust sensor fault accommodation system in flight control system. In: 2017 international conference on intelligent computing and control systems (ICICCS), pp 962–966
2. Parra D, Camargo C (2018) A systematic literature review of hardware neural networks. In: 2018 IEEE 1st Colombian conference on applications in computational intelligence (ColCACI), pp 1–6
3. Clere SR, Sethumadhavan S, Varghese K (2018) FPGA based reconfigurable coprocessor for deep convolutional neural network training. In: 2018 21st Euromicro conference on digital system design (DSD), pp 381–388

4. Medus LD, Iakymchuk T, Frances-Villora JV, Bataller-Mompeán M, Rosado-Muñoz A (2019) A novel systolic parallel hardware architecture for the FPGA acceleration of feedforward neural networks. Access IEEE 7: 76084–76103
5. Wibowo FW (2019) An analysis of FPGA hardware platform based artificial neural network. J Phys Conf Ser 1201:012009
6. Martínez-Prado MA, Rodríguez-Reséndiz J, Gómez-Loenzo RA, Camarillo-Gómez KA, Herrera-Ruiz G (2019) Short informative title: towards a new tendency in embedded systems in mechatronics for the engineering curricula. Comput Appl Eng Educ 27: 603
7. Tank DW, Hopfield JJ (1986) Simple "neural" optimization networks: an A/D converter, signal decision circuit, and a linear programming circuit. IEEE Trans Circ Syst CAS 33: 533–541
8. Chua LO, Lin GN (1984) Non-linear programming without computation. IEEE Trans Circ Syst CAS 31: 182–186
9. Agranat AJ, Neugebauer CF, Nelson RD, Yariv A (1990) The CCD neural processor: a neural network integrated circuit with 65536 programmable analog synapses. IEEE Trans Circ Syst 37:1073–1075
10. Hagan MT, Demuth HB, Beale MH (1996) Neural network design. PWS Publishing, Boston, MA
11. Zhang D (1999) Parallel VLSI neural system design. Springer-Verlag, Singapore

Multi-sized Multiple Text File Encryption System using One-dimensional Chaotic Map

K. Abhimanyu Kumar Patro, Abhinav Agrawal, Raushan Raj, and Bibhudendra Acharya

Abstract This paper proposes a text encryption system to encrypt multiple text files of arbitrary sizes. The proposed work uses one-dimensional piece-wise linear chaotic map (PWLCM) to perform multi-sized multiple text file encryption. In the method, a group of text files is partitioned into number of sections, each one with 4-text elements. The text elements in each of the sections are then arranged into 2×2 dimensional text blocks. All the 2×2 blocks and the remaining text elements (if any) are arranged in a separate array. Finally, the array of blocks and remaining text elements are permuted and diffused using two different PWLCM systems. The SHA-256 hash function is used to generate the keys to prevent the proposed cryptosystem against chosen-plaintext and known-plaintext attacks. The computer simulation reveals the good encryption results of the multi-sized multiple text file encryption system. The security analysis confirms the stronger resistivity of the proposed algorithm against the most widely used security attacks.

Keywords Security · Text encryption · Multi-sized multiple text files · Chaos · Hash algorithm

K. Abhimanyu Kumar Patro · A. Agrawal · R. Raj · B. Acharya (✉)
Department of Electronics and Communication, National Institute of Technology Raipur, Raipur, Chhattisgarh 492010, India
e-mail: bacharya.etc@nitrr.ac.in

K. Abhimanyu Kumar Patro
e-mail: kpatro.phd2016.etc@nitrr.ac.in

A. Agrawal
e-mail: abhinav1291@gmail.com

R. Raj
e-mail: raj31.raushan@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_7

1 Introduction

With the growing use of text data over the Internet, security threats due to third-party attacks are a serious issue. The text data may be in the form of bank transactions, mobile SMS, medical reports, military secrets, industry, etc. [1]. So it is a challenge for all the cryptographic researchers to secure the confidential text information. Encryption is one of the popular approaches to secure the text information. Various encryption techniques like Data Encryption Standard (DES), Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA) [2, 3], etc., have been developed. Among them, DES was used in early stages of research to encrypt image and text information. But, using DNA computing methods, DES and RSA can be broken [4, 5]. Nowadays, AES is used in the encryption field because of their superior properties such as easy implementation, low memory requirement, and high speed [1].

Recently, researchers have suggested the use of chaotic maps in text encryption fields. Chaotic maps have several good properties such as high sensitivity to key parameters, easy implementable software and hardware platforms, ergodicity, determinacy, non-periodicity, and pseudo-random property [6–13]. Several research articles have been published in chaos-based text encryption techniques. Sobhy et al. [14] proposed an image and text file encryption technique using chaos. This algorithm is also implemented in FPGA for real-time applications. Babaei [15] proposed an image and text encryption technique using DNA and chaos. In this technique, researcher used One-Time-Pad (OTP) algorithm to perform encryption operations. In theoretical point of view, OTP algorithm provides better security but in practical, it is very difficult to generate and distribute OTP-based keys. Murillo-Escobar et al. [1] proposed a novel text encryption technique using logistic maps. This algorithm uses 128-bit external secret key and two numbers of logistic maps to perform text encryption operation. The algorithm provides security in just one-round of permutation and diffusion operations.

Aforementioned text encryption techniques have one common problem which is that, to encrypt multiple text files, algorithms should be executed a number of times equal to the number of text files. This problem has been properly addressed in this proposed work.

In this paper, we have proposed a secure multiple text file encryption technique, wherein text files of equal as well as unequal sizes can be encrypted. The security analysis and computer simulations show that the proposed method performs better in terms of security, efficiency, and feasibility in the encryption of multi-sized multiple text files.

The proposed work contributes the followings.

- To get better encryption efficiency, better security, and also to make the algorithm both software and hardware effective, only piece-wise linear chaotic map (PWLCM) systems are utilized in this technique.
- Block-based diffusion and permutation are the only major operations performed to make the algorithm time-efficient.

- Again to make the algorithm time-efficient, one-round permutation and diffusion operations are performed to encrypt multiple text files.
- Hash-based keys are used to protect the algorithm against known-plaintext attack (KPA) and chosen-plaintext attack (CPA).

The rest of the sections are arranged in the following way. The chaotic map is described in Sect. 2. Section 3 describes the proposed methodology. The security analysis and the computer simulations are presented in Sect. 4. Finally, the paper concludes in Sect. 5.

2 Piece-Wise Linear Chaotic Map

In most of the encryption algorithms, PWLCM systems are used to generate better pseudo-random chaotic sequences. PWLCM systems have certain good characteristics such as simplicity, excellent ergodicity, efficient implementation, and good dynamical behavior [16–18]. The PWLCM system is represented as,

$$p_{i+1} = F_c(p_i) = \begin{cases} p_i/c & 0 \le p_i < c \\ (p_i - c)/(0.5 - c) & c \le p_i < 0.5 \\ F_c(1 - p_i) & 0.5 \le p_i < 1 \end{cases} \tag{1}$$

where the control parameter, $c \in (0, 0.5)$ and the initial value, $p \in (0, 1)$.

3 Proposed Methodology

3.1 Key Generation Operation

Followings are the steps for key generation operation.

Step 1 Take k number of original text files of any size. Let $T1, T2, T3, \ldots, Tk$ be the k number of text files having sizes $M1, M2, M3, \ldots, Mk$ bytes (or, ASCII symbols), respectively.

Step 2 Keep all the text files in an array denoted as 'tcell'. The array 'tcell' is defined as,

$$\text{tcell} = \{a1, a2, a3, \ldots, ak\} \tag{2}$$

where $a1, a2, a3, \ldots, ak$ stores the text files $T1, T2, T3, \ldots, Tk$ in read mode.

Step 3 Partition each of the text files of 'tcell' into the number of sections, each having 4-text elements. Reshape the text elements of each of the sections

into 2 × 2 blocks. Keep all the 2 × 2 sized text blocks in an array denoted as 't' and the remaining text elements (if any) in a separate array denoted as 'remcell'. The proposed **Algorithm 1** presents the process of block division operation and block placement operation in the array 't'.

Algorithm 1

1. Set $s \leftarrow 1$ where s is an incremented variable.
2. Input k number of text files from the array 'tcell'.
3. Repeat step-4 to step-12 for k number of text files.
4. Place the elements of text file in an array denoted as $'at'$.
$$at \leftarrow \text{tcell}\{k\}$$
5. Convert each of the ASCII character elements of $'at'$ into *double*.
$$at \leftarrow \text{double}(at)$$
6. Calculate the number of elements of an array $'at'$. Let it is denoted as $'n'$.
7. Store the value of $'n'$ in an array denoted as $'n1'$.
$$n1(k) \leftarrow n$$
8. Store the size of each of the text files in an array denoted as 'sizearr'.
$$\text{sizearr}(k,:) \leftarrow size(\text{tcell}\{k\})$$
9. Set $i \leftarrow 1$. Repeat step-10 to step-12 for $\big(n - \text{mod}(n,4)\big)/4$ times.
10. Divide $'at'$ into 2×2 sized non-overlapping text blocks and place each of the text blocks in an array $'t'$.
$$t\{s\} \leftarrow [at(i{:}\,i+1); at(i+2{:}\,i+3)]$$
11. Increment the value of s by 1, $s \leftarrow s + 1$
12. Increment i by 4, $i \leftarrow i + 4$

Step 4 Convert the text block array 't' into its matrix form '$t1$' and then reshape the matrix '$t1$' into $m \times 4$ where m is the length of the text block array 't'.

Step 5 Generate 64 hex-values of the reshaped matrix '$t1$' through SHA-256 hash algorithm, which is expressed as,

$$\text{hash} = hx_1, hx_2, hx_3, \ldots, hx_{64}$$

Step 6 Generate the keys using the hash values, the given system parameters and initial values of PWLCM systems. The proposed algorithm uses two PWLCM systems (PWLCM-1 and PWLCM-2) to perform text encryption operation. The keys are generated by,

$$\begin{cases} \text{ita11} = \text{ita1} + \text{abs}\Big(\sin\Big(\pi \times \text{sum}\Big(\frac{hx_1:hx_{16}}{1000}\Big)\Big)\Big) \times 0.0096 \\ e11(1) = e1 + \exp\Big(-\frac{(\text{sum}(hx_{17}:hx_{32}))}{1000}\Big) \times 0.0085 \end{cases} \tag{2}$$

$$\begin{cases} \text{ita21} = \text{ita2} + \text{abs}\left(\cos\left(\pi \times \text{sum}\left(\frac{hx_{33}:hx_{48}}{1000}\right)\right)\right) \times 0.0093 \\ e21(1) = e2 + \exp\left(-\frac{(\text{sum}(hx_{49}:hx_{64}))}{850}\right) \times 0.0079 \end{cases} \tag{3}$$

where (ita1, $e1$) and (ita2, $e2$) are the original system parameters and initial values of PWLCM-1 and PWLCM-2, respectively. (ita11, $e11(1)$) and (ita21, $e21(1)$) are the generated system parameters and initial values of PWLCM-1 and PWLCM-2, respectively.

3.2 *Encryption Operation*

Figure 1 shows the text encryption system. The encryption steps are as given below.

Step 1 Repeat **Step 1**–**Step 3** of Sect. 3.1.

Step 2 Keep the number of remaining text elements, $(n - \text{mod}(n, 4) + 1{:}n)$ of each of the text files in the array 'remcell'. Here, 'n' is the number of text elements of the text files. The array 'remcell' is expressed as,

$$\text{remcell} = [\text{remcell}\{1\}, \text{remcell}\{2\}, \text{remcell}\{3\}, \ldots, \text{remcell}\{k\}].$$

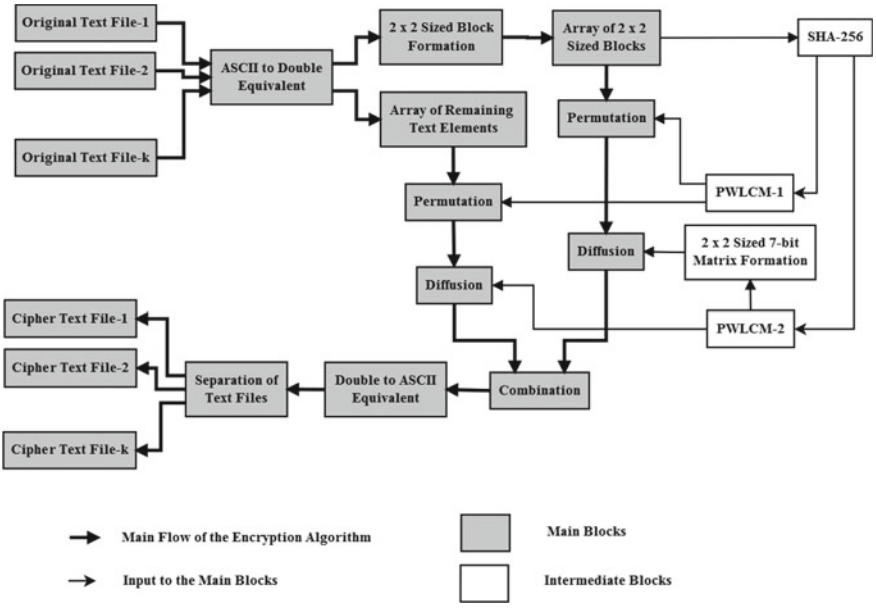


Fig. 1 Text encryption system

where remcell{1}, remcell{2}, remcell{3}, and remcell{k} contains the remaining text elements of T1.txt, T2.txt, T3.txt, and Tk.txt text files, respectively.

Step 3 According to **Step 6** of Sect. 3.1, generate $e11(1)$ and ita11 of PWLCM-1 and also generate $e21(1)$ and ita21 of PWLCM-2.

Step 4 Iterate the PWLCM-1 of Eq. (1) for m times where 'm' is the length of the block array 't'. The generation of PWLCM-1 chaotic sequence is,

$$e11 = (e11(1), e11(2), e11(3), \ldots, e11(m))$$

Step 5 Sort $e11(1:m)$ and $e11$(1:remsize) in ascending order where 'remsize' is the length of the array. The sorting functions are as follows.

$$\begin{cases} [e11\text{sort}b, e11\text{index}b] = \text{sort}(e11(1{:}m)) \\ [e11\text{sort}t, e11\text{index}t] = \text{sort}(e11(1{:}\text{remsize})) \end{cases}$$

where $e11\text{sort}b$ and $e11\text{sort}t$ are the sorting order sequences of $e11(1{:}m)$ and $e11$(1:remsize), respectively. $e11\text{index}b$ and $e11\text{index}t$ are the index values of the sorting sequences $e11\text{sort}b$ and $e11\text{sort}t$, respectively.

Step 6 According to the index value $e11\text{index}b$, shuffle the blocks of the array 't'. Similarly, according to the index value $e11\text{index}t$, shuffle the text elements (if any) of the array 'remcell'. Let the shuffled text block array and the shuffled remaining text element array are denoted as 'st' and 'sremcell', respectively.

Step 7 Iterate the PWLCM-2 3+remsize times to diffuse the blocks of the array 'st' and the text elements of the array 'sremcell'. The generation of PWLCM-2 chaotic sequence is as follows.

$$e21 = (e21(1), e21(2), \ldots, e21(3 + \text{remsize}))$$

Step 8 Sort $e21$ in ascending order by,

$$[e21\text{sort}, e21\text{index}] = \text{sort}(e21)$$

where, $e21\text{sort}$ and $e21\text{index}$ are the sorting and indexing sequences, respectively.

Step 9 Preprocess the $e21\text{index}$ values to generate a 2×2 diffusion matrix to diffuse the blocks of the array 'st' and also generate diffusion text elements to diffuse the text elements of the array 'sremcell'. The pre-processing operation is as follows.

$$\begin{cases} e21\text{index} = \text{mod}\big(\text{mod}\big(e21\text{index} \times 9^{10}, 128\big) \times 9^{9}, 128\big) \\ e22\text{index} = \text{double}(\text{reshape}(e21\text{index}(1:4), [2, 2])) \end{cases}$$

Step 10 Execute bit-XOR operation between the blocks of the array 'st' and the 2×2 matrix 'e22index'. Let the diffused blocks are stored in an array 'dst'. The bit-XOR operation is as follows.

$$\begin{cases} \text{dst}(1) = e22\text{index} \oplus \text{st}(1) \\ \text{dst}(2) = \text{dst}(1) \oplus \text{st}(2) \\ \text{dst}(3) = \text{dst}(2) \oplus \text{st}(3) \\ \qquad\qquad \vdots \\ \text{dst}(m) = \text{dst}(m-1) \oplus \text{st}(m) \end{cases}$$

Execute one-to-one bit-XOR operation between the shuffled text elements of the array 'sremcell' and the pre-processed text elements of $e21\text{index}(5 : 4 + \text{remsize})$. Let the diffused text elements are placed in an array denoted as 'dsremcell'. The text element diffusion process is as follows.

$$\begin{cases} \text{dsremcell}(1) = e21\text{index}(5) \oplus \text{sremcell}(1) \\ \text{dsremcell}(2) = e21\text{index}(6) \oplus \text{sremcell}(2) \\ \qquad\qquad \vdots \\ \text{dsremcell}(\text{remsize}) = e21\text{index}(4 + \text{remsize}) \oplus \text{sremcell}(\text{remsize}) \end{cases}$$

Step 11 Return the diffused text elements to their corresponding text files. Let 'rdsremcell' is an array containing the returned diffused text elements. The array 'rdsremcell' is defined as,

$$\text{rdsremcell} = \begin{bmatrix} \text{rdsremcell}\ \{1\}, \text{rdsremcell}\ \{2\}, \text{rdsremcell}\{3\}, \ldots, \\ \text{rdsremcell}\ \{k\} \end{bmatrix}$$

where rdsremcell {1}, rdsremcell{2}, rdsremcell{3}, rdsremcell{k} contain the diffused text elements of T1.txt, T2.txt, T3.txt, and Tk.txt text files, respectively.

Step 12 Form k number of encrypted text files. The **Algorithm 2** presents the process of forming k number of encrypted text files.

Algorithm 2

1. Set $q \leftarrow 1$ where q is an incremented variable.
2. Let k is the number of text files as an input to the algorithm.
3. Repeat step-4 to step-10 for k number of text files.
4. Take the size of each of the k number of text files (output of step-8 of **Algorithm 1**).
   $M \leftarrow$ sizearr $(k, 2)$
5. Set $i \leftarrow 1$. Repeat step-6 to step-9 for $\big(M - \mathrm{mod}(M, 4)\big)/4$ times.
6. Place the $\times$ sized diffused text blocks of the array $'\mathrm{dst}'$ into a matrix $'\mathrm{tmp}'$.
   $\mathrm{tmp} \leftarrow \mathrm{dst}\{q\}$
7. Perform $\mathrm{das}(i: i+3, :) \leftarrow [\mathrm{tmp}(1:2)\ \mathrm{tmp}(3:4)]$
8. Increment q by 1, $q \leftarrow q + 1$
9. Increment i by 4, $i \leftarrow i + 4$
10. Keep the matrix $'\mathrm{das}'$ in an array $'\mathrm{tcell1}'$.
    $\mathrm{tcell1}\{k\} \leftarrow \mathrm{das}$
11. Repeat step-12 to step-17 for k number of text files.
12. Again take the size of each of the k number of text files. (output of step-8 of **Algorithm 1**)
    $M \leftarrow$ sizearr $(k, 2)$
13. Perform $as \leftarrow \mathrm{tcell1}\{k\}$.
14. Follow step-14 to step-17, if the condition $(\mathrm{mod}(M, 4) \sim = 0)$ satisfies for the text files.
15. Keep the remaining diffused text elements of the text files in a memory element denoted as $'\mathrm{con1}'$.
    $\mathrm{con1} \leftarrow$ rdsremcell$\{k\}$
16. Perform $as(M - \mathrm{mod}(M, 4) + 1: M) \leftarrow \mathrm{con1}$
17. Perform $\mathrm{tcell1}\{k\} \leftarrow as$

Step 19 Convert the elements of tcell1$\{k\}$ into ASCII symbols.

The steps of decryption operation are the reverse of the steps of encryption operation.

4 Security Analysis and Computer Simulation results

In this system, four arbitrary-sized text files from the article [19] are used to encrypt and decrypt the text elements, simultaneously. The text elements of text files are the set of symbols in the ASCII format. The given text files are shown in Fig. 2 and the size (or, length) of all the text files are shown in Table 1. The text files are simulated in MATLAB R2012a. System parameters, initial values of PWLCM systems, and

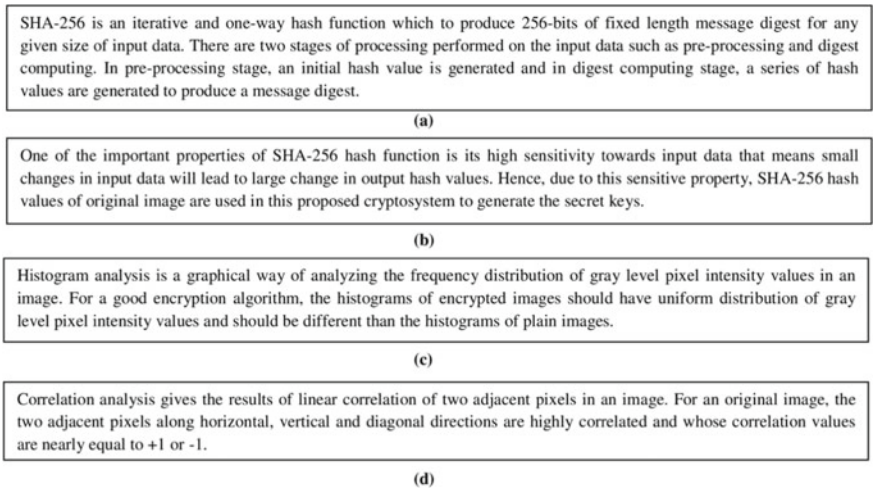


Fig. 2 Original **a** T1.txt file, **b** T2.txt file, **c** T3.txt file, **d** T4.txt file

Table 1 Size and length of the text files

S. No.	Text files	Size (bytes)	Length (No. of ASCII symbols)
1	T1.txt	409	409
2	T2.txt	334	334
3	T3.txt	328	328
4	T4.txt	277	277

the hash value of combined multiple text files are shown in Table 2. Figures 3 and 4 shows the encryption and decryption outputs of the text encryption system. The encryption results show that the proposed cryptosystem encrypts any number of text files of arbitrary sizes. The decryption results show that the proposed cryptosystem decrypts the encrypted text files using the right secret keys. It shows that the proposed scheme is effective for encryption and decryption of multi-sized multiple text files.

The security analysis based on the simulation results are as follows.

Table 2 Initial keys of the encryption/decryption system

Secret keys	Initial values	System parameters
PWLCM-1	$e1 = 0.215663478737535$	$ita1 = 0.317455312945677$
PWLCM-2	$e2 = 0.213891157342188$	$ita2 = 0.318573715124739$
Hash value (64-hex values)	'85241BEAEF498EA203CB793C83F37801 0F35F0F4785CE38E4E5092145CA88CE4'	

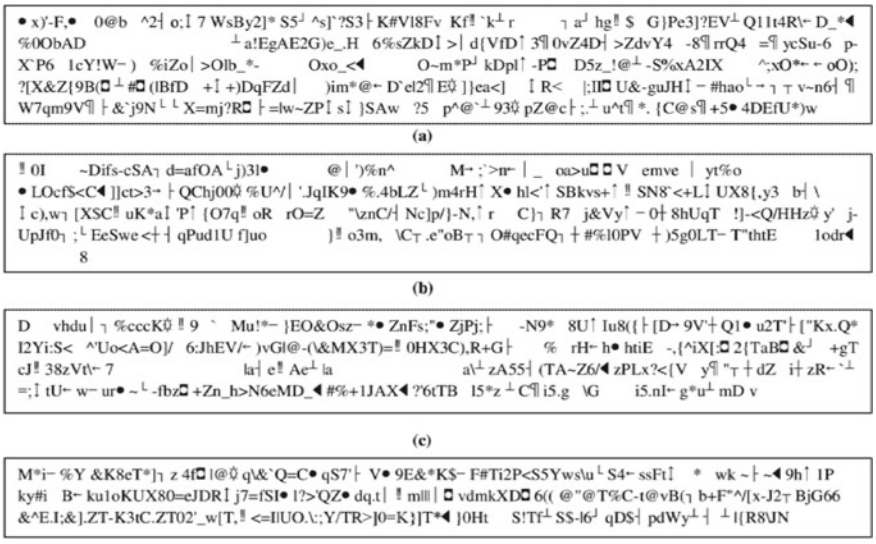


Fig. 3 Simulation outputs: encrypted **a** T1.txt file, **b** T2.txt file, **c** T3.txt file, **d** T4.txt file

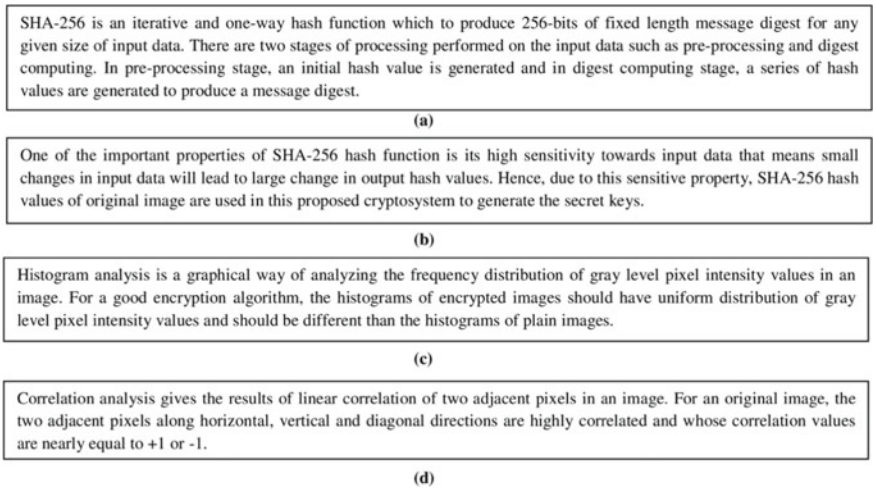


Fig. 4 Simulation outputs: decrypted **a** T1.txt file, **b** T2.txt file, **c** T3.txt file, **d** T4.txt file

4.1 Key Space Analysis

Larger is the key space to 2^{128}, higher is the resistance to brute-force attack [20, 21]. The keys used in this algorithm are as follows.

- System parameters ita1 and ita2 of PWLCM-1 and PWLCM-2, respectively; initial values $e1$ and $e2$ of PWLCM-1 and PWLCM-2, respectively.
- 256-bit hash values.

The proposed work uses a key space of 10^{15} for each of the individual keys of PWLCM systems. This is because 64-bit floating-point standard is used by the algorithm and the IEEE suggested a key space of 10^{15} for the 64-bit floating-point standard [22]. In addition to this, the proposed scheme uses a key space of 2^{128} for the "Secure Hash Algorithm SHA-256" for resisting the best attack in that algorithm. Therefore, the overall key space is $\left(10^{15} \times 10^{15}\right) \times \left(10^{15} \times 10^{15}\right) \times 2^{128} = 1.2446 \times 2^{327} \left(> 2^{128}\right)$, which to resist the brute-force attack efficiently.

4.2 Histogram Analysis (or Frequency Analysis)

Histogram is a graph that shows the frequency distribution of all the ASCII symbols of a text file. The algorithm is resistive to frequency attack when all the ASCII symbols are uniformly distributed in a ciphertext. In addition to this, the histogram of a ciphertext should be different from the histogram of a plaintext. Figure 5 shows the histogram results of the multiple text file encryption system. In histogram figures, x-axis represents the decimal values of all the ASCII symbols of an ASCII table and y-axis represents the frequency distribution of the corresponding ASCII symbols in the ciphertext. By observing the results, we can realize that all the ASCII symbols in cipher text files (Fig. 5b, e, h, k) are uniformly distributed and also we can see that the histograms of plaintext files (Fig. 5a, d, g, j) are totally different from the histograms of ciphertext files. This concludes the high resistivity of frequency attack using the proposed scheme.

4.3 Information Entropy Analysis

Information entropy measures the probability of occurrence of all the ASCII symbols in a text file. Higher the closeness of probabilities of ASCII symbols in a text file, larger is the information entropy. In this work, we are using 128 ASCII symbols. Hence, for an ideal text encryption system, the probability of each of the ASCII symbols is 1/128. Therefore, the information entropy of an ideal system is,

$$\begin{aligned} H(t) &= -\sum_{i=0}^{127} p(t_i) \log_2 p(t_i) \\ &= -128 \times \frac{1}{128} \log_2 \frac{1}{128} = -\left(-\log_2 128\right) = 7 \end{aligned} \tag{4}$$

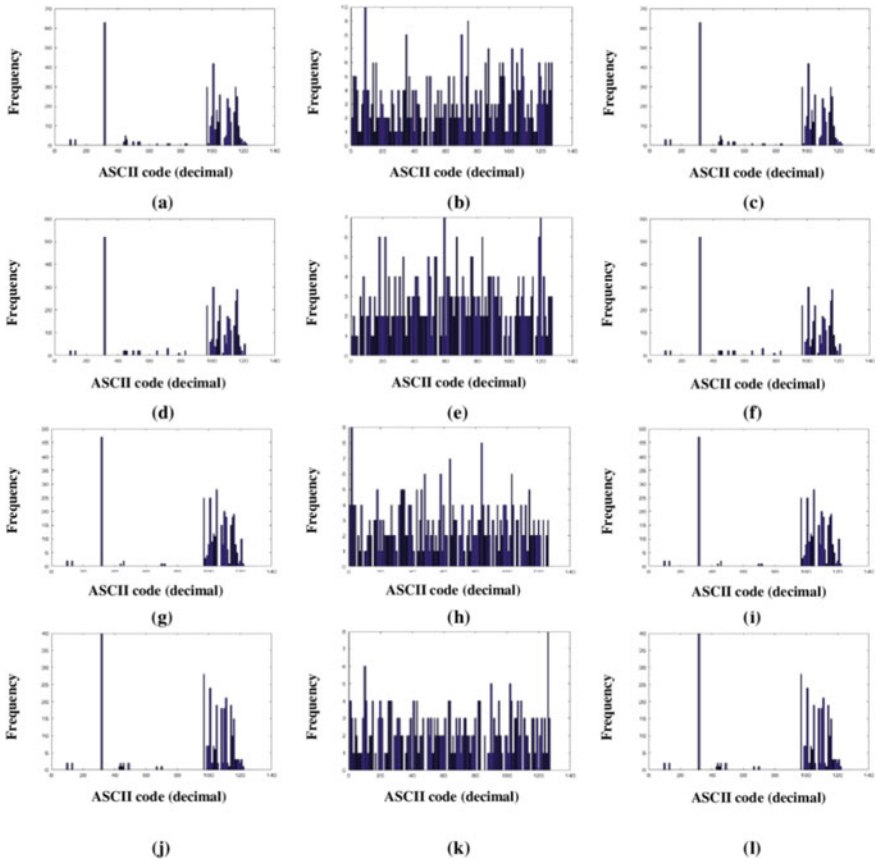


Fig. 5 Histograms of **a** original T1.txt file, **b** encrypted T1.txt file, **c** decrypted T1.txt file, **d** original T2.txt file, **e** encrypted T2.txt file, **f** decrypted T2.txt file, **g** original T3.txt file, **h** encrypted T3.txt file, **i** decrypted T3.txt file, **j** original T4.txt file, **k** encrypted T4.txt file, **l** decrypted T4.txt file

where $p(t_i)$ is the probability and $H(t)$ is the entropy of the information source t. The entropy results of various text files are as shown in Table 3. The results show that all the encrypted text outputs have information entropy values very close to the ideal value. This proves that the ASCII symbols have appeared with almost the

Table 3 Information entropy results of multiple text files

S. No	Text files	Original text files	Encrypted text files
1	T1.txt	4.3099	6.8502
2	T2.txt	4.8760	6.8872
3	T3.txt	4.4066	6.8764
4	T4.txt	4.6727	6.8533
Average		4.5663	6.8668

same probability in encrypted text files. Hence, the proposed text encryption system strongly resists entropy attack.

4.4 Differential Attack Analysis

In differential attack, attackers try to capture some useful information from the ciphered text to make a relationship between the plaintext and the corresponding ciphered text. This is done by first changing a random text element in an original text file and then applying the proposed encryption algorithm on this changed text file to get a new ciphertext file. Finally, attackers compare the two text files (original ciphered text file and new ciphered text file) to realize the relationship between the original text file and the ciphered text file. "Unified Average Changing Intensity (UACI)" and "Number of Pixel Change Rate (NPCR)" are the two measures commonly used to measure the effectiveness of differential attack [23].

Table 4 shows the average UACI and NPCR results of multi-sized multiple text files by using the proposed encryption technique. In this table, the average UACI and NPCR results are calculated by changing the random text elements 100 times in each of the text files. The results indicate that the proposed text encryption system is highly sensitive to all the plaintexts with respect to only one text element change in one text file.

4.5 Known-Plaintext Attack (KPA) and Chosen-Plaintext Attack (CPA) Analysis

Basically, most of the algorithms have been broken by these two types of attacks. But, the proposed scheme is resisted by these attacks. In this work, the chaotic map (PWLCM system)-based keys are generated from the combination of given key values and the hash values (256-bit) of the plaintext. So, when the plaintext changes, its key value is also change. For different key values, different chaotic sequences are produced and also different ciphertexts are obtained. Hence, attackers cannot obtain ciphertexts by selecting some known plaintexts. This concludes the strong resistivity of KPA and CPA using the proposed scheme.

4.6 Computational Time Analysis

A cryptosystem should have less computational time (encryption and decryption time) when used in real-time applications. The computational analysis is done in a system with 3.40 GHz processor, 4 GB of RAM, and 64-bit operating system. For the

Table 5 NPCR and UACI results

S. No.	Text files	Size (bytes)	Text element changed in T1.txt file		Text element changed in T2.txt File	
			NPCR (ideal value: approx. 100%)	UACI (ideal value: approx. 34%)	NPCR (ideal value: approx. 100%)	UACI (ideal value: approx. 34%)
1	T1.txt	409	99.2249	33.4819	99.1932	33.5436
2	T2.txt	334	99.0599	33.3770	99.0629	33.6623
3	T3.txt	328	99.2134	33.1493	99.1585	33.1783
4	T4.txt	277	98.9639	33.0161	98.8809	32.6654
Average			99.1155	33.2547	99.0739	33.2624

S. No.	Text files	Size (bytes)	Text element changed in T3.txt file		Text element changed in T4.txt file	
			NPCR (ideal value: approx. 100%)	UACI (ideal value: approx. 34 %)	NPCR (ideal value: approx. 100%)	UACI (ideal value: approx. 34 %)
1	T1.txt	409	99.2249	33.4819	99.1932	33.5436
2	T2.txt	334	99.0599	33.3770	99.0629	33.6623
3	T3.txt	328	99.2134	33.1493	99.1585	33.1783
4	T4.txt	277	98.9639	33.0161	98.8809	32.6654
Average			99.1155	33.2547	99.0739	33.2624

analysis, different-sized multiple text files are used as inputs to the algorithm. Table 5 shows the result of total encryption and decryption time to encrypt and decrypt multi-sized multiple text files using the proposed cryptosystem. The results show that the proposed system takes very less time to encrypt and decrypt arbitrary-sized multiple text files.

4.7 Comparison Analysis

The comparisons of the proposed method with the method in [1] are as follows. The algorithm in [1] uses 95 printable ASCII symbols to form 1126 characters (abstract of the said article) as a plaintext.

- The proposed method has capability to encrypt and decrypt either a single text file or more than one text files of arbitrary sizes in a single execution of the algorithm. However, the techniques in [1, 14, 15] used only a single text file at a time to encrypt and decrypt their text elements.

- The total key space of the proposed algorithm (1.2446×2^{327}) is larger than 2^{128} and also better than the key space result of the algorithm in [1] (2^{128} or 2^{256}) to resist brute-force attack efficiently.
- The total encryption and decryption time of the proposed algorithm to encrypt and decrypt 1126 symbols (abstract section of Ref [1]) are 0.004284 s and 0.003285 s, respectively. On the other side, the total encryption and decryption time using the algorithm in Ref [1] are 0.140 s and 0.109 s, respectively. This reveals that the proposed algorithm performs faster than the algorithm in [1].
- The entropy of the encrypted text output using the proposed algorithm and the algorithm in [1] are 6.9113 (ideal value = 7) and 6.48 (ideal value = 6.56), respectively. The proposed algorithm deviates 1.267% from the ideal value and the algorithm in [1] deviates 1.220% from the ideal value. So, both the information entropy values are very close to the ideal values; hence, both the algorithms resist entropy attack efficiently.
- Both the algorithms resist CPA and KPA and are also sensitive to the keys.
- In both the algorithms, the ASCII symbols are uniformly distributed in the cipher text files (shown in histogram images). This proves that the proposed algorithm prevents histogram attack or frequency attack.

5 Conclusion

This paper proposes a text file encryption technique using PWLCM systems. The proposed technique uses arbitrary-sized multiple text files to perform encryption operation. The technique performs permutation operation on 2×2 sized text blocks and remaining text elements using a PWLCM system. Then, it performs diffusion operation on both text blocks and remaining text elements using another PWLCM system. The computer simulation shows the good encryption and decryption results of the proposed cryptosystem. The security analyses show the uniformity of ASCII symbols in the histogram of cipher text files. Also there is stronger information entropy and better NPCR and UACI. In addition, it also protects both CPA and KPA. The comparison analyses show that the proposed algorithm is applicable to any size of multiple text files and produces better security as compared to other reported works.

References

1. Murillo-Escobar MA, Abundiz-Pérez F, Cruz-Hernández C, López-Gutiérrez RM (2014) A novel symmetric text encryption algorithm based on logistic map. In: Proceedings of the international conference on communications, signal processing and computers, pp 04953, Feb 2014
2. Coppersmith D (1994) The data encryption standard (DES) and its strengths against attacks. IBM J Res Dev 38(3):243–250

3. Pub NF (2001) 197: Advanced encryption standard (AES). In: Federal Information Processing Standards Publication, vol 197, pp 441–0311
4. Boneh D, Dunworth C, Lipton R (1995) Breaking DES using a molecular computer. In: Proceedings of DIMACS workshop on DNA computing, pp 37–65
5. Cui G, Li C, Li H, Li X (2009) DNA computing and its application to information security field. In: Fifth international conference on natural computation (ICNC'09) IEEE, vol 6, pp 148–152, Aug 2009
6. Wang XY, Zhang YQ, Zhao YY (2015) A novel image encryption scheme based on 2-D logistic map and DNA sequence operations. Nonlinear Dyn 82:1269–1280
7. Ye G, Zhao H, Chai H (2016) Chaotic image encryption algorithm using wave-line permutation and block diffusion. Nonlinear Dyn 83:2067–2077
8. Wang X, Liu C, Zhang H (2016) An effective and fast image encryption algorithm based on Chaos and interweaving of ranks. Nonlinear Dyn 84:1595–1607
9. Li C, Luo G, Qin K, Li C (2017) An image encryption scheme based on chaotic tent map. Nonlinear Dyn 87:127–133
10. Ullah A, Jamal SS, Shah T (2017) A novel construction of substitution box using a combination of chaotic maps with improved chaotic range. Nonlinear Dyn 88:2757–2769
11. Patro KAK, Acharya B (2019) A simple, secure, and time-efficient bit-plane operated bit-level image encryption scheme using 1-D chaotic maps. In: Innovations in soft computing and information technology. Springer, Singapore, pp 261–278
12. Patro KAK, Acharya B, Nath V (2019) A secure multi-stage one-round bit-plane permutation operation based chaotic image encryption. Microsyst Tech: 1–8
13. Patro KAK, Acharya B (2019) An efficient colour image encryption scheme based on 1-D chaotic maps. J Inf Secur Appl 46:23–41
14. Sobhy MI, Shehata AE (2001) Chaotic algorithms for data encryption. IEEE Int Conf Acoust Speech Signal Process 2:997–1000
15. Babaei M (2013) A novel text and image encryption method based on chaos theory and DNA computing. Nat Comput 12(1):101–107
16. Zhang X, Wang X (2017) Multiple-image encryption algorithm based on mixed image element and permutation. Opt Laser Eng 92:6–16
17. Xiang T, Liao X, Wong K-W (2007) An improved particle swarm optimization algorithm combined with piecewise linear chaotic map. Appl Math Comput 190(2):1637–1645
18. Wang X, Xu D (2014) A novel image encryption scheme based on Brownian motion and PWLCM chaotic system. Nonlinear Dyn 75:345–353
19. Patro KAK, Acharya B (2018) Secure multi–level permutation operation based multiple colour image encryption. J Inf Secur Appl 40:111–133
20. Kulsoom A, Xiao D, Abbas SA (2016) An efficient and noise resistive selective image encryption scheme for gray images based on chaotic maps and DNA complementary rules. Multimed Tools Appl 75:1–23
21. Patro KAK, Banerjee A, Acharya B (2017) A simple, secure and time efficient multi-way rotational permutation and diffusion based image encryption by using multiple 1-D chaotic maps. In: international conference on next generation computing technologies. Springer, Singapore, pp 396–418
22. Floating-Point Working Group (1985) IEEE standard for binary floating-point arithmetic. ANSI. IEEE Std., pp 754–1985
23. Enayatifar R, Hanan A, Isnin IF (2014) Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence. Opt Lasers Eng 56:83–93

Relative QoS Investigation of VoIP Over LTE Networks

Smita Avinash Lonkar and K. T. V. Reddy

Abstract The Voice over the Internet Protocol (VoIP) for international calls to either family or friends or business is the most common form of communication in the last period. Since the Internet is easily accessible to maximum people, VoIP is a modest way to lower expensive fares. However, because of the delay and low speed, VoIP systems over mobile networks can achieve higher data rates than the regular 3G network. In order to deal with such high data speeds, greater capacity and less latency, long term evolution (LTE) were created. VoIP is required to allow LTE network voice calls. This paper looks at VoIP service quality (QoS) through the LTE network by providing two case studies on the topology of the LTE network. Relative VoIP QoS analysis is carried out by parameter measurement.

Keywords LTE · VoIP · QoS

1 Introduction

Over recent years, we have seen the incredible rise in the use of smartphone and mobile data. This increase in use is causing a huge increase in mobile network Internet traffic.

LTE is standard for mobile and data terminals to handle high speed wireless communication. It is also the first fully IP-based 3GPP wireless standard. It offers up to 50 Mbps on the uplink (UL), and downlink (DL) at 100 Mbps [1]. Engineers are presented with a critical problem to improve, automate and audit voice traffic over

S. A. Lonkar (✉)
U.M.I.T., S.N.D.T. Women's University, Mumbai, India
e-mail: smitalonkar@gmail.com

K. T. V. Reddy
Director-Technical, PRES, Loni, India
e-mail: ktvreddy@gmail.com

Imm. Past President, IETE HQ, New Delhi, India

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_8

data networks. Based on the study of virtual traffic or actual terms, several schemes were suggested.

In [2], they addressed how to handle the most popular VoIP codecs on a Wi-MAX network utilizing different service groups. The different types of network groups were contrasted to QoS. The performance of unified LAN/WAN wireless VoIP scenarios is tested using an optimization framework in [3]. Various QoS parameters are calculated and the proper VoIP codec was suggested on the basis of this evaluation.

The parameters that affect VoIP's quality through heterogeneous networks, such as Wi-Fi, Wi-MAX, and beyond that, can be identified by using the OPNET simulation tool and are evaluated in [4] for QoS performance simulation through UMTS and Wi-Fi (IEEE 802.11x) alone [5]. VoIP service quality (QoS) is analyzed on the LTE network in [6]. The performance of the test is simulated by measuring parameters and different scenarios.

The accompanying paper examines comparative QoS for VoIP over the LTE network utilizing bandwidth, jitter, end-to-end latency and packet failure measurements.

The paper is organized in this way. The LTE design is listed in Sect. 2. Section 3 explains LTE network topology simulation. Section 4 discusses VoIP QoS specifications. The results and conclusions of Sect. 5 are published.

2 LTE Architecture

The LTE network is designed from three main components. Two different scenarios are applied in case studies to compare the performance of QoS. The LTE network planning is examined and simulated. One-to-one client initialization is done first and then the team link is implemented on track the quality improvement.

The Evolved Universal Terrestrial Radio Access Network is connected to a user equipment (E-UTRAN), which is subsequently connected to the evolved core (EPC) packets and then to the Internet.

The planning of the LTE network is put up from three main parts. A user equipment (UE) is linked with the Evolved Universal Terrestrial Radio Access Network (E-UTRAN), which is then linked to the evolved packet core (EPC) then to the Internet. The links are as shown in Fig. 1.

As shown in Fig. 1, user equipment (UE) is a mobile phone that is used directly by a user to talk. It is connected to the evolved NodeB (eNodeB). The E-UTRAN attaches each UE to the LTE network. It contains eNodeBs which can help many cells in the LTE network. They are interlocked together via the X2 interface as well as connected to the EPC by the S1 interface.

Five main nodes are included in EPC: the serving gateway (SGw), packet data network (PDN), the home subscriber station (HSS) and mobility management entity (MME). Primary network of EPC is the evolving packet core (EPC) and involves many devices. The SGW functions as a router to direct data packets to eNodeBs between users. The inter-eNodeB handover also functions as an anchor. The PDN

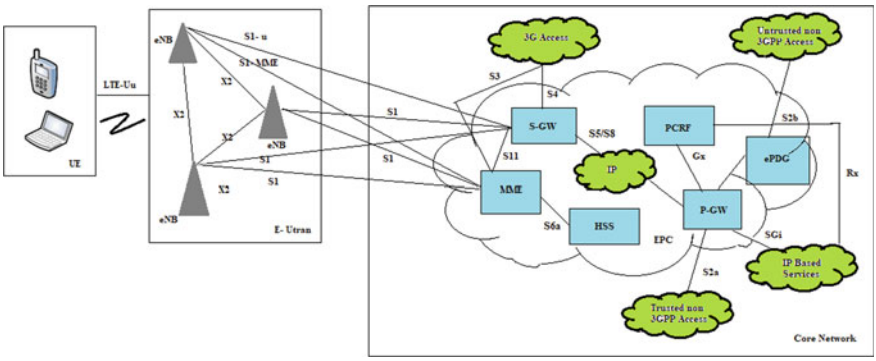


Fig. 1 LTE architecture connection [7]

connects to the Internet the LTE network. Services such as packet filtering, traffic shaping and user loading are controlled, and the GPRS tunneling protocol (GTPs) are established, maintained and deleted for the SGW. MME shakes EU conditions including UE name, phone status and protection definitions. MME shakes client data via HSS.

3 LTE Network Topology

Two instances are being examined for LTE network topology. The LTE network design as shown in Fig. 2 reflects the topology of the first case. This model is made up of 2 servers, 3 AGWs, 3 improved base transceivers (EBTs) and 9 consumer equipment (UEs). Two AGWs are available on server 0. There is an EBT in each AGW. EBT (0) has two users and EBT (1) has three users. There is 1 AGW and 1 EBT for 4 clients on Database 1.

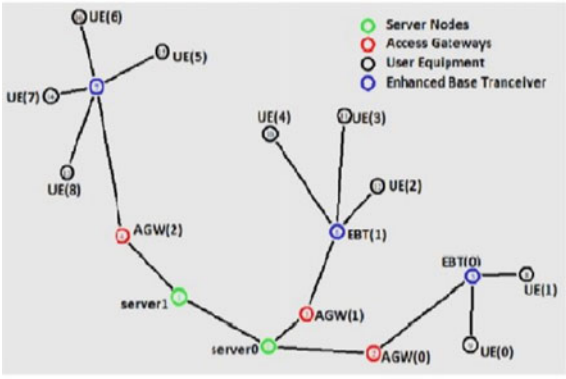


Fig. 2 LTE network topology: case 1 [6]

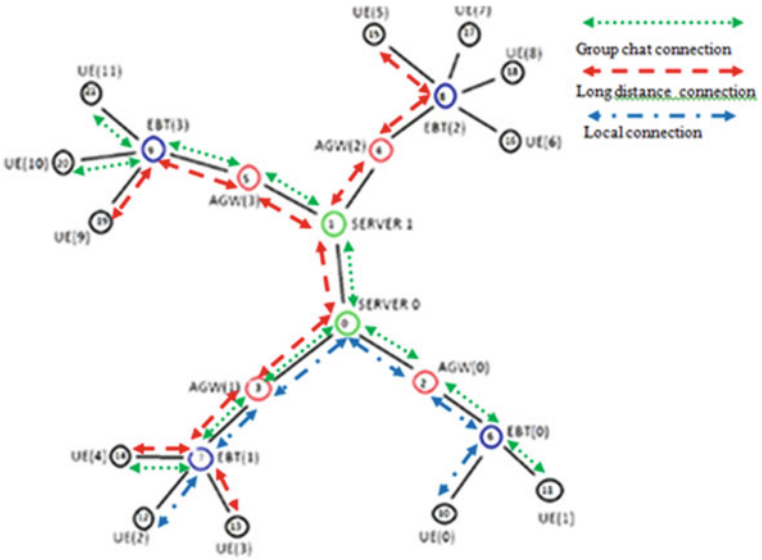


Fig. 3 LTE network topology: case 2

The LTE network model as shown in Fig. 3 topology for the second case. The model is made up of two servers, 4 access gateways (AGWs), 4 enhanced EBTs and 12 user devices. Two AGWs are deployed on database 0. Each AGW is fitted with EBT. EBT (0) contains two UEs and EBT (1) three UEs. The EBT (2) and 4 UE servers are AGW (3) and AGW (3) servers are EBT (3) and three UEs.

Servers obtain incoming and outgoing information and essential feedback. AGW transmits data to the servers so that the data can reach them. EBTs provide mobile devices with signals to allow the Internet to interchange data. The general EBT is called the LTE eNodeB. UE is any device for creating inter-user communication.

The LTE setup on NS2 provides the connections between AGWs and EBTs as well as the wireless connection between UEs. For linking the part, duplex ties are used, as each attachment is twofold. There is a long delay between two databases of this topology. This means a long-distance link, and the AGW is interrupted in a small amount of time. In this example, VoIP is introduced using the UDP protocol. Both users 'tools for sending and receiving voice data' are connected to UDP agents and sinks.

For VoIP configuration, Cisco determines packet size and time between the transmission cycles. The average size of the packet is 480 bytes with an interval of 0.03 ms.

Two separate setups are to compare the performance of QoS from VoIP with LTE in two instances. One-to-one client situation will be performed first and then the quality improvement classes will be identified for group connection.

Two client pairs are used for single user cases in the LTE network topology case1 [6]. The first pair is a regional VoIP connection between the same server's two client

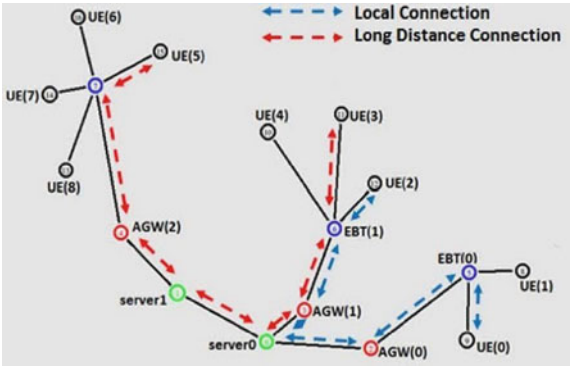


Fig. 4 Local connection and long-distance connection for LTE network topology case 1 [6]

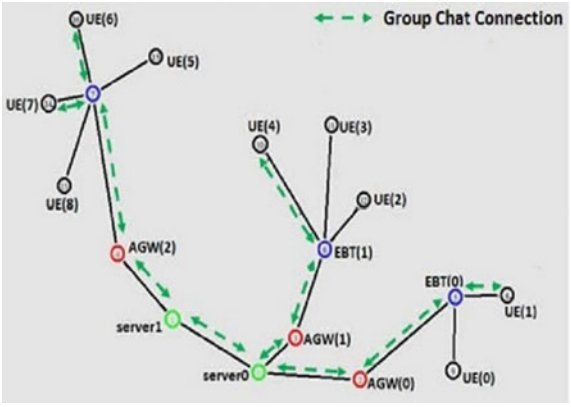


Fig. 5 Group connection for LTE network topology case 1 [6]

nodes. For this scenario, UE (0) and UE (2) are chosen for each AGW connected to Server0. This reference is shown in Fig. 4 in blue. The one-to-one contact simulation starts at 1.0 s and finishes at 19.0 s.

The second scenario is the long-distance VoIP communication on different servers between the two client nodes. For this case, UE (3) and UE (5) are chosen in server 1 where UE (3) is linked to server 0 and UE (5). This reference is shown in red in Fig. 4.

For the local chat link in LTE network topology case 1, community chat is introduced between 4 users: UE (1), UE (4), UE (6) and UE (7). In Fig. 5, this connection is a mix of local and long-distance users. This is shown in green. The chat interval is 20.0 to 39.0 s.

In LTE network topology case 2, two device pairs are used for one-to-one example. A regional VoIP call between two client nodes belonging to the same network is the first pair. UE (0) and UE (2), which are both linked by various AGW's to server 0. The one-to-one call simulation begins at 1.0 s and terminates at 19.0 s

The second scenario is a long-distance VoIP call between user nodes on different servers. For this scenario, the UE (4), UE (5) and UE (9) are selected, where UE (4)

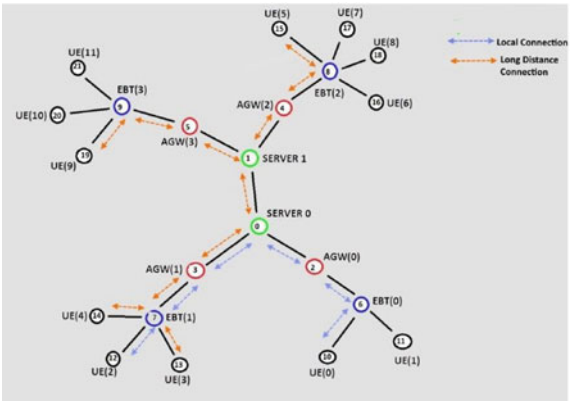


Fig. 6 Local connection and long-distance connection for LTE network topology case 2

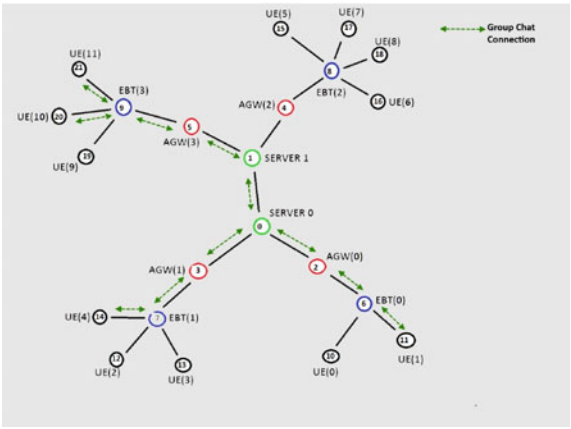


Fig. 7 Group connection for LTE network topology case 2

is connected to server 0, the server 0 and UE (5) and UE (9). The link appears in red in Fig. 6.

For LTE network communication link case 2, group chat between 4 users is implemented: UE (1), UE (4), UE (10) and UE (11). In Fig. 7, it is a mix of local and long-haul users. This connection is shown in green. The message interval in team communication varies between 20.0 and 39.0 s.

4 VoIP QoS Parameters

VoIP QoS is described by several factors including jitter, latency, lag from end to end and lack of packets. For each VoIP situation, they can analyze these variables.

Table 1 Lists network quality parameters below.

Table 1 VoIP quality parameter measures [6]

Network Parameter	Good	Acceptable	Poor
End-to-end delay (ms)	0–150.0	151–300	>300
Jitter (ms)	0–20.0	21–50	>50
Packet loss	0–0.5%	0.6–1.5%	>1.5%
Throughput (Mbps)	0–50.0	51–144	>144

5 Results and Discussion

CBR and UDP agent are implemented in VoIP. The loss monitor class of NS2 was used to perform all data collections. The sum of the packets missing, the amount and the number of bytes are determined from last arrival time of package. Such variables are shown with simulation tests.

End-to-End Delay

The spikes occur at the beginning of each scenario (1 and 21 ms) due to adjusting the containment window as shown in the simulatory graph in Fig. 8a. Delays from 2 to 20 ms are not stable and the delay can be tolerated at 30 ms.

The findings for case 2 appear in Fig. 8b. Overlaps of the two scenarios indicate also that there are no delay effects on each scenario, compared with Fig. 8a.

Performance: Performance is indicated by throughput which is the amount of data that can be transferred in a given amount of time from the source to the destination. The unit is bits per second for performance.

The throughput rate in Figs. 9a, b is as high as the plateaus. The simulation is accurate because only one speech (local and/or international) calls happens at the end. A much lower peak emerges at 20 s. The increase in data packets from the group chat, consisting of many nodes that send, receive and receive each other, evidently results in this plateau.

Three times the traffic was introduced by group chat, than single calls. Figure 9b shows the increase of data packets for one call, group chat and one call per group. The performance is 65 and 190 Mbps for group chat. This indicates that the simulation complies with theory.

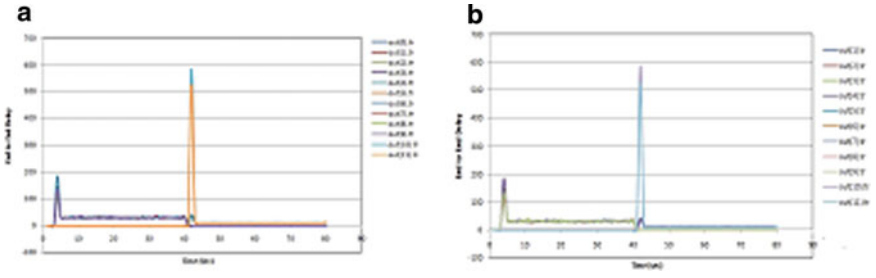


Fig. 8 **a** End-to-end delay case 1, **b** end-to-end delay case 2

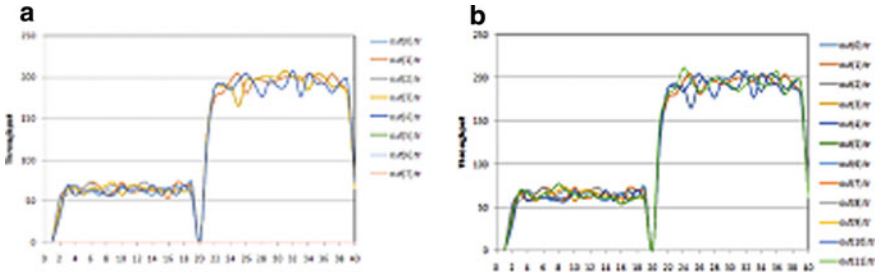


Fig. 9 **a** Throughput case 1, **b** throughput case 2

Packet loss: Packet loss is the percentage of packets lost through the network when they are transferred from source to destination. Nonetheless, a proportion of up to 1.5% is appropriate. Hopefully, there should be no packet failure.

Figure 10a, b simulation chart shows a significant loss in the packet when group chat is developed which goes back 20.0 s. These results are consistent, as there is considerably more pressure on wired and wireless networks for transmission of more packets once group chat sessions happen. All situations are also overlapped to show that the frequency of packet loss is not influenced by each scenario. For one call, the loss of a packet is low and increases with time in group chat.

Jitter: Jitter simply uses a latency folder for jitters and plot jitter map, as shown in Fig. 11. Like the latency diagram, at the end of each operation, jitter shows a number of peaks attributable to the change of the contents range, and at the other time, jitters are smooth and tolerable.

The results of network quality parameters for LTE network topology are summarized below in Table 2. The comparative analysis of QoS for VoIP over LTE network topology for both cases is done. The throughput shows increase in packets of data for one-to-one call and group chat for more number of user equipment. The significant packet loss occurs in both cases. The end-to-end delay falls in good range with more spikes at 1 and 21 ms in case 2.

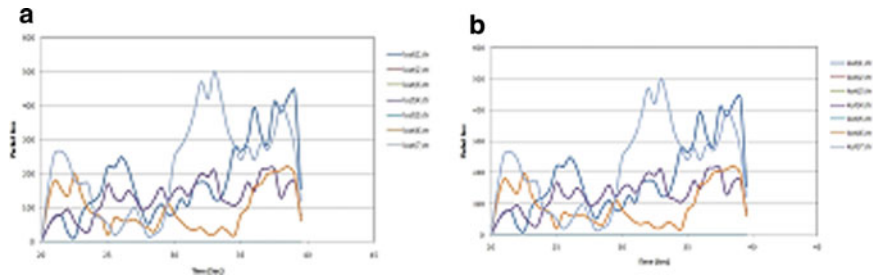


Fig. 10 **a** Packet loss case 1, **b** packet loss case 2

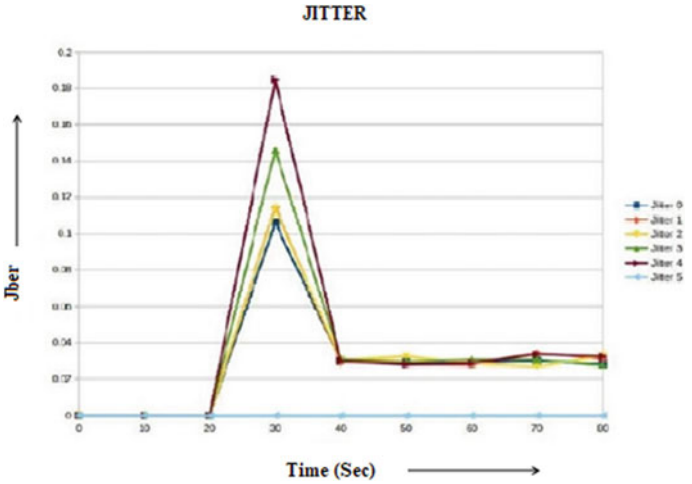


Fig. 11 Jitter for LTE network topology

Table 2 VoIP quality parameter measures for LTE network topology case 1 and case 2

Network parameter	Good	Acceptable	Poor
End-to-end delay (ms)	30 ms	–	–
Jitter (ms)	Less than 2 ms	–	–
Packet loss	–	–	2%
Throughput (Mbps)	–	65 Mbps for one-to-one call 190 Mbps for group chat	–

6 Conclusion

This article presents a comparative study of VoIP QoS over LTE network by latency, jitter, end-to-end lag and packet failure measurements.

In two case studies, two different scenarios are used to compare the performance of VoIP's QoS with LTE. A single user scenario is first used, accompanied by teams to monitor the change in performance. The end-to-end lag and jitter collapse into good space, with the exception for a few bursts from the start of each scenario. QoS guidelines are provided. Nevertheless, the packet failure is of poor quality.

Emerging technology for voice over data networks gives companies the opportunity to reduce costs and allow new applications that maintain good voice performance.

References

1. 3rd generation partnership project (Release 8) (2009)
2. Anouari T, Haqiq A (2012) Performance analysis of VoIP traffic in WiMAX using various service classes. Int J Comput Appl 52(20):29–34
3. Alsahlany AM (2014) Performance analysis of Voip traffic over integrating wireless LAN and wan using different CODECS. Int J Wireless Mobile Netw IJWMN 6(3)
4. Miraz MH, Molvi SA, Ali M, Ganie MA, Hussein AH (2014) Analysis of QoS of VoIP traffic through WiFi-UMTS networks. In: Proceedings of the world congress on engineering, vol I. WCE. London, UK, 2–4 July 2014
5. Miraz MH et al (2017) Simulation and analysis of quality of service (QoS) parameters of voice over IP (VoIP) traffic through heterogeneous networks. Int J Adv Comput Sci Appl 8(7)
6. Lonkar S, Reddy KTV (2018) Performance analysis of voice over internet protocol (VoIP) over LTE networks. In: International conference on recent innovations in electrical, electronics and communication engineering—(ICRIEECE). IEEE, July 2018, ISBN CFP18P98-PRT/978-1-5386-5994-6
7. Adu OI, Oshin BO, Alatishe AA (2013) VoIP on 3GPP LTE network: a survey. J Inf Eng Appl 3(11). www.iiste.org. ISSN 2224-5782 (print), ISSN 2225-0506 (online)

Detection and Prevention of Vampire Attack for MANET

Vanita Verma and Vijay Kumar Jha

Abstract MANET is mobile ad hoc network where mobile means moving, ad hoc means impromptu. Thus, MANET is a kind of transient networks in which nodes are able to move without any fixed infrastructure or centralized administration. A MANET can be used in special situation where the structure of installation is very complex like disaster area, battle field, etc. The battery power of nodes decides the life span of wireless ad hoc network. The recharge or battery change option seems to be infeasible in most of the applications. Therefore, data packet needs to be sent from one node to the other using minimal energy utilization. But, malicious nodes in case of vampire attacks spend more energy than that of normal and thus leading to power drooping of node and breakdown of the network. Vampire attack is the resource depletion attack and thus resulting in early drainage of the battery. The main concept is about honest and dishonest nodes. If there is a dishonest node present in the middle of honest nodes, then it will consume more energy than the normal in order to transmit the data packets to the neighboring node. In this paper, we have presented a new detection and prevention algorithm for vampire attack for MANET based on the concept of hash value and timestamp.

Keywords MANET · Vampire attack · IDS · Blockchain · Network security · Hash value · Timestamp

1 Introduction

MANETs are very supportive networks for the areas where structure of installation is very complex and battery recharge or replace option is hardly available. In these situations data should be transferred from one node to other node by utilizing minimal

V. Verma (✉) · V. K. Jha
Department of CSE, Birla Insitute of Technology Mesra, Ranchi, Jharkhand 835215, India
e-mail: vanitaaa.12@gmail.com

V. K. Jha
e-mail: vkjha@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_9

energy. But in case of vampire attack in the network, malicious node will acquire more energy than normal and thus leading to power drooping. The structure of MANET is highly volatile and thus vulnerable to attack mainly network layer or any of the layers. Intrusion detection and prevention is very crucial for achieving secure MANETs. Intrusion detection system (IDS) is usually made up of three components: 1. data collection, 2. detection, and 3. response. Detection of intrusion in MANET is more demanding and composite when compared to fixed networks because of the below-listed features:

1. MANETs does not have concentration points where audit data and monitoring can be conducted.
2. The routing protocols of MANET require cooperation from the nodes to act as routers which creates chances for attackers to attack.
3. 3. Process of IDS becomes complicated because of the mobility of the nodes, and hence, the topology of the network is unpredictable and dynamic.
4. Due to the limited computational ability of the nodes, IDS in MANETs become more complex [1].

2 Types of Vampire Attack [2]

2.1 Carousel Attack

In this attack, a data packet is sent by the attacker with route having sequence of loops and one node appears multiple times which increases the length of the route beyond the nodes count in the network that is generally limited in the source route by the number of entries allowed (Fig. 1).

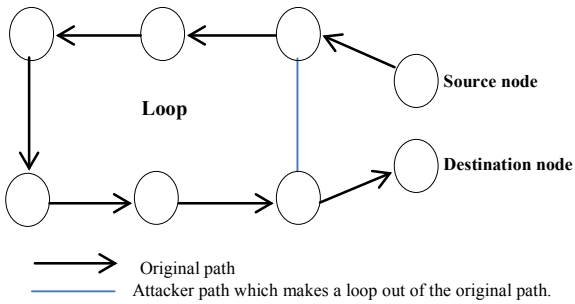


Fig. 1 Pictorial representation of Carousel attack

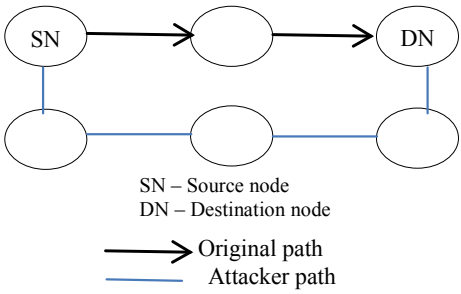


Fig. 2 Pictorial representation of Stretch attack

2.2 Stretch Attack

In stretch attack, attacker chooses the longest path for transmitting data from source to destination node rather than the optimal predefined source path and makes data packets to travel across unnecessarily many other nodes in the network (Fig. 2).

2.3 Blockchain

Blockchain is a ledger of digital transactions which is decentralized and not under the control of any single authority. It is structured and quite difficult to change the rules and content without the consensus among the people who are using it. Newer blocks are linked to the older ones forming a chain-like structure & that is why the term blockchain [3]. It can also be contemplated as a database which is used to store data in a network that is decentralized. However, blockchain is a technology that is not just getting used in application related to finance. Moreover, a transaction can also be designed to match the application according to our need.

2.4 Blockchain Structure [4]

See Fig. 3.

2.5 Related Work

See Table 1.

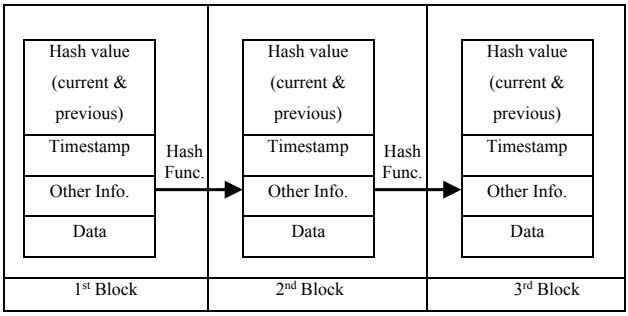


Fig. 3 Basic structure of Blockchain

3 Proposed Algorithm for Detection and Prevention of Vampire Attack for MANET

The proposed approach describes the method to detect and prevent vampire attack in MANET. In this approach, our main work is to identify the malicious node by using hash and timestamp concept of blockchain during draining the battery life of the other nodes that are genuine and then removing the malicious node for the sake of improving our network and saving other nodes battery life.

3.1 Assumption

Source node specifies the hash value of destination node.

3.2 Algorithm

- User authentication is done for each node by the server.
- Every node is linked with the hash function to the next succeeding node.
- Source node contains the current hash value, and the previous hash value will be zero since it is the first node.
- Second node contains the current hash value and previous hash value, i.e., the current hash value of the source node and so on.
- For each node, the validation is done with the current hash value of current node against the previous hash value of succeeding node.
- If both the hash values are same, it means current node is honest node. If different, it means current node is malicious node.
- As soon as malicious node is detected, alert message will be sent to the administrator and is removed from the network.

Table 1 Related works

S. No.	Authors, year, and paper	Method used	Remarks
1	Eugene Y. Vasserman et al. [5]—Vampire attacks: draining life from wireless ad hoc sensor networks	Sensor network routing protocol PLGPa	PLGPa protects from vampire attack by making sure that data packets are consistently in progress toward the destination
2	Mrs. R. Abirami et al. [6]—Depletion of vampire attacks in medium access control level using interior gateway routing protocol	Interior gateway routing protocol (IGRP)	The author has proposed protection in opposition to some attacks that are forwarding phase attacks and has also explained about IGRP
3	Amee A. Patel et al. (2015)—A novel proposal for defending against vampire attack in WSN	Threshold energy of node	1. The draining of energy of the nodes in WSN can be protected from vampire attack, and the network lifetime can be increased efficiently 2. Future work according to the author is implementation of the proposed method in NS2 for AODV routing protocol
4	Ankita Shrivastava et al. [7]—Detection of vampire attack in wireless ad hoc network	Packet monitoring technique	The author has mentioned the objective of the proposed work which is finding a solution that is optimum for detecting vampire attack in wireless ad hoc networks
5	Manish Soni et al. [8]—Detection and removal of vampire attack in wireless sensor network	Variance difference of node's energy at distict time	As per the proposed solution, an attacker can be detected with the help of broadcast rate of data packets and the parameters of energy among the nodes of the network. In future, they will come up with the simulation work in NS2
6	Harsha Nanwani et al. (2015)—Detection of vampire attack and prevention for MANET	Waiting time of a node and count of no. of requests sent	The author has proposed the algorithm to detect and then remove the node that is malicious from the network along with simulation work

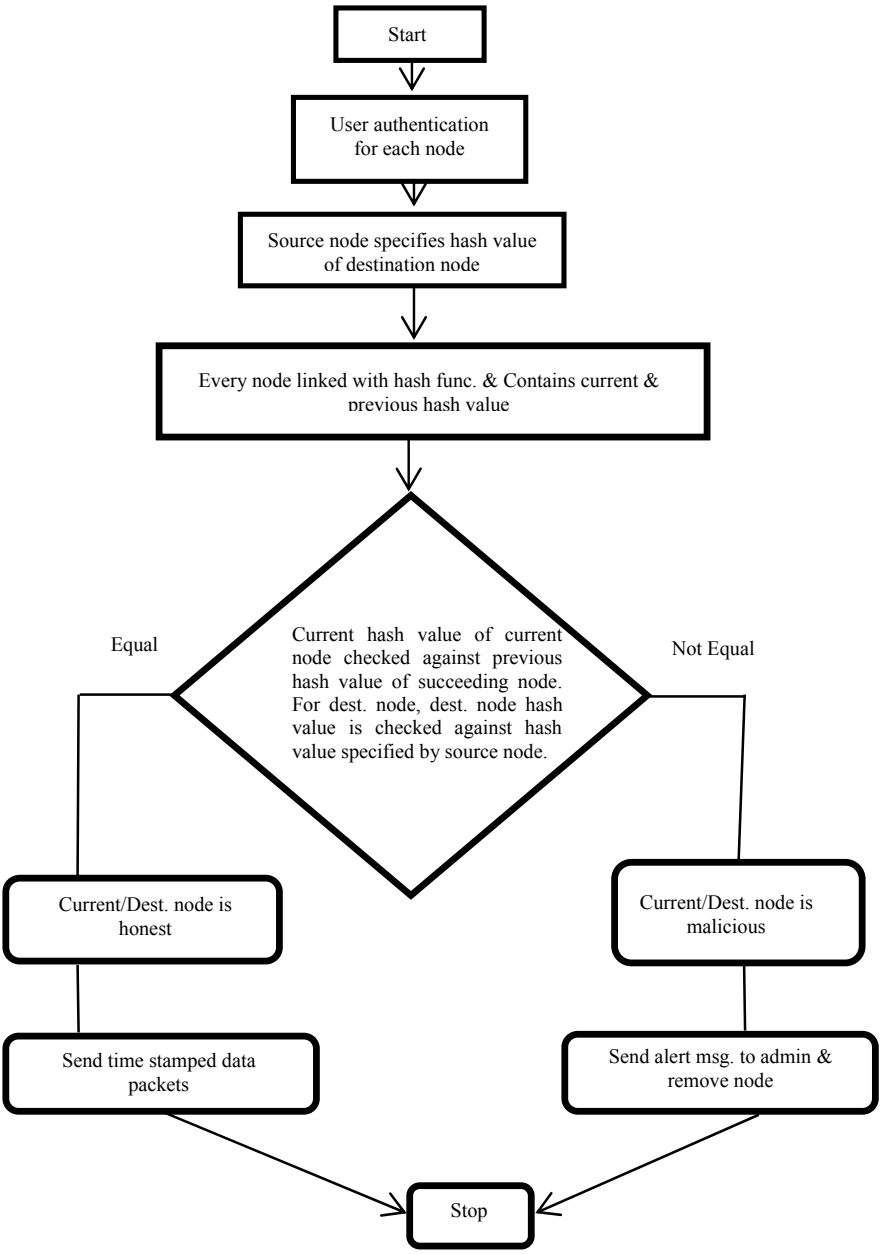
(continued)

Table 1 (continued)

S. No.	Authors, year, and paper	Method used	Remarks
7	Prof. Gayatri A. Jagnade et al. [9]—Secure VANET from vampire attack using LEACH protocol	LEACH protocol	Through this methodology, the time requirement of each node will be less, and therefore, communication can be done securely. When communication can be done in that time if heard node's energy is less on the LEACH protocol can select alternate node which has nearer energy with respect to main header node and for that specific communication this node is then considerd as temporary header
8	Mayank Kumar sharma et al. [10]—Detection and prevention of vampire attack in wireless sensor networks	Draining rate prediction method	In the first phase, deployment is done in AODV by configuration of vampire AODV a new protocol, for analyzing how it will affect the sensor nodes availability. After this, the mitigation approach is unified in AODV by observing the difference in battery consumption for victim nodes in normal scenario and in case of vampire attack
9	Anithaashri T. P. et al. [11]—Security enhancement for wireless network cmmunication	Algorithm -Elliptic curve digital signature, DCFM in addition to GTP	As per the author the proposed system is highly secure for wireless network. For vampire attack and so, it includes trust model and status management schemes

- At destination node, the hash value of destination node is compared with the specified hash value by the source node, and if both are equal, it means that it is an honest node else malicious.
- In case if it is malicious node, alert message will be sent to the administrator and is removed from the network.
- Data packets which move from one node to other will also be timestamped to enhance the security of the network.

4 Architecture for Detecting Malicious Node



5 Pseudo Code for Detection and Prevention of Malicious Node in Vampire Attack

5.1 Assumptions

n = Set of nodes in the network.
N = Destination node
h = hash of each node present in the network
H = Previous hash of each node in the network
x = Value of hash
Hash value of destination node is known through the source node.
1. Initialize h, H and x
2. For n = 1
  h[n] = x[n]
  H[n] = 0 (Since for the 1st node there is no previous node)
3. Loop starts from n > 1 and ends at n-1.
4. For n = 2
  h[n] = x[n]
  If {H[n] = h[n-1]}
    {Node 1 is honest node}
    {Timestamped data packet is sent}
  Else
    {Node 1 is malicious node}
    {Send alert message to the administrator}
    {Remove malicious node from the network}
  End-If
5. For n = 3
  h[n] = x[n]
  If {H[n] = h[n-1]}
    {Node 2 is honest node}
    {Timestamped data packet is sent}
  Else
    {Node 2 is malicious node}
    {Send alert message to the administrator}
    {Remove malicious node from the network}
  End-If
6. So on for rest of the nodes except the destination node.
7. For destination node i.e., n = N, If h[N] != Known hash value of destination node
  Then destination node itself is malicious Else honest
    {Send alert message to the administrator}
    {Remove malicious node from the network}

6 Conclusion

In this digital era, networking is playing a very crucial role in connecting people, and if the connection is wireless, then it is more efficient. To establish communication over the network, several nodes interact with each other by sharing information, but if the security of the network is breached by the entry of malicious node into the network, then it is very necessary to detect these nodes at a very early stage and prevent them or block them. The proposed algorithm is an effective approach toward the same.

7 Future Work

In future, the proposed approach will be implemented by NS2 simulator tool and upgraded in various ways.

- Graphs will be presented for proving its efficiency.
- Comparison with existing proposed algorithm results.

References

1. Sreenivasa Chakravarthi S, Veluru S (2014) A review on intrusion detection techniques and intrusion detection systems in MANETS. In: 2014 sixth international conference on computational intelligence and communication networks. Bhopal, India, 14–16 Nov 2014
2. Patel AA, Soni SJ (2015) A novel proposal for defending against vampire attack in WSN. In: 2015 Fifth international conference on communication systems and network technologies. Gwalior, India, 4–6 Apr 2015
3. Kumar S et al (2019) A survey paper on blockchain technology, challenges and opportunities. Int J Comput Trends Technol IJCTT 67(4)
4. Joshi AP, Han M, Wang Y (2018) A survey on security and privacy issues of blockchain technology. Math Found Comput 1(2): 121–147
5. Vasserman EY, Hopper N (2011) Vampire attacks: draining life from wireless ad-hoc sensor networks. IEEE Trans Netw Secur Tech Details 12(2):318–332
6. Abirami et al (2014) Depletion of vampire attacks in medium access control level using interior gateway routing protocol. In: International Conference on Information Communication and Embedded Systems (ICICES2014). IEEE
7. Shrivastava A, Verma R (2015) Detection of vampire attack in wireless ad-hoc network. Int J Softw Hardw Res Eng 3(1):43–48
8. Manish Soni et al (2015) Detection and removal of vampire attack in wireless sensor network. Int J Comput Appl 126(7)
9. Gayatri A. Jagnade et al (2016) Secure VANET from vampire attack using LEACH protocol. In: 2016 International conference on signal processing, communication, power and embedded system (SCOPES). IEEE
10. Sharma MK, Joshi BK (2017) Detection and prevention of vampire attack in wireless sensor networks. In: IEEE, International Conference on Information, Communication, Instrumentation and Control (ICICIC- 2017)

11. Anithaashri TP, Ravichandran G (2018) Security enhancement for wireless network communication. In: Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI) IEEE 2018
12. Cao N, Wang Y (2017) A novel approach to improve robustness of data mining models used in cyber security applications. In: 2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC), vol 2. IEEE
13. Clincy V, Shahriar H (2018) Web application firewall: Network security models and configuration. In: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). vol. 1
14. Gayathri E, Geetanjali N (2017) A resilient strategy to overcome stretch attacks by using kmidpoint—medoid gateway clustering algorithm. In: International Conference on Current Trends in Computer, Electrical, Electronics and Communication (ICCTCEEC-2017)
15. Aras ST, Kulkarni V (2017) Blockchain and its applications—a detailed survey. Int J Comput Appl 180(3):29–35
16. Alhadhrami Z, Alghfeli S, Alghfeli M, Abedlla JA, Shuaib K (2017) Introducing blockchains for healthcare. In: 2017 international conference on electrical and computing technologies and applications (ICECTA), IEEE, Ras Al Khaimah, United Arab Emirates, 21–23 Nov 2017
17. Tasatanattakool P, Techapanupreeda C (2018) Blockchain: challenges and applications. In: 2018 International Conference on Information Networking (ICOIN). IEEE
18. Abbas QE, Sung-Bong J (2019) A survey of blockchain and its applications. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE
19. Trabelsi Z, Molvizadah V (2016) Edu-firewall device: An advanced firewall hardware device for information security education. In: 2016 13th IEEE Annual Consumer Communications and Networking Conference (CCNC). IEEE
20. Sarkar BK (2017) Big data for secure healthcare system: a conceptual design. Complex Intell Syst 3(2):133–151
21. Saha S (2018) Web application firewall-dot defender
22. Goliwale P et al (2018) Intrusion detection system using data mining. Manage 5(3)

Mathematical Modelling of Magnesium Block-Driven NMDA Receptor Response in CA1 Pyramidal Neuron for Alzheimer's Disease

Vijay Dave, Arpit D. Shrimankar, Devanshi Gokani, and Abha Dashora

Abstract N-methyl-D-aspartate receptors (NMDA) located on the postsynaptic terminal of CA1 pyramidal neuron plays an essential role in synaptic plasticity, and the calcium(Ca) current flowing through it drives several intracellular calcium mechanisms. We have developed a single-compartment model involving least mechanisms to study the NMDA receptor response to different input condition and to obtain the voltage-dependent relationship of magnesium(Mg) block in the NMDA receptor. Using this interdependence, we have replicated a pathological condition of Alzheimer's diseases (AD) to obtain total NMDA receptor current in AD. The results obtained at different Mg concentration suggested that the blocking rate increases linearly with the Mg concentration. In AD, there is dysfunction of Mg block which leads to the early opening of the NMDA receptor. So, using a drug which controls the blocking mechanism of Mg leads to control of NMDA receptor current. This results in the control of intracellular Ca current.

Keywords NMDA receptor · CA1 pyramidal neuron · Calcium · Magnesium · Alzheimer disease

V. Dave (✉) · A. D. Shrimankar · D. Gokani · A. Dashora
Department of Biomedical Engineering, Government Engineering College, Sector 28, Gandhinagar 382028, Gujarat, India
e-mail: vijaydave12@gmail.com

A. D. Shrimankar
e-mail: arpitshrimankar@gmail.com

D. Gokani
e-mail: devanshigokani18@gmail.com

A. Dashora
e-mail: pari1998dashora@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_10

1 Introduction

Neurons found in the region of the cerebral cortex, the hippocampus, and the amygdala area of the brain are recognized as pyramidal neurons.The CA1 pyramidal cells are primary output neurons of the hippocampus [1] critical to object differentiation in long-term memory. Their relationship with Schaffer collateral in the CA3 region of the hippocampus leads to long-term potentiation(LTP) which is crucial for information storage. These neurons are majorly affected by neurological disorders like Alzheimer's diseases(AD) [2]. At the cellular level, AD is identified by a gradual loss of cortical neurons, particularly pyramidal cells, that regulate higher cognitive functions. The symptom of neurodegenerative disease declines over time from early stages of forgetfulness to the gradual change in language, orientation, and behavior and severe loss of memory and some bodily function until the ultimate death. Thus, there are both structural and functional abnormalities found in AD [3]. AD neuropathology is usually identified by the accumulation of insoluble amyloid protein that originates from the amyloidogenic processing of a much larger metalloprotein—amyloid precursor protein (APP) that leads to the formation of extracellular neuritic amyloid plaques containing the peptide-beta-amyloid peptide ($A\ \beta$).

The other major AD pathological hallmarks include neurofibrillary tangles (NFTs) that consist of misfolded, abnormal phosphorylated microtubule-associated tau protein. It has been theorized that the neuronal damage seen in AD is identified by deposition of abnormal proteins both within and outside of neurons [4]. N-methyl-D-aspartate receptors (NMDA) play significant physiological as well as pathological roles in the central nervous system (CNS). NMDA receptors are ionotropic glutamate receptors or ligand-gated ion channels that open when ligand interacts with the receptor to create an internal response. Initially, the NMDA receptors are blocked by physiological magnesium (Mg) concentration. This block is voltage-dependent thus allowing NMDA receptor to conduct only when the membrane is depolarized

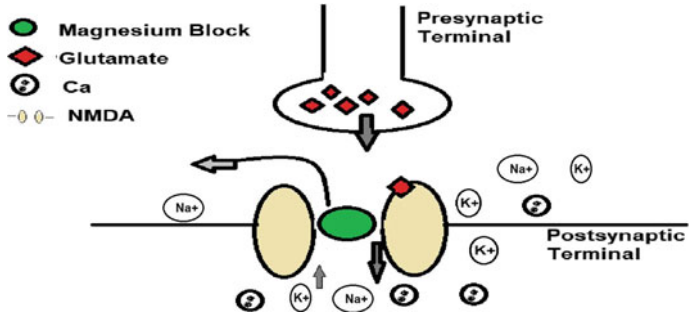


Fig. 1 Presynaptic action potential triggers glutamate release from vesicles that diffuse into the synaptic cleft and bind to NMDA receptor present on a postsynaptic terminal that is permeable to Na+ , K+ and Ca ions

as shown in Fig. 1. Since for the current to flow, it requires activation of both pre and postsynaptic terminal, the receptors behave like co-incidence detectors [5].

The activation of the receptor causes the opening of an ion channel that is selective for cations, that leads toan influx of sodium (Na+),Ca ions, and efflux of potassium (K+) ions [6]. These receptor are found on synaptic and extrasynaptic locations of postsynaptic membrane. In the postsynaptic membrane of excitatory neurons, the density of NMDA receptors is higher in dendritic spines, within the postsynaptic density (PSD), considered synaptic NMDA receptors while on further sides of PSD, on the spine neck, the dendritic shaft or soma, extra synaptic NMDA receptors are localized [3]. The Ca influx through NMDA receptors is the critical trigger for the induction of synapse-specific plasticity [7]. Modest activation of NMDA receptor induces long-term depression (LTD), whereas strong activation produces LTP [8–10]. Long-term potentiation is a continuous strengthening of connections between synapses of existing neurons based on recent activities. It is the mechanism underlying memory and learning [5, 11].The stimulus driven by presynaptic terminal results in the release of neurotransmitters such as glutamate that activates alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor(AMPAr) that causes depolarization of postsynaptic terminal. Sufficient depolarization and binding of glutamate can relieve the Mg block and allow Ca influx through synaptic NMDA receptor. This promotes the recruitment of AMPAr and growth of dendritic spines thus increases synaptic strength and cell survival [5, 12]. Activation of synaptic NMDA receptor and increase in calcium influx results in LTP. However, inpathological conditions like AD, due to accumulation of A β the Mg block of a channel is partially removed and abnormally increases the NMDA receptor activity [13, 14]. This leads to increased intracellular Ca concentration that results in cell excitotoxicity or celldeath [15–17].

2 Methodology

Through a literature review, we have adopted a simple kinetic model of synaptic transmission [18] in which a single compartment is used to represent a dendritic section of length 50 μm and diameter 1 μm. The passive parameter was set as membranne resistance (Rm) of 28 kΩ cm^2 and membrane capacitance(Cm) of 1.5 μF/cm^2 [19]. In addition to that, NMDA receptor was introduced to the section to obtain the total NMDA receptor current (I_{NMDA}) and NMDA receptor-mediated Ca currents (I_{Ca}) that are important for various intracellular mechanisms. In the present study, we have modelled synaptic conductance (g_{NMDA}) using single exponential function, alpha function, and double exponential function. We have blocked other postsynaptic mechanism, i.e., AMPAr and voltage-gated calcium channels (VDCC) to study the NMDA receptor activity in response to different input condition as well as toward varying Mg concentration that gets affected in AD. NMDA receptor is a transmitter-gated channel, i.e., it opens when transmitter molecule binds to it. It is described by the following equation:

$$C + T \underset{\beta}{\overset{\alpha}{\longleftrightarrow}} O \tag{1}$$

where C and O refer to the closed state and open state of receptor, respectively. T is the neurotransmitter that binds to receptors and α and β are the voltage-independent forward and backward rate constants, respectively ($\alpha = 0.072$/ms mM and $\beta = 0.0066$/ms). If r is the fraction of channels in open state, then it is described by the following equation:

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \propto [T](1 - r) - \beta r \tag{2}$$

The current through the receptors takes into account voltage-dependent Mg block $[B(V)]$:

$$B(V) = 1/(1 + \exp(-0.062V)([Mg2+]o/3.57) \tag{3}$$

Here, V is the membrane potential in mV and $[\mathrm{Mg}]_\mathrm{o}$ is extracellular Mg concentration (1–2 mM in physiological conditions).

g_{NMDA} is modelled using different functions [20]. Three commonly used functions are as follows:

(A) Single exponential decay:

$$g_{\mathrm{NMDA}}(t) = g_{\max} \exp\left(-\frac{t - t_s}{\tau}\right) \tag{4}$$

where onset (t_s) is the time after which conductance of postsynaptic terminal changes (100 ms).

(B) Alpha function:

$$g_{\mathrm{NMDA}}(t) = g_{\max} \frac{t - t_s}{\tau} \exp\left(-\frac{t - t_s}{\tau}\right) \tag{5}$$

(C) Dual exponential function:

$$g_{\mathrm{NMDA}}(t) = g_{\max} \frac{\tau_\mathrm{f}\tau_\mathrm{s}}{\tau_\mathrm{f} - \tau_\mathrm{s}} \left(\exp\left(-\frac{t - t_\mathrm{s}}{\tau_\mathrm{f}}\right) - \exp\left(-\frac{t - t_\mathrm{s}}{\tau_\mathrm{s}}\right)\right) \tag{6}$$

Here, $g_{\max}$ is maximum NMDA receptor conductance (range: 0.01–0.6 nS), and τ_f (fast) and τ_s (slow) are the time constants [21]. Among all the above Eqs. (4), (5), and (6), the dual exponential function which proves to be a more realistic representation of synaptic conductance, fits the best with experimental results. The voltage-dependent NMDA current I_{NMDA} that takes into account the Mg block is given

By:

$$I_{\mathrm{NMDA}} = (g_{\mathrm{NMDA}}) * B(V) * r * (V - E_{\mathrm{NMDA}}) \tag{7}$$

where E_{NMDA} is the NMDA receptor reversal potential (0 mv).

The Ca current flowing through NMDA receptor I_{Ca} is described by the following equation [19]:

$$I_{\mathrm{Ca}} = Y(V - E_{\mathrm{Ca}})\left(\exp\left|\left(-\frac{t - t_{\mathrm{s}}}{\tau_{\mathrm{f}}}\right) - \exp\left|\left(t - \frac{t - t_{\mathrm{s}}}{\tau_{\mathrm{s}}}\right)\right)\right.\right. \tag{8}$$

$$Y = P_{\mathrm{NMDA}} P_{\mathrm{Ca}} B| \tag{9}$$

where $\mathrm{P_{NMDA}}$ is the maximum permeability of NMDA receptor, P_{Ca} is the Ca permeability through the NMDA receptor, and E_{Ca} is the reversal potential of Ca channels (130 mv).

3 Results

Through the literature review, we optimized the parameters used in our simulation results with a wide range of experimental and simulation results. The simulation work is done in NEURON software [Michael Hines, John W. Moore and Ted Carnevale]. Results are shown below in Figs. 2, 3, 4, 5, 6, and 7. Table 1 shows the comparison of simulation parameters of the synaptic model for three different input functions, i.e., alpha function, single exponential function, and double exponential function are used obtain the total NMDA receptor current in presence of 1 mM Mg (Fig. 2).

In the experimental condition, the input is provided with a square pulse (1 ms) pulse of 1 mM of glutamate neurotransmitter. The range of g_{max} nS) is between 0.01 and 0.6, the amplitude of I_{NMDA} is 80 pA, and duration is 600 ms [18]. From the above three functions, it has found that dual exponential fits the best with that

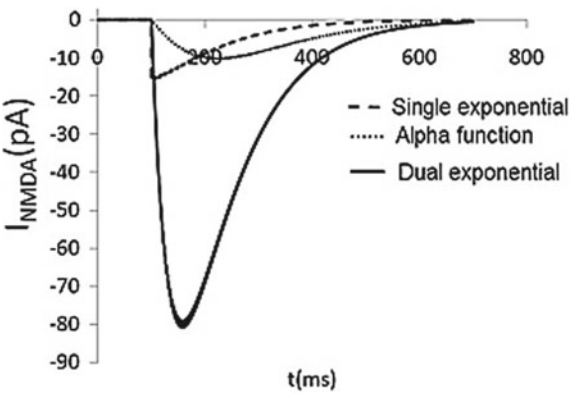


Fig. 2 Total NMDA receptor current v/s time for three different input functions in the presence of 1 mM Mg concentration (refer to Table 1)

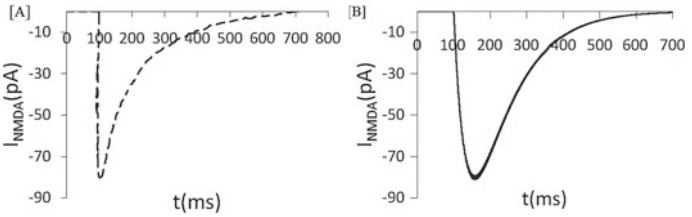


Fig. 3 Total NMDA current (influx of Na+ , Ca and outflux of K+ ions), through postsynaptic terminal in 1 mM Mg concentration. [A] Experimental result [18], [B] simulation result [using Eq. (7)], refer Table 1

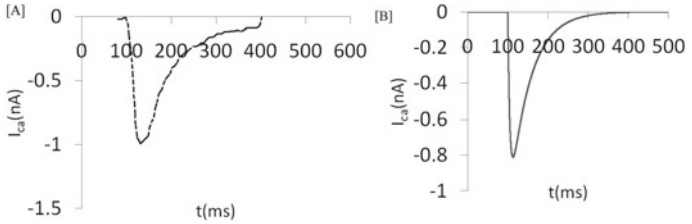


Fig. 4 Ca current (I_{Ca}) through NMDA. [A] Experimental [22], [B] Simulation [using Eq.(8)]

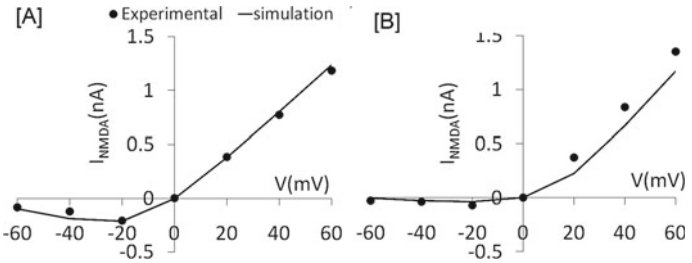


Fig. 5 I/V relationship of NMDA currents for [A] 1 mM and [B] 10 mM Mg concentration [23, 24]

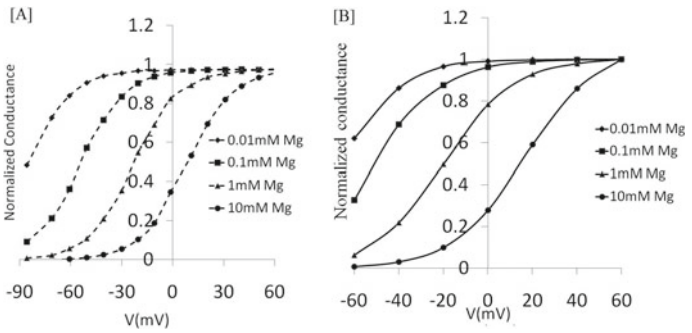


Fig. 6 Normalized conductance-voltage relationship of NMDA receptor responses at four Mg concentrations. [Refer Eq. (8)]. [A] Experimental result [23], [B] simulation result

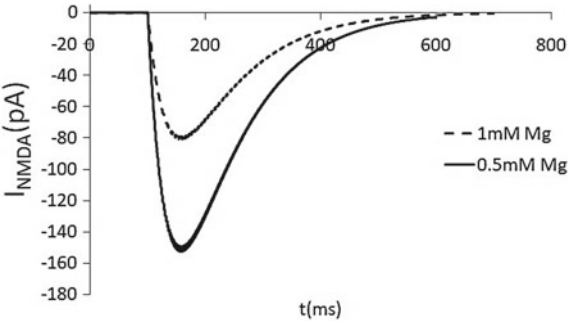


Fig. 7 NMDA receptor current in normal and AD pathological condition

Table. 1 Comparison of amplitude and duration of three different functions to obtain NMDAr current

Function	Amplitude (pA)	Duration (ms)
Single exponential decay ($\tau = 90$ ms)	15	615
Alpha function ($\tau = 90$ ms)	10	867
Double exponential function (τ_f) = 24 ms and (τ_s) = 90 ms)	80	656

of experimental results, and it is shown in Fig. 3. Here, synaptic conductance has been modelled as the double exponential function(as it resembles NMDA receptor synaptic conductance) with an onset of 100 ms.

These ionotropic channels are highly permeable to Ca ions that drives other intracellular mechanisms, and we have plotted these current in Mg-free solution which is shown in Fig. 4 with an onset of 100 ms. Here, synaptic conductance has been modelled as a double exponential function (as it resembles NMDA receptor synaptic conductance). The parameters are given in Table 2, and it compares the simulation parameters of our model with the experimental parameters done previously [22]. The experiment was performed by providing current injection (IClamp) of 0.1 nA amplitude and duration of 2 ms.

We further studied the effects of Mg on NMDA responses, and this was to obtain the voltage-dependent relationship of Mg block in NMDA receptor as shown in Fig. 5. The simulation was done by considering two different magnesium concentrationNMDAr mediated postsynaptic current shows a negative slope conductance for potentials between −40 and −20 mV for 1 mM Mg and between −20 and 0 mV for 10 mM Mg. This is because at high Mg concentrations blocking rate

Table. 2 Comparison of experimental and simulation parameters of the synaptic model to obtain NMDAr mediated calcium current

Parameter	Experimental values [22]	Simulation values (our model)
P_{NMDA}	1e−6	1e−10
I_{ca} (nA)	−1	−0.8
Duration (ms)	300	270

is high, thus requires more positive potential to unblock the receptors. Normalized conductance-voltage relationship of NMDA receptor responses is plotted at four different Mg concentrations as shown in Fig. 6. The curves become less steep with decreasing concentration of Mg. Since blocking rates increase linearly with increased Mg concentration, thus, at higher Mg concentrations, more positive potential is required to relieve the block. Both of the two simulations were recorded at a membrane potential from −60 to +60 mv in 20mv increments at 1000 ms duration.

Along with this, results are obtained for NMDA current in healthy and under AD pathological condition as shown in Fig. 7. In normal conditions, assuming extracellular Mg concentration to be 1 mM. While under AD pathological conditions, the Mg block is partially relieved, thus considering 0.5 mM Mg. The amount of current flowing through receptors is increased by twofold [13, 25].

4 Conclusion

To summarize, we have developed a simple compartment model with NMDA receptor on a dendritic section of the postsynaptic terminal of CA1 pyramidal neuron. The reason for considering the least mechanism and structure is to replicate the same biological responses which are complex in nature with least components. The postsynaptic terminal is activated using three different inputs conditions, i.e., single exponential function, alpha function, and double exponential function, to obtain the NMDA receptor current which plays an important role in synaptic plasticity. These ionotropic channels are highly permeable to Ca ions that drives other intracellular mechanisms, and we have plotted these current in Mg-free solution. We also studied the effects of Mg on NMDA receptor responses, and this was to obtain the voltage-dependent relationship of Mg block in the NMDA receptor. The current–voltage relationship and normalized conductance voltage curves suggested that the blocking rate increases linearly with the Mg concentration, and thus at higher Mg concentrations, more positive potential is required to relieve the block. In addition to this, we plotted the NMDA receptor current in healthy and under AD pathological condition. The receptor channels are initially blocked in normal conditions.

But in the case of AD, the $A\ \beta$ chronically depolarize neurons through its action on the ionotropic glutamate receptors. It partially relieves the block causing overactivation of receptors. These results in excessive current (i.e., calcium overload) to flow inside the cell and give rise to cell excitotoxicity. We have simulated this situation by varying Mg concentration that becomes half of that assumed in healthy condition (1 mM Mg). Also, we have obtained the NMDA receptor current in AD. Since the glutamatergic neurotransmission plays a crucial role in the process of memory and learning, dysfunctioning of these glutamatergic receptors can lead to neuronal injury and death and cognitive deficits associated with dementia. So in order to perform proper functioning of NMDA receptor, Mg block is the key component that regulates NMDA current. To avoid the condition of AD, a drug can be developed that can regulate the functioning of Mg block in the NMDA receptor.

References

1. Bloss EB, Cembrowski MS, Karsh B, Colonell J, Fetter RD, Spruston N (2016) Structured dendritic inhibition supports branch-selective integration in CA1 pyramidal cells. Neuron 89:1016–1030
2. Spruston N (2008) Pyramidal neurons: dendritic structure and synaptic integration. Nat Rev Neurosci 9:206
3. Wang R, Reddy PH (2017) Role of glutamate and NMDA receptors in Alzheimer's disease . J Alzheimer's Dis 57:1041–1048
4. Korolev IO (2014) Alzheimer's disease: a clinical and basic science review. Medical Stud Res J 4:24–33
5. Olivares D, Deshpande VK, Shi Y, Lahiri KD, Greig HN, Rogers TJ, Huang X (2012) N-methyl D-aspartate (NMDA) receptor antagonists and memantine treatment for Alzheimer's disease, vascular dementia and Parkinson's disease. Curr Alzheimer Res 9:746–758
6. Zito K, Scheuss V (2009) NMDA receptor function and physiological modulation
7. Rackham O, Tsaneva-Atanasova K, Ganesh A, Mellor J (2010) A Ca2+-based computational model for NMDA receptor-dependent synaptic plasticity at individual post-synaptic spines in the hippocampus. Front Synap Neurosci 2:31
8. Shouval HZ, Bear MF, Cooper LN (2002) A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. Proc Nat Acad Sci 99:10831–10836
9. Malinow R, Otmakhov N, Blum KI, Lisman J (1994) Visualizing hippocampal synaptic function by optical detection of Ca^{2+} entry through the N-methyl-D-aspartate channel. Proc Nat Acad Sci 91:8170–8174
10. Urakubo H, Honda M, Froemke RC, Kuroda S (2008) Requirement of an allosteric kinetics of NMDA receptors for spike timing-dependent plasticity. J Neurosci 28:3310–3323
11. Lynch MA (2004) Long-term potentiation and memory 84: 87–136
12. Ozawa S (1996) Permeation of calcium through glutamate receptor channels. Elsevier 8
13. Wenk GL, Parsons CG, Danysz W (2006) Potential role of N-methyl-D-aspartate receptors as executors of neurodegeneration resulting from diverse insults: focus on memantine. Behav Pharmacol 17:411–424
14. Zeevalk GD, Nicklas WJ (1992) Evidence that the loss of the voltage-dependent Mg2+ block at the N-methyl-d-aspartate receptor underlies receptor activation during inhibition of neuronal metabolism. J Neurochem 59:1211–1220
15. Latulippe J, Lotito D, Murby D (2018) A mathematical model for the effects of amyloid beta on intracellular calcium. PLoS ONE 13:e0202503
16. Berridge MJ (2011) Calcium signalling and Alzheimer's disease. Neurochem Res 36:1149–1156
17. Zipfel GJ, Babcock DJ, Lee J-M, Choi DW (2000) Neuronal apoptosis after CNS injury: the roles of glutamate and calcium. J Neurotrauma 17:857–869
18. Destexhe A, Mainen ZF, Sejnowski TJ (1998) Kinetic models of synaptic transmission. Methods Neuronal Model 2:1–25
19. Narayanan R, Johns D (2010) The h current is a candidate mechanism for regulating the sliding modification threshold in a BCM-like synaptic learning rule. J Neurophysiol 104(2)
20. Sterrett D, Graham B, Gillies A, Willshaw D (2011) Principles of computationalmodelling inneuroscience. CambridgeUniversity Press, Cambridge
21. Hestrin S, Sah P, Nicoll RA (1990) Mechanisms are generatingthetime courseofdual-component excitatorysynapticcurrents. Neuron 5
22. Bollmann JGGBJH, Sakmann B (1998) Postsynaptic Ca^{2+} influx mediated by three different pathways during synaptic transmission at a calyx-type synapse. J Neurosci 18(24):10409–10419
23. Jahr CE, Stevens CF (1990) A quantitative description of NMDA receptor-channel kinetic. J Neurosci 10:1830–1837

24. Ambert N, Great R, Haeberlé O, Bischoff S, Berger TW, Bouteiller JM, Baudry M (2010) Computational studies of NMDA receptors: differential effects of neuronal activity on the efficacy of competitive and non-competitive antagonists. NIH Public Access Open Access Bioinform 2:113
25. Molnar Z, Sos K, Lengyel I, Penke B, Szegedi V, Budai D (2004) Enhancement of NMDA responses by beta-amyloid peptides in the hippocampus in vivo. NeuroReport 15:1649–1652

Hybrid Model for Stress Detection in Social Media by Using Dynamic Factor Graph Model and Convolutional Neural Networks

N. Prabakaran, L. Ramanathan, and R. Kannadasan

Abstract In present days, mental pressure is getting to be real medical problems. It distinguishes to recognize pressure opportune in proactive way. By extending the predominance digital life, person using to sharing their step-by-step activities and speaking to companions by means of electronic systems administration locales in various ways, so it makes viable to utilize online electronic life data for stress acknowledgment. Here, we discover clients' stress state which is immovably related to that of companions in their Web/public activity, and also, we use different sizes of dataset in various states which is identified with public activity to productively look at the relationship of clients' stress states and connecting within public activity. First, we identify the clients' stress states like visual pictures, writings and group properties with alternate points of view and further move on novel model to show. Dynamic factor graph model (DFGM) is combined with convolutional neural network (CNN) to utilize tweets and group cooperation information for stress disclosure and furthermore which can identify some dependencies locally and scale of invariance in discourse of speech recognitions and also in image recognition (i.e., smileys). The most part of FGM focuses on mood cast technique which is dependent on unique persistent in demonstrating and predicts the user's emotions in social life. By further examining the social life of the user's data, we can likewise find many scenarios and the amount of social models in form of scattered connection of stress related to users and also with non-stressed users.

Keywords Convolutional neural network (CNN) · Dynamic factor graph model (DFGM) · Mood state detection · Social media

N. Prabakaran (✉) · L. Ramanathan · R. Kannadasan
SCOPE, VIT University, Vellore, India
e-mail: dhoni.praba@gmail.com

L. Ramanathan
e-mail: lramanathan@vit.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_11

1 Introduction

Feeling can be characterized as an unpredictable arrangement psychophysical invigorate that emerges suddenly multiple times quicker than reasonable idea. Such inclination invokes by positive or negative response by physical explanations. Quite a while back, scientists trusted that our choice procedure was primarily made by discerning musings. Reality, in any case, ended up being entirely different. Instantly, idea drives human being to be intrigued, yet it is a feeling that moves to be distinctly. Advertisers currently understand that people are very enthusiastic in nature and they take their decisions either by choice or emotionally. It is intriguing to see how a person's feeling is impacted by different variables. Already, it was extremely hard to think about the issue because of the absence of accessibility of information. Recently the achievement of some expansive scale, online informal organizations, for example, Facebook, MySpace and Twitter, numerous implicit networks where approximately shaped and regularly to show basic alternates, for example, discussions on a broad assortment of problem running from item audits to crusades. In implicit networks, the difference in a person's passionate stages can impact in constructive methods and the feeling may additionally change substantial course selections of satisfaction and sadness.

The spread of feeling differs profoundly which affects the aggregate opinions in informal organizations. It understands the elements of human feelings in online interpersonal organizations and can give efficient data to applications, for example, spread of political perspectives and shopper purchasing behaviors. To be sure, there have been subjective investigations on the spread of joy or misery. For instance, it was discovered that an individual could spread her passionate state through connections up to three types of partition.

Features can be extracted by considering important attributes, since it can be applicable on various types of data to separate among different user stress states. Here, we use state of mind type for dynamic factor diagram show for mindset discovery of client tweets on day by day and week after week premise, and furthermore, CNN strategy is utilized for feature extraction of local dependency and furthermore for temporal dependency.

2 Related Works

Adversarial networks were effectively utilized for picture age (Good fellow 2014; Denton et al., 2015), area adaption and semi-directed learning. The key thought of adversarial networks for space adaption is to build invariant highlights by advancing the component extractor as an adversary against the area classifier (Zhan et al., 2017).

Parts-of-speech form labeling is an essential preprocessing step and can give important data for different common dialect preparing undertakings. As of late, profound learning calculations have been effectively utilized for POS labeling.

Various methodologies have been proposed and have accomplished some advancement by utilizing a character-based convolutional neural system in POS labeling issue. Bi-LSTMs with word, character or Unicode byte implanting were likewise acquainted with accomplish the POS labeling and named element acknowledgment tasks (Plank 2016; furthermore, Hovy 2016). In this work, we examine the issue from an area adaption point of view. Roused by these works, we additionally propose to utilize character-level techniques to deal with out-of-vocabulary words and bi-LSTMs to display the grouping relations.

Recently, an ever-increasing number of indirectly regulated strategies are proposed. Preparation information comprises tweets with emoji like ":)" and utilizes these emoji as loud names. Davidson (2010) utilizes 50 Twitter labels and 15 smileys as boisterous names to distinguish and group differing feeling kinds of tweets. Different techniques with uproarious names (Moore 2011) are additionally proposed. Every one of these techniques cannot deal with subjectivity characterization well. Besides, these strategies need to slither every one of the information and store them in the nearby circles. This is exceptionally wasteful when millions or even billions of tweets are utilized in light of the fact that ask for rate for creeping tweets which is restricted by Twitter server.

3 Existing Model

Fast augmentation in stress is transformed from unprecedented test to human prosperity and quality of life. Subsequently, there is essential centrality to recognize stress before it changes into extraordinary issues [1]. Stress is mainly used to identify based on direct interviews and self-poll reports. There are likewise some examination works, utilizing user tweeting substance via Web-based networking media stages to recognize clients' stress state [2]. There are no opportune and proactive techniques for stress identification. Firstly, tweets to be constrained among greatest of 100 characters in social stages like Twitter clients do not utilize in every case to expressing the stress states straightforwardly by tweets [3]. Users who have high stress pressure contains low active state on social communities, as revealed during some examination [4]. These miracles achieve the normal data sparse and equivocalness issue, which changes the execution of tweets-based weight area execution (Olszewski et al. 2016).

4 Proposed Model

Impelled by mental theories, we at first consider set of properties in stress area through tweets and client-level points independently (Fig. 1).

- Tweets level to consider features from the user's individual tweets

- A user-level characteristic from the user's week-by-week tweets.

Mainly, we consider client-level properties in two types for evaluating the major dissimilarity among stress and non-stress user based on states that depends on the user's week-by-week tweet postings:

- The level of the user's posting considers the qualities from clients' week-by-week posting of tweets.
- Client-level social collaboration traits by individual client's interaction with social connections with the user's week after week tweet postings (Fig. 2).

Above exhibited structure for recognizing the user's mental pressure states from the user's week-by-week social media data, utilizing tweets' substance and additionally user social communications. While dealing with real world, we needed to consider, Web-based life information as the premise, and we contemplated the relationship between the user's mental pressure states and their social communication practices. The proposed model is a mixture to demonstrate which joins the dynamic factor diagram display (DFGM) with a convolutional neural network (CNN).

5 Results and Analysis

The stress status of individual user by their tweets in social networking is represented in Fig. 3.

Stress status of individual user weekly and their daily tweets on social sites verify their stress levels. We consider sample of users with different states of stress levels; in graph, it states that out of six users two users are non-stressed users with level 0, whereas other users contain different stress levels as 1, 2, 3, etc. (Fig. 4).

If the user's stress level is above 5, it clearly indicates that a particular group of users contain more stress than that of normal users (Fig. 5) by sending some notifications to users for motivating him/her to reduce their stress by watching some inspirational videos, cartoons show, etc. In the research, we found that by following this procedure, the stress levels of stressed users are reduced to some extent and it increased non-stressed users to above 20% in average period of time.

6 Conclusion

The existing techniques utilize either physically named information check or by considering noisy information of interpersonal interaction site, yet while building up a standard or model, some considers the two kinds of named information esteems. Here, in this paper, we suggest a novel approach called hybrid data model to notice the stress levels in individual user and group of users, while considering the tweets of users on social networking. Our hybrid model works for only for one particular

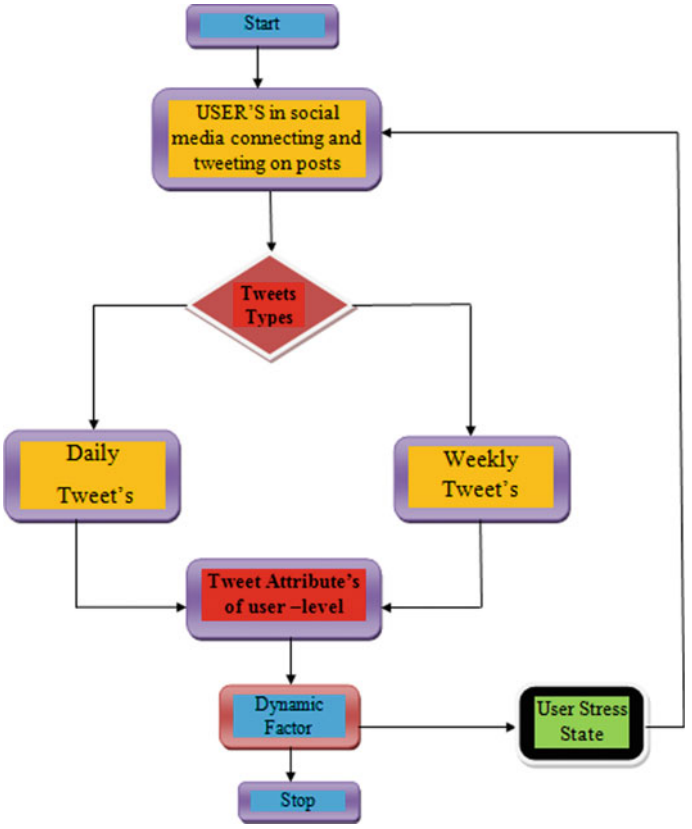


Fig. 1 Flow of user social networking usage of hybrid model

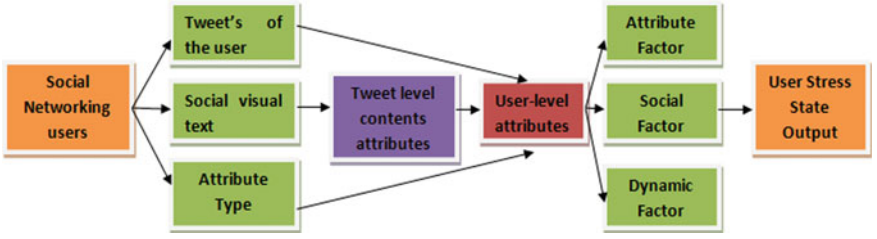


Fig. 2 Architecture of hybrid model for finding stress of users by using tweets

account of user social site; in the future, we mainly focus on multiple accounts of users to find out the stress state.

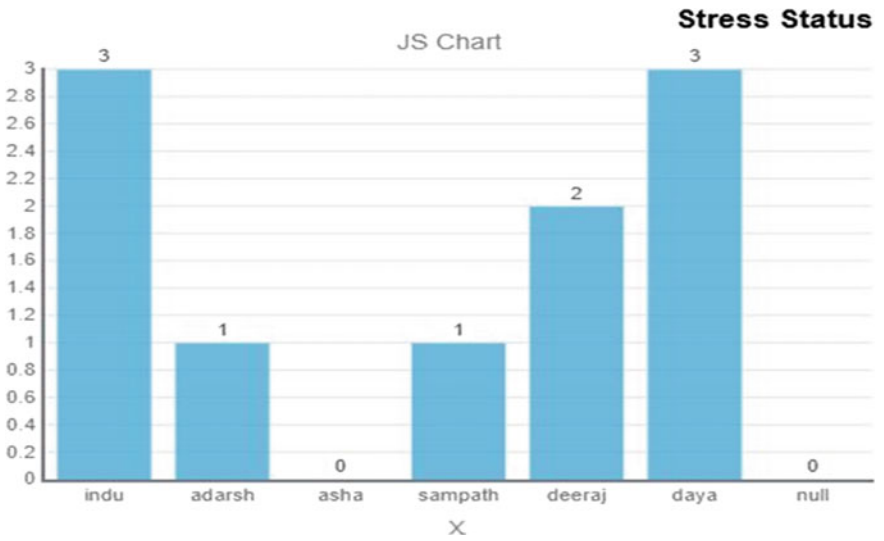


Fig. 3 Stress status of users on weekly and day-to-day tweets in social sites

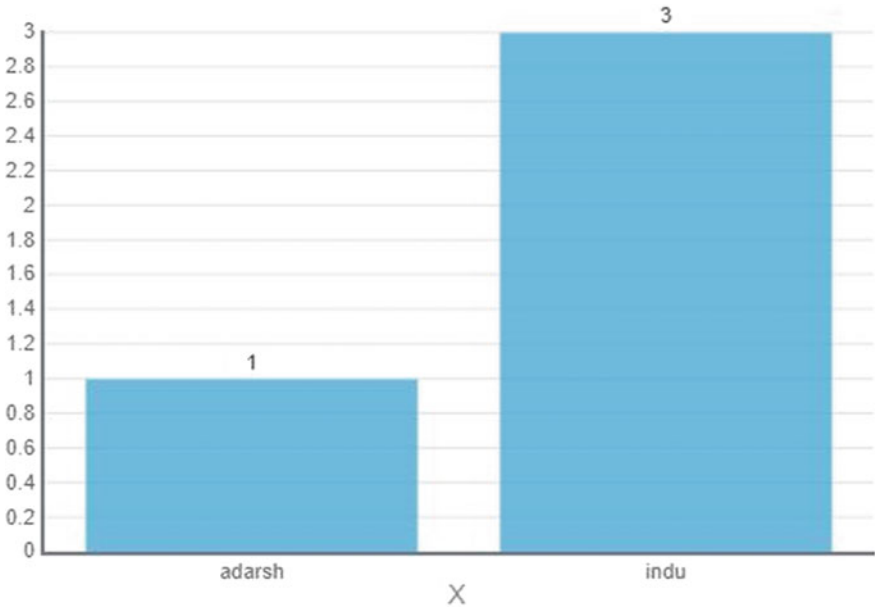


Fig. 4 Stress status of users based on daily tweets

Fig. 5 Top level stressed and non-stressed users

Id	Username	Stress Status
1	indu	3
2	adarsh	1
3	asha	0
4	sampath	1
5	deeraj	2
6	daya	3
7	null	0

References

1. Williams RJ, Zipser D (1989) A learning algorithm for continually running fully recurrent neural networks. Neural Comput 1(2):270–280
2. Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16(12):2639–2664
3. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc Ser B (Stat Method) 73(3):273–282
4. Zhao J, Dong L, Wu J, Xu K (2012) Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1528–1531. ACM; Ng J, Branul T (2007) Perf Comp Bug Navigation Alg J Intell Robot Syst 50(1):73–84
5. Zhang Y, Tang J, Sun J, Chen Y, Rao J (2010) Moodcast: emotion prediction via dynamic continuous factor graph model. In: 2010 IEEE 10th international conference on data mining (ICDM). IEEE, pp 1193–1198
6. Zhang Q, Gonçalves B (2016) Topical differences between chinese language Twitter and sinaweibo. In: Proceedings of the 25th international conference companion on world wide web. International World Wide Web Conferences Steering Committee, pp 625–628
7. Liu X, Li K, Zhou M, Xiong Z (2011) Collective semantic role labeling for tweets with clustering. In: IJCAI, vol 11, pp 1832–1837.

Novel Compact Planar Four-Element MIMO Antenna for 4G/5G Applications

Arumita Biswas and Vibha Rani Gupta

Abstract A low-profile four-element MIMO antenna is proposed in this paper. The designed antenna is planar and can resonate over frequency ranging between 3–5 GHz, covering frequencies allotted for 4G Bands [B22, B42, B43, B48 and B49] and 5G New Radio bands [n77, n78 and n79]. The antenna structure comprises of identical U-shaped and inverted U-shaped antenna elements placed over etched ground slot of 10×5 mm^2. The antenna prototype is designed on FR4 substrate using IE3D simulation software. The MIMO antenna system exhibits good isolation above threshold and envelop correlation coefficient lower than 0.012 without the requirement of any additional decoupling structure.

Keywords Planar antenna · MIMO antenna · 4G antenna · 5G antenna · Envelop correlation

1 Introduction

The mobile communication industry is gearing toward the next big technological leap in form of 5G communication system. The current fourth generation of mobile communication is unable to satisfy the increasing demand of data speed [1]. Researchers and policy makers are working toward achieving a network that will meet the growing data requirements of the subscribers. 5G standard is expected to enhance the data rate up-to 10 Gbps [2, 3]. 3GPP has proposed two broad frequency spectrum that can be used for 5G services—FR1 and FR2 with corresponding frequency range 450 MHz to 6 GHz and 24.25 GHz to 52.6 GHz, respectively [4, 5]. The allocation of 5G C-band for implementing beam-forming and for designing multiple antenna

A. Biswas (✉) · V. R. Gupta
Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: arumita111@gmail.com

V. R. Gupta
e-mail: vrgupta@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_12

system for 5G at World Radio Communication Conference 2015 further triggered the research around the globe in the lower frequency range [6, 7].

Antenna systems have also witnessed enormous transformation to fit to the requirements of advancing mobile communication generation. Modern antenna system must be compact, low-profile and planar, must have ease of fabrication and should employ multiple elements at transmitting and receiving ends [8, 9]. Use of multiple antennas helps to increase the channel capacity; however, due to close placement of antennas inside wireless device, the mutual coupling between them also increases. The increase in mutual coupling degrades antenna performance. Extensive research has been carried out to identify ways of reducing the mutual coupling between antenna elements. Some popularly identified methods are using EBG structures [10, 12, 13], designing decoupling networks [11], including neutralization line [14], using polarization and pattern diversity [15], introduction of extended ground plane [16] and introduction of ground slot [17]. However, these techniques increase the complexity of the final design. Designing compact MIMO antennas with mutual coupling between antenna elements lower than threshold value of −10 dB without increasing the complexity of design is the need of the hour.

In this paper, planar four-element MIMO antenna is designed. The antenna can resonate between frequency range of 3 and 5 GHz covering frequencies allotted for 4G [LTE B-22, B-42, B-43, B-48 and B-49] and 5G [n77, n78 and n79]. Table 1 lists the frequency range of each 4G/5G band supported by the antenna. The mutual coupling between the antenna elements is lower than the threshold without requirement of any extra isolation improvement technique. In the following sections, antenna configuration is presented; antenna characteristics regarding reflection coefficient, mutual coupling, radiation patterns and gain are discussed; envelope correlation coefficient (ECC) calculated from S-parameter is presented, and a comparison of designed antenna with existing work is discussed.

Table 1 Mobile standards supported by designed antenna

S.No.	Supported bands		
	Generation	Band No.	Frequency range (MHz)
1	4G	B22	3410–3500 3510–3600
2		B42	3400–3600
3		B43	3600–3800
4		B48, B49	3550–3700
5	5G	n77	3300–4200
6		n78	3300–3800
7		n79	4400–5000

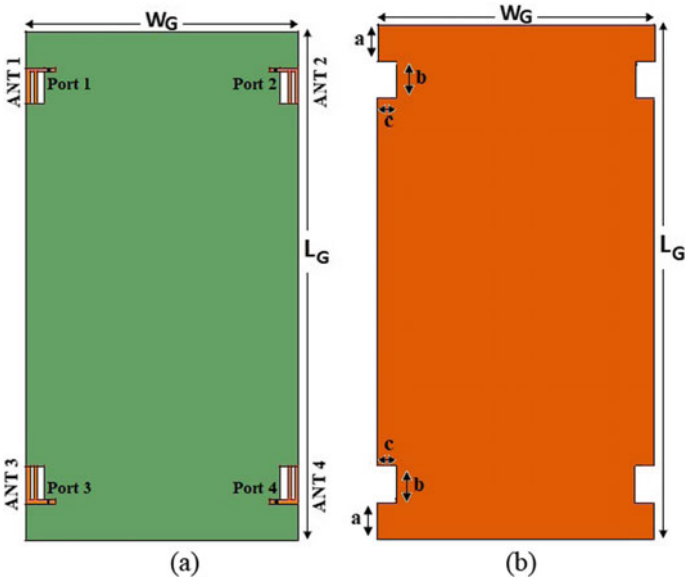


Fig. 1 Geometry of proposed antenna: **a** front plane, **b** back plane

2 Antenna Configuration

The geometry of the proposed MIMO antenna is illustrated in Fig. 1. FR4 substrate having dielectric constant 4.4, loss tangent 0.02 and thickness 1.6 mm is used as the base. The substrate is selected to have a size of 140 mm × 75 mm. The system ground as illustrated in Fig. 1b has same dimension as that of the FR4 substrate. Four slots of dimension 10 mm × 5 mm are etched at a distance of 10 mm from the edges in the ground plane.

Figure 2 illustrates the U-shaped and inverted U-shaped monopole antenna elements designed to resonate over the required frequency spectrum. These antennas are designed to cover the frequency spectrum between 3 and 5 GHz (considering S_{xx} reference as −6 dB).

The unit antenna elements are printed on the top surface of FR4 substrate over the etched ground slots as illustrated in Fig. 1a. Feed-line of dimension 5 × 1 mm^2 is used for connecting the 50 Ω probe feed to each of the unit antennas. Table 2 lists the dimensions of the final designed antenna prototype.

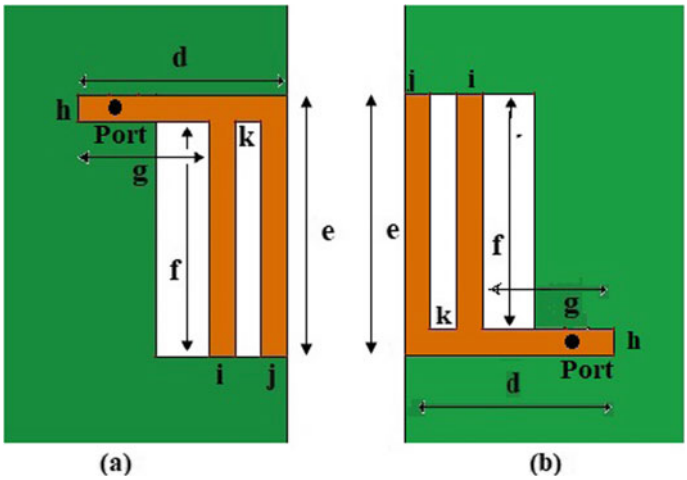


Fig. 2 Geometry of unit antenna: **a** inverted U-shaped antenna, **b** U-shaped antenna

Table 2 Detailed dimension of the proposed antenna

Parameter	L_G	W_G	a	b	c	d	e
Unit (mm)	140	75	10	10	5	8	10
Parameter	**f**	**g**	**h**	**i**	**j**	**k**	
Unit (mm)	9	5	1	1	1	1	

3 Results and Discussion

A. *Reflection Coefficient and Mutual Coupling*

The designed four-element MIMO antenna was designed and optimized using IE3D simulator. Figure 3 illustrates the reflection coefficients (S_{11}, S_{22}, S_{33} and S_{44}) characteristics obtained for the four designed antennas in MIMO system. It can be noted from the figure that for S_{xx} reference taken less than −6 dB (VSWR 3:1), the designed four-element MIMO antenna system can resonate over the desired frequency range between 3 and 5 GHz covering multiple 4G and 5G bands.

The mutual coupling between the antennas was considered next. Figure 4 illustrates the mutual coupling between the respective antenna elements in terms of transmission coefficients. It can be observed that due to strategic placement of the antenna over the long arm of the ground no further isolation improvement technique is required. Mutual coupling value between all the antennas for the designed frequencies is much lower than the threshold of -10 dB.

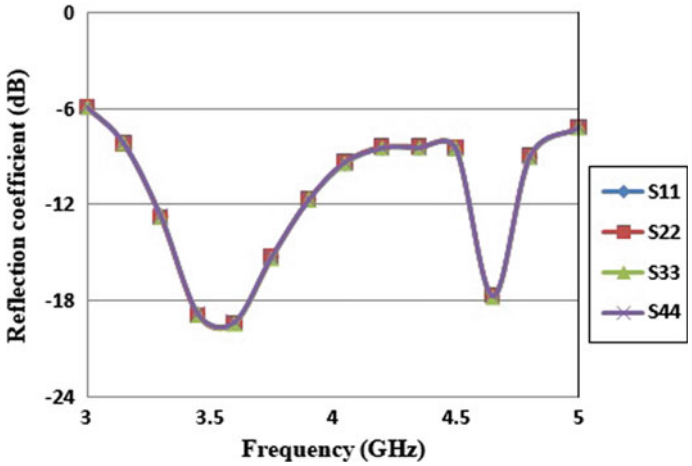


Fig. 3 Reflection coefficient of MIMO antenna

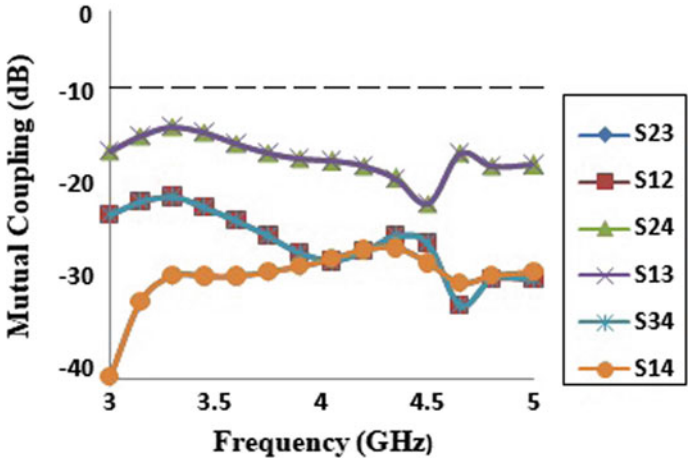


Fig. 4 Mutual coupling between the antenna elements

B. *Radiation Pattern and Gain*

The radiation pattern of the MIMO antenna system was recorded in both azimuth plane and elevation plane for two sample frequencies $f_1 = 3.5$ GHz and $f_2 = 4.8$ GHz. Figure 5a, b illustrates the obtained normalized radiation patterns. In the azimuth plane, the radiation is uniform in all direction, while in elevation plane, the maximum radiation is obtained at 30° and 150°, indicating an omnidirectional radiation pattern. In wireless communication, omnidirectional radiation is considered most suitable. The gain for the antenna was studied next,

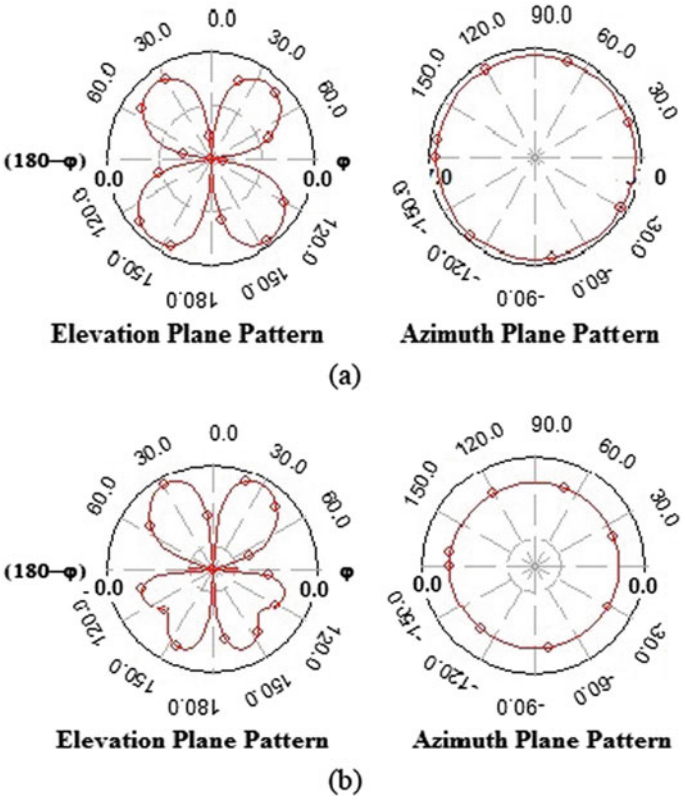


Fig. 5 Radiation pattern in azimuth and elevation plane: **a** at 3.5 GHz; **b** at 4.8 GHz

and maximum gain corresponding to every 4G/5G bands covered by the designed prototype was recorded. Table 3 lists the maximum gain value for each band. It can be noted that the maximum gain of 4.58 dBi was recorded at 4.05 GHz in 5G n77 band.

Table 3 Maximum gain in bands covered

LTE/5G bands	Maximum gain (dBi)
B-22	3.54
B-42	3.54
B-43	4.25
B-48	4.25
B-49	4.25
n77	4.58
n78	4.25
n79	2.33

C. Diversity Performance

The envelope correlation coefficient is a tool to judge the diversity performance between the identical antennas in MIMO system. As per industrial standard, ECC value lower than 0.5 is considered desirable [18–20].

ECC (ρ) between two antenna elements Ant1 and Ant 2 can be computed from S-parameters using the Eq. (1). The equation is used to compute the ECC value between all the possible combinations in four-element MIMO antenna system. Figure 6 illustrates the computed ECC value corresponding to the frequencies between the antennas. It was obtained that for the entire range of frequency under consideration maximum ECC value was obtained as 0.012 at 4.05 GHz between antenna element—Ant1 and Ant 3 as well as Ant 2 and Ant 4. For all other combination of antenna elements, ECC value was lower than 0.005.

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

D. Comparison with existing works

Table 4 draws a comparison of the proposed four-element MIMO antenna with some works proposed by other researchers in same domain. It can be observed that the proposed antenna covers more number of 5G bands proposed for future 5G services with at par computed ECC value between the identical antenna elements, indicating that the design is efficient for future wireless device.

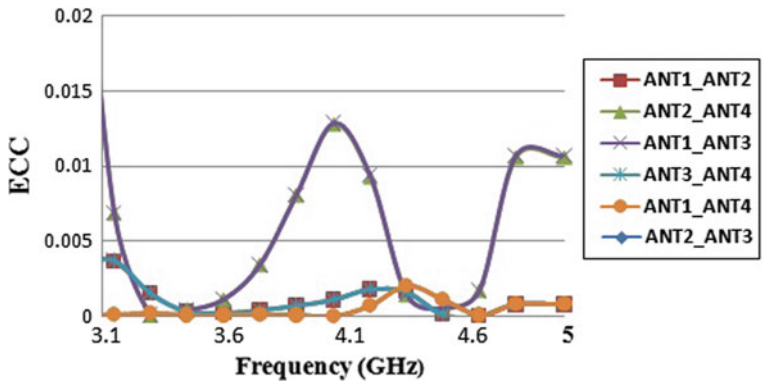


Fig. 6 Computed envelope correlation coefficient of proposed MIMO antenna

Table 4 Comparison of proposed antenna with existing antennas

Design	Year	Substrate used	Frequency covered (GHz)	5G bands covered	ECC (Less than)
Proposed	2019	FR4 substrate (h-1.6 mm, ε_r-4.4)	3–5	n77, n78 and n79	0.012
[3]	2019	FR4 substrate (h-1.5 mm, ε_r-4.4)	2.7–3.6	Partially n77 and n78	0.009
[6]	2017	FR4 substrate (h-0.8 mm, ε_r-4.4)	3.4–3.6	Partially n77 and n78	0.027
[7]	2019	FR4 substrate (h-0.8 mm, ε_r-4.4)	3.4 –36	Partially n77 and n78	0.02
[9]	2019	FR4 substrate (h-1.6 mm, ε_r-4.4)	3.3–3.9	Partially n77 and n78	0.01

4 Conclusion

This chapter proposes a simple planar four-element MIMO antenna system that can resonate between 3 and 5 GHz covering 4G LTE bands [B22, B42, B43, B48 and B49] and 5G bands [n77, n78 and n79]. The designed unit antenna elements comprised of U-shaped and inverted U-shaped antennas placed over etched ground slots. Owing to the careful placement of antennas along the long arm of substrate, no further isolation improvement technique was required. Omnidirectional radiation pattern, maximum gain of 4.58 dBi at 4.05 GHz and ECC value lower than 0.012 obtained over the entire range of frequency under consideration qualify the proposed antenna as a good candidate to be integrated inside future wireless devices.

References

1. Chen WS, Lin YC (2017) Design of 2 × 2 microstrip patch array antenna for 5G C-band access point application. In: 2017 international conference on computing methodologies and communication (ICCMC). Erode, India. https://doi.org/10.1109/ICCMC.2017.8282576
2. Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong ACK, Zhang JC (2014) What will 5G be? IEEE J Sel Areas Commun 32(6):1065–1082. https://doi.org/10.1109/JSAC.2014.2328098
3. Chatta HT (2019) 4-Port 2-element MIMO antenna for 5G portable applications. IEEE Access 7:96516–96520. https://doi.org/10.1109/ACCESS.2019.2925351
4. 3rd Generation Partnership Project—5G; NR; User Equipment (UE) radio transmission and reception; Part1: Range 1 Standalone, TS 38.101–1 version 15.2.0 Release 15, July 2018
5. 3rd Generation Partnership Project—5G; NR; Base Station (BS) radio transmission and reception, ETSI TS 138 104 v15.3.0, Oct 2018

6. Abdullah M, Ban YL, Kang K, Kf Sarkodie OK, Li MY (2017) Compact 4-port MIMO antenna system for 5G mobile terminal. In: International applied computational electromagnetics society symposium, Florence, Italy.https://doi.org/10.23919/ROPACES.2017.7916045
7. Liu Y, Lu Y, Zhang Y, Gong S (2019) MIMO antenna array for 5G smartphone applications. In: 13th European conference on antennas and propagation, Poland
8. Yang HH, Quel YQS (2017) Massive MIMO meet small cell. In: Springer briefs in electrical and computer engineering. Springer International Publishing. https://doi.org/10.1007/978-3-319-43715-6
9. Parchin NO et al (2019) Eight-element dual-polarized MIMO slot antenna system for 5G smartphone applications. IEEE Access 7:15612–15622. https://doi.org/10.1109/ACCESS.2019.2893112
10. Yang F (2003) Microstrip antennas integrated with electromagnetic band-gap structures: a low mutual coupling design for array applications. IEEE Trans Antennas Propag 51(10):2936–2946. https://doi.org/10.1109/TAP.2003.817983
11. Zou XJ, Wang GM, Wang YW, Li HP (2019) An efficient decoupling network between feeding points for multielement linear arrays. IEEE Trans Antennas Propag 67(5):3101–3108. https://doi.org/10.1109/TAP.2019.2899039
12. Ghosh S, Tran TN, Ngoc TL (2014) Dual-layer EBG-based miniaturized multi-element antenna for MIMO systems. IEEE Trans Antennas Propag 62(8):3985–3997. https://doi.org/10.1109/TAP.2014.2323410
13. Islam MT, Alam MS (2013) Compact EBG structure for alleviating mutual coupling between patch antenna array elements. Progr Electromagn Res 137:425–438
14. Zhang S, Pedersen GF (2015) Mutual coupling reduction for UWB MIMO antennas with a wideband neutralization line. IEEE Antennas Wirel Propag Lett 15:166–169. https://doi.org/10.1109/LAWP.2015.2435992
15. Boukarkar A, Lin XQ, Jiang Y, Nie LY, Mei P, Yu YQ (2017) A miniaturized extremely close-spaced four-element dual-band MIMO antenna system with polarization and pattern diversity. IEEE Antenna Wirel Propag Lett 17(1):134–137. https://doi.org/10.1109/LAWP.2017.2777839
16. Chen WS, Chang YL (2018) Small size 5G C-band/WLAN 5.2/5.8 GHz MIMO antenna for laptop computer applications. In: Proceedings of IEEE international workshop on electromagnetics: applications and student innovation competition, Japan
17. Sharawi MS, Ikram M, Shamim A A two concentric slot loop based connected array MIMO antenna system for 4G/5G terminals. IEEE Trans Antennas Propag 65(12):6679–6686. https://doi.org/10.1109/TAP.2017.2671028
18. Votis C, Tatsis G, Kostarakis P (2010) Envelope correlation parameter measurements in a MIMO antenna array configuration. Int J Commun Netw Syst Sci 3(4):350–354
19. Biswas A, Gupta VR (2019) Design of penta-band MIMO antenna for GPS/2G/3G/4G and 5G NR applications. Int J Recent Technol Eng 8(1):1935–1940
20. Sharawi MS (2014) Printed MIMO antenna systems: performance metrics, implementations and challenges. Forum Electromagn Res Methods Appl Technol 1:1–11

Design of High-Speed Binary Counter Architecture for Low-Power Applications

Mangal Deep Gupta, Saurabh Kumar Singh, and R. K. Chauhan

Abstract This chapter presents a design of high-speed binary counter architecture using clock gating for low-power applications. Clock gating techniques enables in improving the latency and power dissipation of proposed binary counter. The latency of proposed architecture is lower as compared to conventional architecture, which shows that the proposed design can be operated at high input frequencies. The proposed binary counter design of 4-, 8-, and 16-bit has been built by Verilog HDL code and simulated using Questa Simulator of Mentor Graphics. For synthesis of proposed design, LeonardoSpectrum tool by mentor Graphics is used and synthesis of it is based on CMOS process TSMC 0.35 μm, Spartan 6, and Spartan 3E FPGA. The Semicustom physical layout for Proposed 8-bit counter architecture using 350-nm Standard CMOS process is also obtained in this work.

Keywords Binary counter · Clock gating · Latency · Power dissipation · Operating frequency · FPGA · Layout · Verilog HDL

1 Introduction

In many VLSI system, binary counters are basic building blocks. A n-bit binary counter design consist a series of k-flip-flops, and its count value can be 0to$(2^k - 1)$ [1]. Fast binary counter design is the basic point of concern, when designing of high-speed digital system for various applications. Such as counting time for process allocation in scheduling, analog to digital conversion [2–4], time to digital

M. D. Gupta (✉) · S. K. Singh · R. K. Chauhan
Department of Electronics and Communication Engineering, MMMUT, Gorakhpur, U.P, India
e-mail: mangaldeepgcet@gmail.com

S. K. Singh
e-mail: srbhsingh448@gmai.com

R. K. Chauhan
e-mail: rkchauhan27@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_13

conversion [5–9], etc. It can be used as clock dividers (used in on-chip processor because sometimes processor work at lower processor than actual frequency of the processor), etc.

Design techniques of counter based on clock gating are categorized into asynchronous and synchronous counter. In synchronous counter, all flip-flops are operated by same clock pulse; however, in asynchronous counters design, LSB flip-flop operated by master clock signal and output of each flip-flop act as a clock signal for succeeding flip-flops. Asynchronous counters are fundamentally different from synchronous counter because it can offer advantages [10] in the following areas such as low power dissipation; high operating speed; improved composability and modularity; no clock skew and clock distribution problems. Performance of synchronous counter such as delay is dependent on the counter bit length, thus latency of synchronous counter increases linearly with bit length. Another major issue, when modulus counters system is made by synchronous counters, is that it requires the fast detection circuits for modulus counting. And time period of clock signal is limited by detection circuits.

Counters design with low power dissipation, high operating speed and high reliability is important designing parameters. Its power dissipation can be decreases by decreasing the power associated with clock signal, so that wrong counter design decision may result in degradation of performance of overall system. The major source of power dissipation in counter is clock source due to high driving load from its nets or wire, and clock signal is continuously present across flip-flops which are not required to change their state, contributing a large power consumption. In counter application, clock signal consumes on average 25–40% of total power which is observed [11]. To optimize the power dissipation of counter due to clock signal, clock gating is appropriate techniques to overcome the high power dissipation by excluding unnecessary clock activities.

There are several efficient techniques for counter design present in literature. A counter design using clock gating is presented by Wu et al. [12]. A power-efficient Johnson counter with clock gating is proposed in [13]. however, [12, 13] required extra hardware to implement clock gating network, interconnect complexity and clock buffer network complexity increase with the counters width. A priority encoder-based counter design for low-power application was proposed in [14]; it is based on compressed the numerous inputs into single outputs. A pulse latch base ring counter design is presented in [16]. But drawback associated with [14–16] required additional hardware in terms of pulse generator and priority encoder; it is enforcing the restriction on design of wide bit binary counters. For optimization of power, a quasi-synchronous counter design is presented in [16]. In this approach, clock signal for flip-flops is derived from main source of clock signal, i.e., isolating the flip-flops from unwanted transitions of clock signal for optimizing the power consumption. Adiabatic logic and pass transistor-based counter design for low power dissipation is proposed in [17–19]. Katreepalli and Haniotakis proposed a power-efficient techniques for counters design. It is reducing the power consumption due to clock distribution arrangements for different flip-flops [19]. But downside of this architecture

is clock gating complexity which increases with counter bit length. So, it degrades the timing performance of counter with bit length of counters.

To design of VLSI circuits, mainly two possible solutions are used such as application-specific integrated circuits (ASICs) and another, i.e., programmable logic device (PLD, FPGA)-based design [20]. ASICs designs have many advantages with respect to other solutions such as low power dissipation, high speed, low cost for mass production, better controlling characteristics of input–outputs and high compact ICs design. However, there are some disadvantages like costly production for low volume due to high investment in CAD tools, workstations, and production manpower while FPGA can be implement just any hardware design. Circuits implementation in FPGA can be implemented in short amount of time comparatively, because it does not require fabrication process such as physical layout process, mask making, low NRE costs, and short time to market. So, FPGAs have been progressively used as the final product platforms.

In this brief, we present a design of high-speed asynchronous binary counter with efficient clock gating techniques for low-power applications. Here, we used the tristate buffer to minimize clock activity at each flip-flops. Clock gating techniques enable in improving the latency and power dissipation of the proposed binary counter. The remaining brief is organized as follows: Sect. 2 presents the proposed binary counter design technique. Section 3 presents a proposed binary counter design for 4-, 8-, and 16-bit. Section 4 presents a semicustom physical layout for proposed 8-bit counter architecture. Section 5 presents the performance comparison between 0.35 μm-CMOS process, Spartan 6, and Spartan 3E FPGA. Finally, Sect. 6 concludes the chapter.

2 Proposed Design of Binary Counter

To overcome drawback of previously reported work, we propose a power efficient, high speed design of asynchronous counter with minimum hardware overhead. The proposed design of counter employs an efficient tristate buffer logic for clock gating. There are various methods to design a binary counter, among them D flip-flops are used widely since to reduce hardware complexity of counter, convert the D-flip-flops into T-flip-flops using XOR gate that offers toggling at each clock pulse. It is the fundamental step to design a counter. For n-bit counter design, '*n*' number of flip-flops are required. Figure 1 shows the 4-bit proposed asynchronous up–down counter in which every flip-flop is made by D-flip-flops and XOR gate. First input each XOR gate is fixed by logic one, and second input becomes a feedback through output (Q). It can be observed that XOR gate's inputs of all the flip-flops are set to logic '1'; therefore, in every negative edge (for up-counter), the output will be inverted. In conventional architecture LSB flip-flop is drive by system clock signal, and rest of flip-flop is driven by previous output signal directly. But in proposed design, all flip-flops (except LSB), driven by tristate buffer outputs and each tristate buffer, are controlled by previous output signal. In this operation, tristate buffer passes the

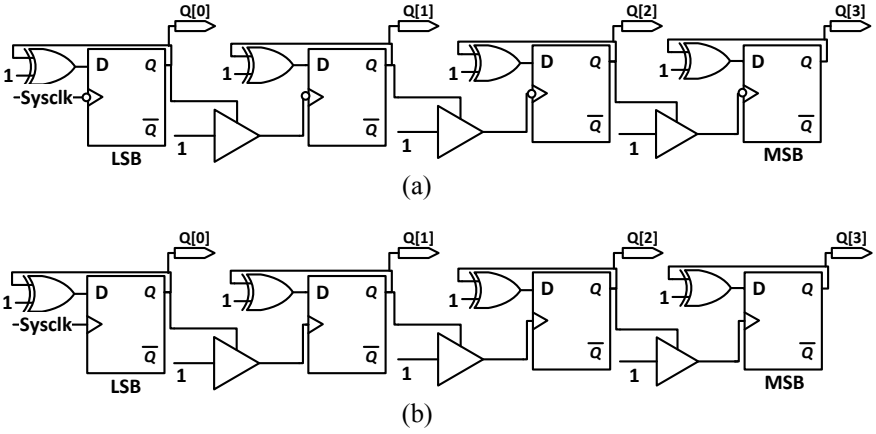


Fig. 1 Proposed architecture of 4-bit binary counter. **a** Up-counter. **b** Down-counter

logic '1' when it is active; otherwise, its output is high impedance state (z). It is effectively optimizing the power dissipation due to reduce transition activity at each flip-flops. This work has been design by Verilog HDL, and its synthesis was done by LeonardoSpectrum tool. Figure 2 shows the schematics of proposed 4-bit binary counter.

3 Methodology for Layout Design

Toolbar at the top of this work was evaluated on Mentor Graphics tool. The summary of work flow details is mentioned here as:

(i) HDL code in Verilog for proposed algorithm is written using Questa Simulator tool of mentor graphics. Functional verification is carried out using register transfer level (RTL) design.

(ii) Synthesis on the RTL code using LeonardoSpectrum tool is carried out by generating Netlist and Standard Delay Format (SDF) file. Here, Netlist refers to textual description of connection of gates. It is then used to produce the layout of the chip. SDF file is the representation and interpretation of timing data for use at any stage of an electronic design process, and it is used for dynamic and static time analysis. For this synthesis operation, select the technology file as TSMC 0.35μm and specify the frequency.

(iii) After synthesis on RTL, we perform the post-synthesis verification of Netlist that can be done by Equivalence Checking or FormalPRO tool, which operates and verifies the Netlist which is generated from LeonardoSpectrum tool. For verification of Netlist, one requires dot v (.v) and sdf file.

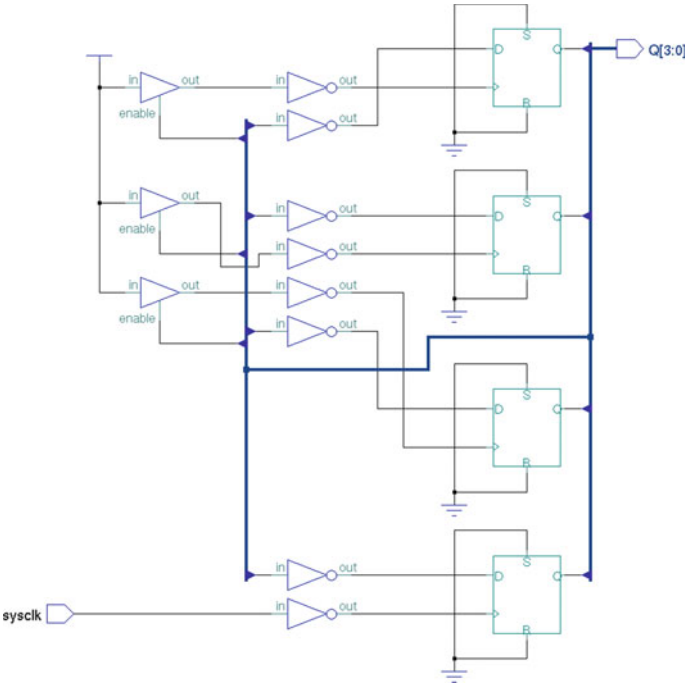


Fig. 2 Schematics diagram of proposed 4-bit binary counter

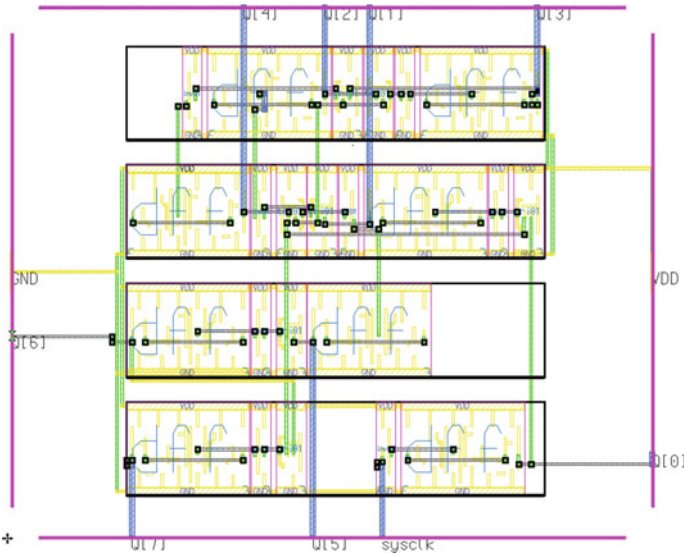


Fig. 3 Semicustom layout of proposed 8-bit binary counter using 350 nm standard CMOS process

(iv) To generate layout for Netlist of HDL code, one can invoke Pyxis Design tool. In this operation, first we select the files to initiate layout operation such as Verilog Netlist, process and rule file. At this stage, the gate-level Netlist is changed into a complete physical geometric representation. For physical design, the following steps are necessary: floorplanning for placing the various blocks, I/O pads across the chip area based on the design constraints, placement of physical elements inside each block are done. Thereafter, the global and detailed routing has been done to connect each element.

(v) After layout, import GDSII (GDS2) file from this tool, which is used by the foundry to fabricate the silicon. The layout should be done according to the silicon foundry design rules.

(vi) At this stage, a physical verification is required using calibre DRC (Design Rule Check) to ensure the layout phase.

4 Result and Discussion

The proposed design of asynchronous counter using tristate buffer has been built by Verilog HDL code and simulated using Questa Simulator tool of Mentor Graphics. For synthesis of proposed design, we used LeonardoSpectrum tool, and it is based on TSMC 0.35 μm-CMOS technology, Spartan 6, and Spartan 3E FPGA. To estimate the on-chip power dissipation corresponding FPGA design, we used the X-power analyzer tool provided by Xilinx. For this analysis, we write the Verilog test bench to generate VCD file for 4-, 8-, and 16-bit proposed binary counter. In this paper, we also analyzed the latency and maximum operating frequency of proposed and previous reported binary counter [19] for both ASIC- and FPGA0-based design. Maximum operating frequency will be calculated by critical path delay. And latency will be calculated by delay between two consecutive counting sequence of counter. For synthesis of counter, we used the 100 and 50 MHz operating frequency correspondingly FPGA- and ASIC-based design.

Table 1 shows the synthesis results based on ASIC design of proposed and existing counter. For this synthesis, we used TSMC 350-nm CMOS technology. In this synthesis, we explore the latency, maximum operating frequency and area complexity

Table 1 Synthesis report at 50 MHz clock frequency using 350-nm CMOS technology

Counter	Proposed counter			Counter [19]		
	Latency (ns)	Maximum operating frequency (MHz)	No. of gates	Latency (ns)	Maximum operating frequency (MHz)	No. of gates
4-bit	0.77	1488.5	26	1	1000	53
8-bit	0.77	1488.5	53	1.04	960	116
16-bit	0.77	1488.5	106	1.53	892	235

Table 2 Spartan 6 FPGA synthesis report at 100 MHz clock frequency

Counter	Proposed counter			Counter [19]		
	Latency (ns)	On-chip power dissipation (mW)	Maximum operating frequency (MHz)	Latency (ns)	On-chip power dissipation (mW)	Maximum operating frequency (MHz)
4-bit	1.95	93.56	512.794	2.048	94.65	488.317
8-bit	1.95	98.23	512.794	2.078	98.74	481.231
16-bit	1.95	98.61	512.794	2.106	98.94	474.721

(in terms of utilized logic gates). In this analysis, we find the better latency and maximum operating frequency due to less hardware complexity as compare with previous counter [19]. Also, we found the latency of proposed counter does not depend on counter bit length but latency of clock gating-based synchronous counter [19] is directly proportional to the counter bit length.

Table 2 shows the synthesis results based on FPGA design for proposed and existing counter. For this synthesis, we used the Spartan-6 FPGA. Similarly, in this synthesis, we explore the latency and maximum operating frequency for proposed counter and existing counter [19]. From this analysis, we find the better latency and maximum operating frequency due to less LUTs utilize in FPGA chip as compare with previous counter [19]. Also, we found the latency of proposed counter does not depend on counter bit length while latency of clock gating-based synchronous counter [19] is directly proportional to the counter bit length. On-chip power dissipation of proposed counter is approximately equal or less than clock gating-based synchronous counter [19] while their speed has been enhanced due to utilizing the tristate buffer and arrangement of flip-flops.

Table 3 shows the synthesis results based on Spartan-6 FPGA for proposed and existing counter. From this analysis, we explore the latency, maximum operating frequency, and on-chip power dissipation for proposed counter and existing counter [19]. From this analysis, we find the better latency and operating frequency due to less LUTs utilize in FPGA chip as compare with previous counter [19]. Moreover, we find the latency of proposed counter does not depend on counter bit length while

Table 3 Spartan 3E FPGA synthesis report at 100 MHz clock frequency

Counter	Proposed counter			Counter [19]		
	Latency (ns)	On-chip power dissipation (mW)	Maximum operating frequency (MHz)	Latency (ns)	On-chip power dissipation (mW)	Maximum operating frequency (MHz)
4-bit	1.689	111.91	592.084	2.916	112.87	342.977
8-bit	1.689	116.58	592.084	2.949	117.63	339.12
16-bit	1.689	116.66	592.084	2.949	118.13	339.12

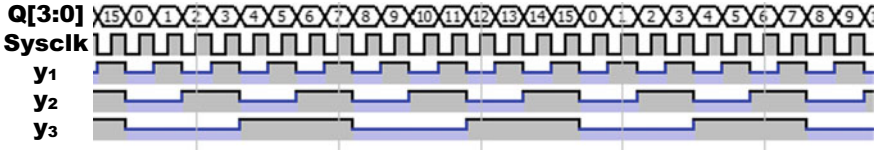


Fig. 4 Simulation waveform of proposed 4-bit proposed binary counter

latency of clock gating-based synchronous counter [19] is directly proportional to the counter bit length. On-chip power dissipation of proposed counter is slightly less than clock gating-based synchronous counter [19]. Finally, we enhanced the performance of proposed counter as compare with previous counter utilizing the tristate buffer and modified arrangement of flip-flops.

Figure 4 shows simulation result of proposed 4-bit binary counter where 'sysclk' is the clock signal to drive LSB flip-flop. Signals y_1, y_2 and y_3 are the clock source of consecutive flip-flop. And Q[3 : 0] is the output of 4-bit binary counter. This waveform clearly shows that signals y_1, y_2 and y_3 do not have transition from 1 to 0 due to behavior of tristate buffer that overcomes the dynamic power dissipation by transitions of clock signals. This signal is active only when flip-flops are required to change their state, and rest of the cases, it behaves as a high impudence value. This approach effectively overcomes the high power dissipation.

Figure 5 shows the compression between proposed and previous reported counter [19] in terms of on-chip power dissipation, maximum operating frequency, and latency. Figure 5a represents the graphical analysis of on-chip power dissipation of counter architecture corresponding Spartan-6 and Spartan-3E FPGA. In this analysis, we find the Spartan-6 FPGA dissipate less power than Spartan-3E. Figure 5b shows the analysis of maximum operating frequency correspondingly 350-nm CMOS technology, Spartan-6, and Spartan-3E FPGA. From this graph, first we analyze the ASIC design using 350-nm CMOS have been operated at higher frequency than FPGA-based design. Also, we found the operating frequency of proposed counter is higher than clock gating-based synchronous counter [19]. And Fig. 5c represents the latency analysis of counter with respect to 350-nm CMOS technology, Spartan-6, and Spartan-3E FPGA. It is concluded that the CMOS-based ASIC design has a minimum latency than FPGA-based design. Finally, we find the 350-nm CMOS-based counter designs have been superior in terms of latency and maximum operating frequency than FPGA. Once we compare between Spartan-6 and Spartan-3E FPGA, we find the latency and maximum operating frequency of binary counter are better for Spartan-3E FPGA and on chip-power dissipation of counter is better for Spartan-6 FPGA.

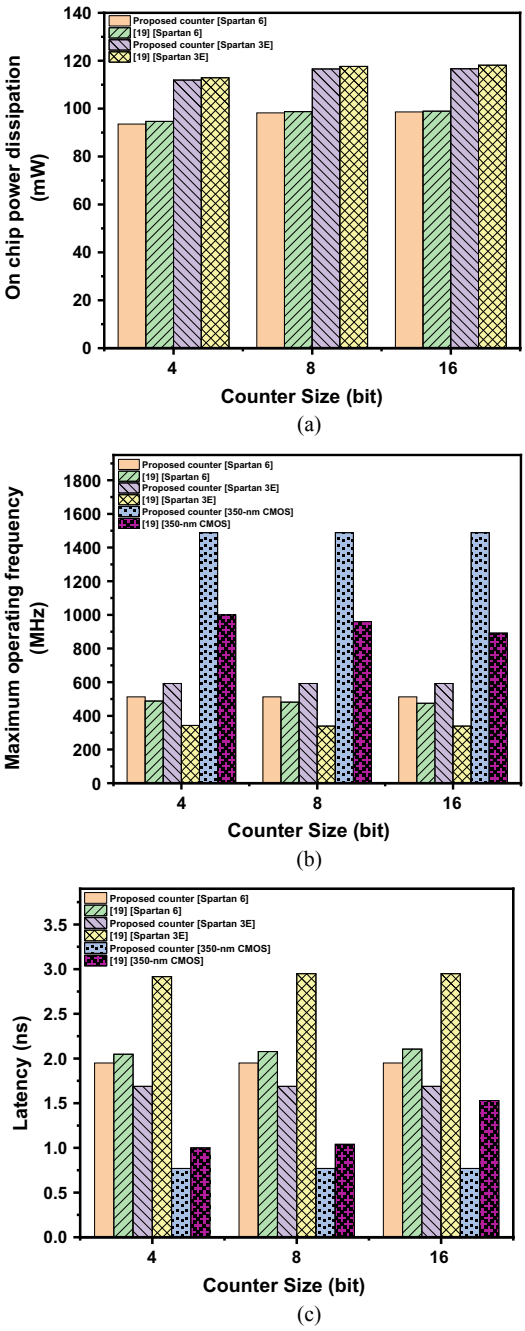


Fig. 5 Compression between proposed and previous reported counter [19]. **a** For on-chip power dissipation. **b** For maximum operating frequency. **c** For latency

5 Conclusion

In this work, a design of binary counter architecture using tristate buffer is proposed to minimize the latency and maximize the operating frequency of counter with minimum on-chip power dissipation. Synthesis of proposed counter with different bit length has been done using 0.35 μm CMOS technology, Spartan 6, and Spartan 3E FPGA. We find the 0.35 μm CMOS-based counter design has been superior in terms of latency and maximum operating frequency than FPGA. Once we compare between Spartan-6 and Spartan-3E FPGA, we find the latency and maximum operating frequency of binary counter are better for Spartan-3E FPGA and on-chip power dissipation of counter is better for Spartan-6 FPGA. The synthesis results show that the latency and maximum operating of proposed counter does not change with counter bit length. While, latency and maximum operating frequency of clock gating-based counter [19] degraded with the bit length of counter. When proposed counter compared with the earlier reported counter [19], we find the latency of proposed counter reduces by 33% when using 0.35 μm CMOS technology, whereas maximum operating frequency increases by 48%. Latency of proposed counter reduces by 42% when using Spartan-3E FPGA, whereas maximum operating frequency increases by 72%.

References

1. Rabaey JM, Chandrakasan A, Nikolic B (2003) Digital integrated circuits: a design perspective. Prentice Hall
2. Peizerat A, Rostaing JP, Zitouni N, Baier N, Guellec F, Jalby R, Tchagaspanian M (2012) An 88 dB SNR, 30 μm pixel pitch infra-red image sensor with a 2-step 16 bit A/D conversion. In: Symposium on VLSIC, pp 128–129
3. Matsuo S, Bales TJ, Shoda M, Osawa S, Kawamura K, Andersson A, Haque M, Honda H, Almond B, Mo Y, Gleason J, Chow T, Takayanagi I (2009) 8.9-Megapixel video image sensor with 14-b column-parallel SA-ADC. IEEE Trans Electron Dev 56(11):2380–2389
4. Akyurek F, Bayram B (2017) Digital output ROIC with single slope ADC for cooled infrared applications. Microsyst Technol 3(4):1091–1099.https://doi.org/10.1007/s00542-016-2963-7
5. Wu J, Shi Z The 10-ps wave union TDC: improving FPGA TDC resolution beyond its cell delay. In: Proceedings IEEE nuclear science symposium conference record, Oct 2008, pp 3440–3446
6. Jachna Z, Szplet R, Kwiatkowski P, Ró˙zyc K (2015) Parallel data processing in a 3-channel integrated time-interval counter. Meas Autom Monitor 61(7):308–310
7. Yu W, Kim K, Cho S (2015) A 0.22 ps rms integrated noise 15 MHz bandwidth fourth-order time-to-digital converter using time-domain error-feedback filter. IEEE J Solid-State Circuits 50(5):1251–1262
8. Klepacki K, Szplet R, Pełka R (2014) A 7.5 ps single-shot precision integrated time counter with segmented delay line. Rev Sci Instrum 85(3):034703
9. Szplet R, Klepacki K (2010) An FPGA-integrated time-to-digital converter based on two-stage pulse shrinking. IEEE Trans Instrum Meas 59(6):1663–1670
10. Nielsen LS, Sparso J (1999) Designing asynchronous circuits for low power: an IFIR filter bank for a digital hearing aid. Proc IEEE 87(2):268–281. https://doi.org/10.1109/5.740020
11. Gupta MD, Chauhan RK (2020) Performance Investigation of Binary Counter with Different Clock Gating Networks. J Telecommun Electron Comput Eng (JTEC) 12(2):59–67

12. Wu Q, Pedram M, Wu X (2000) Clock-gating and its application to low power design of sequential circuits. IEEE Trans Circuits Syst I 47(3):415–420
13. Ismail SM, Rahman ABMS, Islam FT (2012) Low power design of Johnson counter using clock gating. In: 2012 15th international conference on computer and information technology (ICCIT). Chittagong, 510–517
14. Wu X, Pedram M (2000) Low power sequential circuit design using priority encoding and clock gating, power electronics and design. In: Proceedings of the 2000 international symposium on ISLPED '00, pp. 143–148
15. Doi T, Niranjan V (2016) Low power and high performance ring counter using pulsed latch technique. In: 2016 international conference on micro-electronics and telecommunication engineering (ICMETE). Ghaziabad, pp 584–586
16. Wu X, Wei J, Pedram M (2000) Low-power design of sequential circuits using a quasi-synchronous derived clock. In: Proceedings 2000. design automation conference (IEEE Cat. No.00CH37106). Yokohama, Japan, pp 345–350
17. Hu JP, Xu TF, Xia YS (2005) Low-power adiabatic sequential circuits with complementary pass-transistor logic. In: IEEE MWSCAS05, pp 1398–401
18. Gautam M, Nirmal U, Jain R (2016) Low power sequential circuits using improved clocked adiabatic logic in 180 nm CMOS processes. In: 2016 international conference on research advances in integrated navigation systems (RAINS), Bangalore, pp 1–4
19. Katreepalli R, Haniotakis T (2019) Power efficient synchronous counter design. Comput Electric Eng 75:288–300. ISSN 0045–7906, https://doi.org/10.1016/j.compeleceng.2018.01.001
20. Synthesis of Arithmetic Circuits. FPGA, ASIC, and Embedded Systems, Wiley, Inc., Publication

Bidirectional Audio Transmission in Optical Wireless Communication Using PWM and Class D Amplifier

Basudeb Das, Shibabrata Mukherjee, and Saswati Mazumdar

Abstract Optical wireless communication (OWC) using light-emitting diode (LED) is a supplement of modern radio frequency wireless communication scheme due to high bandwidth (>300 GHz) of the optical domain. It is less harmful compared to RF wave, and it fulfills both illumination and data transmission at a time in indoor. OWC experiences many schemes starting from amplitude modulation to subcarrier index modulation-orthogonal frequency division multiplexing (SIM-OFDM) technique, and improvements are still going on for achieving more secured and higher data rated (in Gigabit Class) communication system. Here, an audio transmission system in OWC using PWM and class D audio amplifier have been used. Experiments have been performed using 9 W white LEDs (correlated color temperature 2900 K) as downlink maintaining the proper light level in the work plane and 850 nm IR LED as uplink for bidirectional communication. A detailed description of optical channel modeling is also described here which is very important for illumination point of view as well as communication point of view. For transmitting voice, 50 kHz carrier frequency is used, and successfully, a distance of 8 m for communication has been achieved.

Keywords Audio transmission · Class D amplifier · Optical wireless communication · PWM · Visible light communication

1 Introduction

The optical wireless communication (OWC) refers to a communication technology where optical sources are used as signal transmitter, air as a transmitting medium and photodetector as a signal receiver. When a visible light source (380–780 nm) is used as signal transmitter, it is called visible light communication (VLC). In recent decades, the development of light-emitting diodes (LED) in visible spectrum replaces

B. Das (✉) · S. Mukherjee · S. Mazumdar
Department of Electrical Engineering, Jadavpur University, Kolkata, West Bengal, India
e-mail: basudebdas.1991@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_14

the incandescent bulbs and fluorescent lamps, since they have the characteristics of long lifetime, mercury free, color mixing, higher efficacy, good color rendering index (CRI) value and fast switching. Utilizing the property of fast switching of LED, VLC uses white LED as transmitter which can transmit data at very high data rate as well as fulfill the requirement of illumination in indoor.

The concept of optical wireless communication is not the new one. In 1880, Grahambell invented photophone, where sun light was used as optical beam to transmit telephone signal at 200 m [1–3]. Later in 1962, MIT Lincoln Labs built an experimental OWC link using low-power GaAs LED to transmit TV signal [2, 3]. In 1994, OWC link was demonstrated using 806 nm IR LED at 2.9 m distance [4]. However after invention of LED in visible spectrum, VLC link for voice transmission was reported in Hong Kong by Pang et al. [5]. From 1999, many work based on VLC was reported. The revolutionary invention was Horald Hass's pure Li-Fi in 2011, where the data rate was 100 Mbps [6]. OWC became one of the most favorable technologies in the world of communication for its very high data rate, safe to human health compared to RF and point-to-point secure communication.

In this article, a technology has been demonstrated to transmit the audio signal using white and IR LED to establish a bidirectional communication link utilizing PWM scheme and class D audio amplifier. The previous works reported in [5, 7–11, 13] describe only unidirectional communication mode. Their developed systems have the link range from 12 to 100 cm. A work reported in [12], which covered a distance of 8 m, but worked in unidirectional mode also. Compared to the mentioned works, this developed prototype is able to drive high power LEDs, so that larger distance can be achieved and it works in bidirectional mode. A detailed comparison study is shown in Table 1.

Table 1 Comparison of existing audio transmission system using VLC

Author (Year)	Modulation technique	Carrier frequency	Distance	Communication mode
Pang [5]	Frequency modulation	100 kHz ± 50 kHz	77 cm	Unidirectional
Son [7]	S/PDIF	192 kHz	100 cm	Unidirectional
Jia [8]	multi level pulse coded modulation	Bit clock set to 256 kHz	12 cm	Unidirectional
Dong [9]	Voltage to frequency conversion	–	–	Unidirectional
Assabin [10]	Pulse width modulation	87 kHz	–	Unidirectional
Soh [11]	On–OFF keying using pyaudio	Datarate 4Mbps	15 cm	Unidirectional
Taufik [12]	Pulse width modulation	50 kHz	8 m	Unidirectional
Kodama [13]	Pulse width modulation	49.6 kHz	–	Unidirectional
This prototype	Pulse width modulation with class D amplifier	50 kHz	8 m	Bidirectional

2 Standardization of Visible Light Communication

IEEE 802.15.7 standard supports the modulation scheme for data rate from 11.67 Kbps to 96 Mbps for visible light (380–780 nm) communication. It also provides the dimming adoptable mechanisms for flicker-free high data rate visible light communication [14–16].

3 Optical Wireless Communication Channel Modeling

In visible light communication system, the channel can be classified in four different types depending upon the position of the transmitter and the receiver. The basic link types include directed (focused) line-of-sight (LOS), non-directed line-of-sight, directed non-line-of-sight and non-directed non-line-of-sight. Directed (focused) and non-directed (non-focused) channel depend on the direction of transmitter and the receiver, and LOS and non-LOS channel depend on whether there exists a barrier to the block the transmission of light between transmitter and receiver.

As general illumination operates on LOS environment and it is not focused or directed, so a non-directed LOS channel is important for VLC [10, 17–20]. Geometry for an indoor, non directed LOS VLC link is given in Fig. 1.

So, received optical power P_r at the receiver is expressed by

$$P_r = H_0 * P_t \tag{1}$$

H_0 represent the DC gain of the channel, and P_t is the transmitted optical power. H_0 is given by,

$$\begin{aligned} H_0 &= (m+1)A/2\pi\, d^2 \cos^m(\varphi)\, T_s(\psi) g(\psi) \cos(\psi), \quad 0 < \psi < \psi_c \\ 0 \quad \psi &> \psi_c \end{aligned} \tag{2}$$

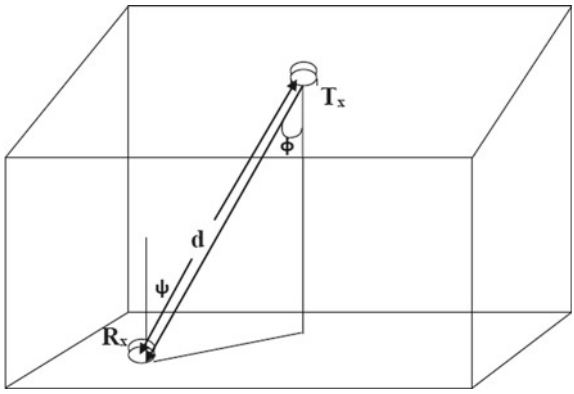


Fig. 1 Geometry of an indoor, non-directed LOS VLC link

A is the physical area of the detector in a photodetector. φ is the angle of irradiance with respect to the axis normal to the transmitter surface, ψ is the angle of incidence with respect to the axis normal to the receiver surface, d is the distance between LED and the detector's surface, $T_s(\psi)$ is the filter transmission, $g(\psi)$ is the concentrator gain, ψ_c is the concentrator field of view, i.e., semi-angle at half power, m is the order of Lambertian emission and is given with the transmitter semi-ngle at half power) $\varphi_{1/2}$ as

$$m = -\ln 2/\ln\left(\varphi_{1/2}\right) \tag{3}$$

where $m = 1$ in the case of $\varphi_{1/2} = 60°$ (Lambertian Surface).
The optical concentrator $g(\psi)$ can be given as [17]

$$\begin{aligned} &g(\psi) = n^2/\sin^2\psi_c \quad 0 < \psi < \psi_c \\ &0 \quad \psi > \psi_c \end{aligned} \tag{4}$$

n is the refractive index of the concentrator.

4 Experimental Study

A. *Block diagram of the system*
In this article, an audio transmission technique is reported for larger distance and noise-free audio transmission based on pulse width modulation (PWM) scheme and class D audio amplifier. The block diagram of the system is given in Fig. 2.

An amplified audio signal is fed to a PWM modulator. Then, the modulated PWM signal drives a MOSFET (as a switch) to switch on and switch off the high wattage LED. The distance of the communication link depends upon the wattage of the LED used. The photodetector is placed at a particular distance from the LED. It may be non-directed channel, but must be in line-of-sight. Output of the photodetector is a slightly distorted PWM signal whose amplitude is in mV. A comparator is used to regenerate the transmitted PWM signal which will drive an output stage MOSFET. A L-C low-pass filter is used to demodulate the PWM

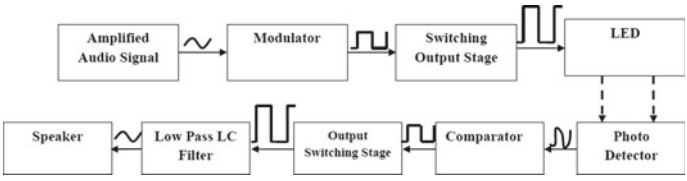


Fig. 2 Block diagram of the whole audio system

signal, and the demodulated audio signal is fed to the speaker. This scheme shows a less noisy optical wireless channel for audio transmission.

B. *Description of the transmitter operation*

The transmitter unit is built with an audio amplifier, a triangular wave generator, a comparator, a switching output stage and a LED. The goal of the transmitter unit is to switch on and off the LED according to the information (audio signal). LED will be on when the signal is in high state and off when the signal is in low state. Here, the PWM signal is generated by comparing the amplified audio signal and a triangular wave act as a carrier wave. The frequency of the PWM signal depends upon the frequency of the triangular wave. In this work, the frequency of operation is 50 kHz, which is the frequency of triangular wave.
A block diagram of the transmitter unit is given in Fig. 3. An audio amplifier is used to amplify the audio signal with a DC bias. LM 358 Op-Amp is used as audio amplifier. The triangular wave generator is made with ADA4625-1 Op-Amp. LM 393 Op-Amp is used as comparator, and to generate PWM signal, IRFZ44N MOSFET is used as a switch to switch on and off the LED. The MOSFET is being chosen depending upon the forward current of the LED. The developed transmitter circuit is shown in Fig. 4.

C. *Description of the receiver operation*
In receiver unit, the transmitted signal is received by a photodetector and the received signal is processed to the speaker. The audio processing is based on the class D audio amplifier. The class D amplifier is operated on PWM signal,

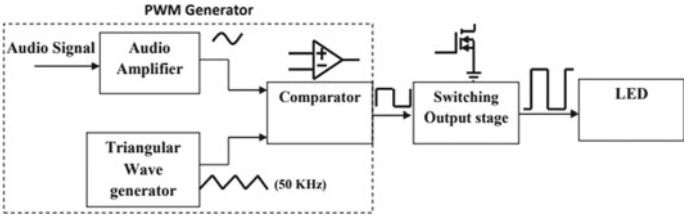


Fig. 3 Block diagram of the transmitter unit

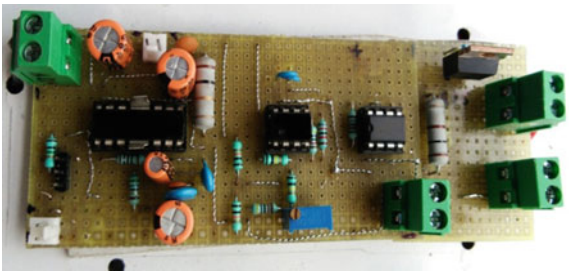


Fig. 4 Developed transmitter circuit

and it dissipates much less power compared to other audio amplifiers like class A, class B, class AB. [21–24]. For these reasons, the class D amplifier has been chosen.

Figure 5 shows the block diagram of the receiver unit. Photodetector converts the received optical signal into current, and using a resistor, it is converted to voltage form. Very low amplitude (mV) PWM signal is received which is fed to a comparator (LM 393) to get transmitted PWM signal. The PWM signal is the input of class D amplifier. A digital audio driver (IRS 20957S) is used for half-bridge operation. A dual power supply is required to drive the driver. Two N-Channel MOSFETs (IRFZ44N) are used for high-side and low-side drive of half bridge. The MOSFETs have to be chosen depending upon the output speaker power. The amplified audio passes through a low-pass L-C filter to demodulate the PWM signal and get the actual audio signal. Single-ended Butterworth low-pass L-C filter is chosen for this purpose. The value of the inductor and capacitor can be calculated if the speaker's resistance and cutoff frequency of the filter is known. The developed receiver circuit is shown in Fig. 6.

D. *Single-ended L-C filter design*

Figure 7 shows the single-ended L-C filter configuration. The equation of the single-ended L-C filter is shown.

$$f_0 = W_0/2\pi = 1/(2\pi\sqrt{LC}) \tag{5}$$

f_0 is the cutoff frequency of single-ended L-C filter.

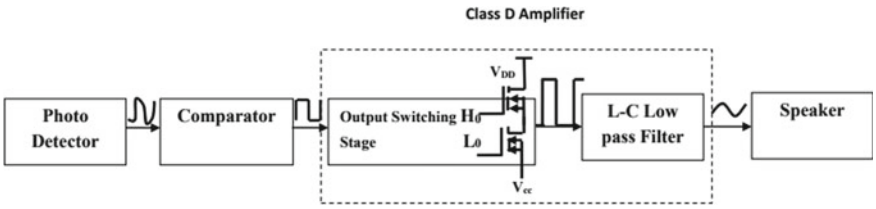


Fig. 5 Block diagram of receiver unit

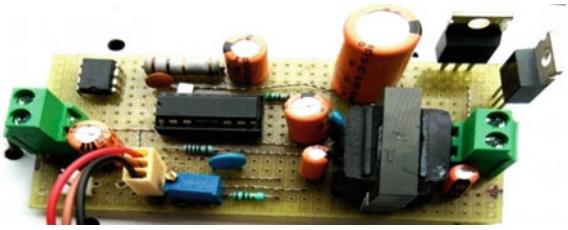


Fig. 6 Developed receiver circuit

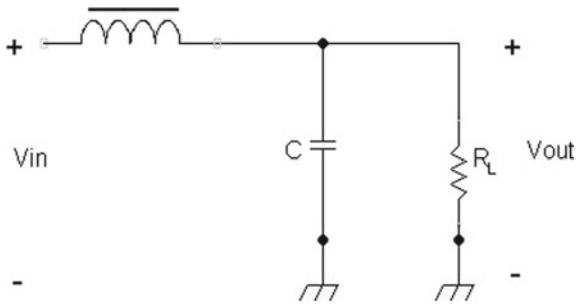


Fig. 7 The single-ended L-C filter configuration

$$\text{Quality factor } Q = R_L \sqrt{(C/L)} \tag{6}$$

$$\text{Damping Ratio } \zeta = 1/2Q = 1/2 * R_L * \sqrt{C/L} \tag{7}$$

It is usually desirable to design a filter that is critically damped with a Butterworth response [25]. For this type of filter, $Q = 1/\sqrt{2}$.
L-C value can be derived for critically damped system, assuming the cutoff frequency is known

$$L = R_L * \sqrt{2}/W_0 \tag{8}$$

$$C = 1/(W_0 * R_L * \sqrt{2}) \tag{9}$$

In this work, for 4 Ω speaker and 20 kHz cutoff frequency is considered. For that $L = 45$ μH and $C = 1.5$ μF are calculated.
A Bidirectional communication scheme

A bidirectional communication system is implemented here. The main purpose of indoor optical communication is to provide sufficient illuminance on the working plane as well as transmission of data and voice through light.

In Fig. 8, TRU1 represents the transceiver unit 1 which comprises of warm white LED, whose CCT is around 2900 k and BPW24R photodetector. TRU 2 represents the transceiver unit 2 which comprises of IR LED, whose wavelength is 850 nm and 1KL3B photodetector. The white LED (2900 k) is used not only for voice transmission, this also is used as illuminant unit whose light falls on the working plane and the voice signal is demodulated through 1KL3B. In case of transceiver unit 2, the low wattage IR LED of wavelength 850 nm is used. Due to invisible IR LED (850 nm) wavelength, this is effectively used as an uplink (no up light is required to illuminate the ceiling here) of the system. This is also carrying the voice signal which is detected by BPW 24R photodetector. So, in this way, a bidirectional optical wireless audio transmission channel is established.

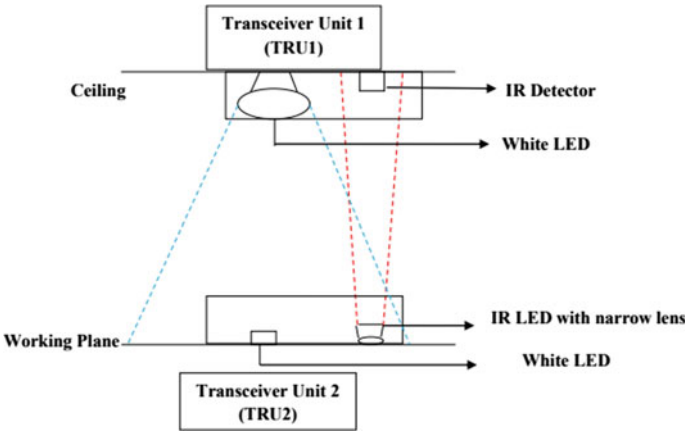


Fig. 8 Bidirectional audio transmission scheme using wireless optical channel

5 Experimental Result

The experiment has been performed at Illumination Engineering Laboratory, Electrical Engineering Dept., Jadavpur University, India. Audio signal is generated and processed through the transmitter circuit. The received signal is processed, and finally, the transmitted audio is audible on the 4 Ω, 3 W speaker. The received sound is noiseless and clear. Figure 9 shows the experimental setup of the system

To clarify the reliability of the optical wireless channel, a 50 kHz square wave pulse, generated by Keysight 33,600A waveform generator is transmitted via LED, and at the receiving end, the waveform is analyzed by the Keysight MSOS054A mixed-signal oscilloscope. The eye diagram of the received 50 kHz square wave signal is shown in Fig. 10 which shows the communication channel is noise less and reliable.

Figure 11 shows the experimental result of the transmitter side. Yellow line shows a 50 kHz triangular signal, which is the reference signal of the comparator. Green line is the amplified audio signal coming from the mobile phone, and it is the input signal of the comparator. Blue signal is corresponding PWM signal which is fed to the gate of switching MOSFET. The horizontal scale is 10 μs/div, and vertical scale is 5 V/div.

The experimental result of receiver side is given in Fig. 12. Yellow signal shows the received signal of the photodetector end. It clearly shows a very low amplitude PWM signal at 50 kHz. By using a comparator, the amplitude of that signal is shifted to 5 V and the frequency remains same which is the blue signal in Fig. 12. That 5 V PWM signal is the input of class D amplifier. Horizontal scale is 10 μs/div, and vertical scale is 2 V/div.

The purpose of this system is not only transmitting the audio signal, but also to maintain the light level on the working plane. As audio is transferred through light

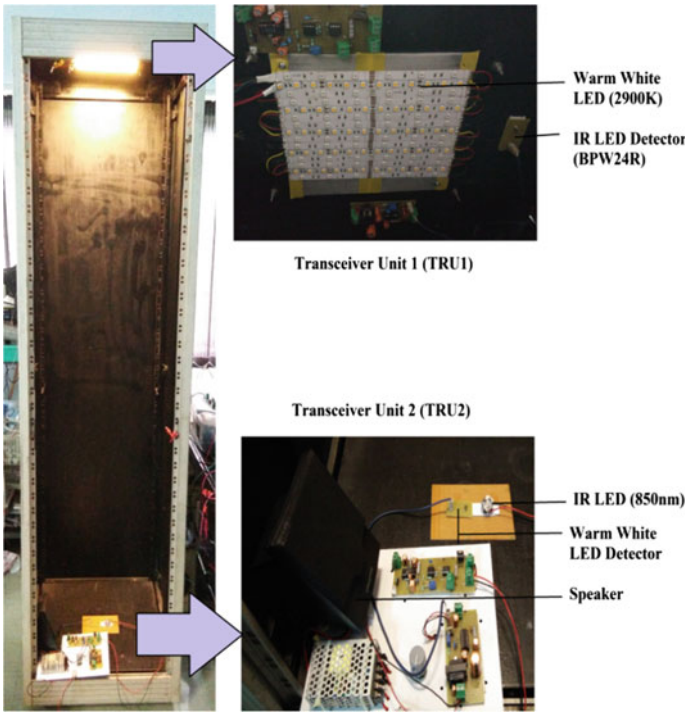


Fig. 9 Experimental setup

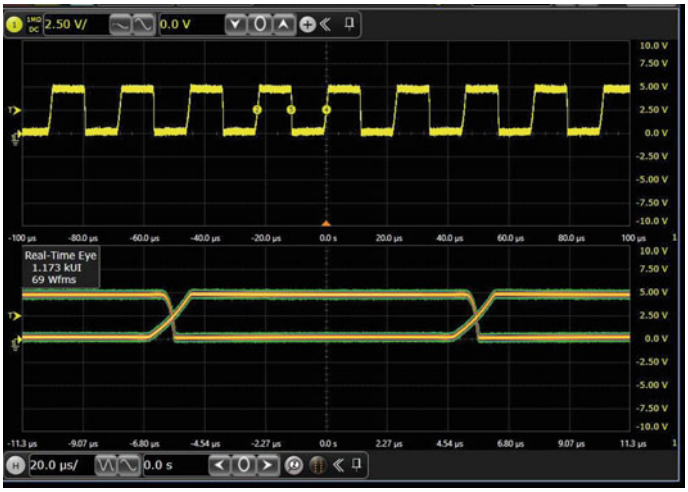


Fig. 10 Received 50 kHz signal and its corresponding eye diagram

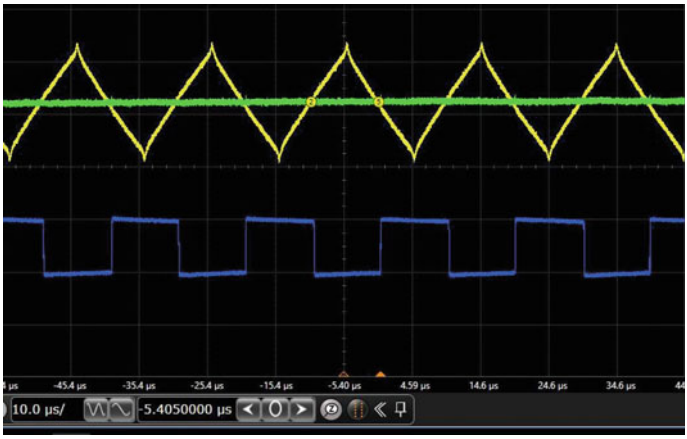


Fig. 11 Experimental data of transmitter side

Fig. 12 Experimental data of receiver side

by pulse width modulation technique, the average current of the LED becomes a variable parameter which depends on the pulse width of the signal.

$$I_{\text{avg}} = (T_{\text{On}}/T_{\text{total}}) * I_{\text{peak}} \tag{10}$$

where I_{avg} is the average current of the LED Luminaire, I_{peak} is the peak current of the LED luminaire, T_{On} is the On time of the PWM signal, T_{total} is the total time period of the PWM signal. Total time period of PWM signal and peak current of LED is constant. So, average current of the LED becomes function of On time of PWM signal.

$$I_{\text{avg}} \infty f(T_{\text{On}}) \tag{11}$$

Illuminance is a function of average current of the LED

$$E_v \propto f\left(I_{\text{avg}}\right) \tag{12}$$

From Eqs. (11) and (12), it can be stated that illuminance on working plane is a function of on time of the PWM signal,

$$E_v \propto f\left(T_{\text{On}}\right) \tag{13}$$

The on time of the PWM signal depends on the volume and frequency of the audio signal. The experimental results of audio input sound level verses illuminance on working plane is shown in Fig. 13 and audio frequency verses illuminance on working plane (in Lux) is shown in Fig. 14.

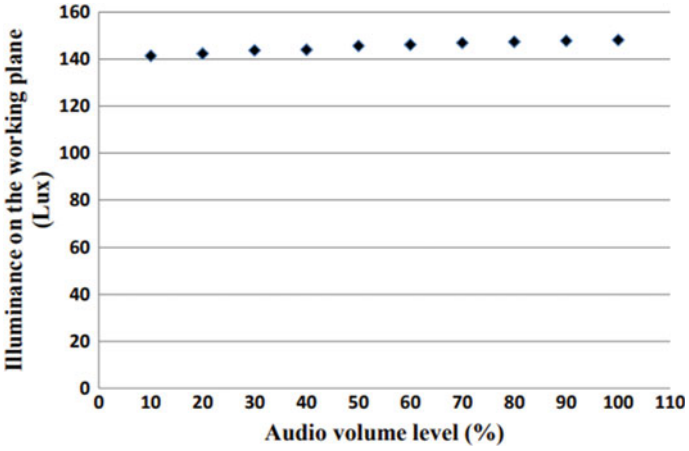


Fig. 13 Audio input sound level (in %) versus illuminance on working plane (in Lux)

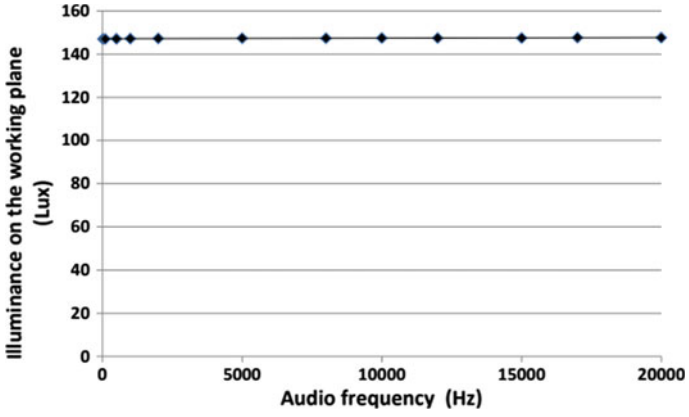


Fig. 14 Audio frequency versus illuminance on working plane (in Lux)

From Fig. 13 it can be stated that for increase in volume level from 10 to 100%, the illuminance level on working plane changes from 141 to 148 lx; only 4% variation is noticed. For variation in frequency of the audio signal from 20 Hz to 20 kHz, the illuminance remains almost constant, as shown in Fig. 14, a change of only ±1 lx is noticed which is around 0.7% of total light level. This slight change in light level does not occur any disturbance or flicker to the users.

6 Application

This system can be used as an optical intercom system in many offices or buildings. Nowadays, the use of LED has became very popular in office, school, colleges, multistoried buildings, housing complex, etc., where the optical intercom system can be implemented very easily with proper networking channel. A prototype of the optical intercom system is shown in Fig. 15.

Two rooms are considered; TRU 11 and TRU 12 are the transceiver modules of room1, and TRU 21 and TRU 22 are the transceiver modules of room 2. TRU 11 and TRU 21 are mounted on ceiling which comprises the white LED and IR photodetector, and TRU 12 and TRU 22 are placed on working plane, which comprises of IR LED and white LED detector. IR photodetector in TRU 11 is connected to the white LED of TRU21 through wire, and IR photodetector of TRU21 is connected to white LED of TRU 11 through wire. Someone sitting in room 1 will speak on microphone which is connected to the IR LED of TRU12. The audio is received by the IR detector of TRU11; accordingly, the white LED of TRU21 will switch on and off and transmit the audio to TRU22. White LED detector of TRU22 receives the audio signal, and it is processed to the speaker which is in room 2, and the overall process is vice versa for room 2 to room 1. In this way, someone sitting in room 1 can talk to another person sitting in room 2 in bidirectional mode. In this way, an optical intercom system may work.

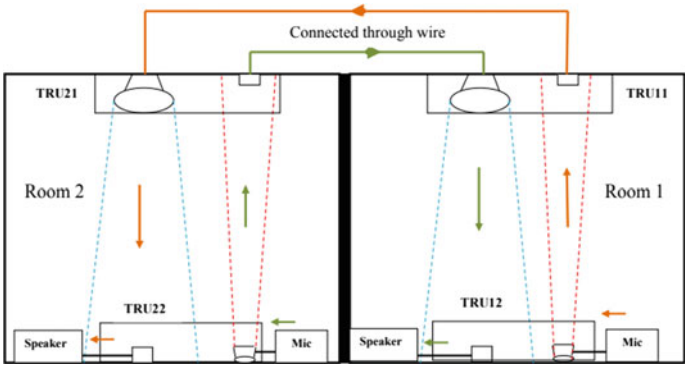


Fig. 15 Optical Intercom System

7 Conclusion and Future Scope

In this paper, an optical wireless communication channel for audio transmission is discussed with necessary circuit description. Both PWM modulation technique and class D amplifier are used in this system, which shows a higher distance and less noisy communication system. The average illuminance level 148 Lux is maintained in the working plane. If the wattage or efficacy of the LED is increased, illuminance on the working plane will also increase. The fundamental experiments are performed and reported in this article. This channel may be used for data transmission and video transmission; for this, a higher channel bandwidth is required. It can be done by modifying the transmitting and receiving hardware with proper communication protocol.

Acknowledgements The authors want to acknowledge Illumination Engineering Laboratory, Electrical Engineering Department, Jadavpur University for the infrastructural facility to carry out this research work.

References

1. Dennis K (2002) Free space optics for laser communication through air. Opt Photon news 13(10):36–42
2. Frank GE (1970) A review of operational laser communication systems. Proc IEEE 58(10)
3. Ali KM, Murat U (2014) Survey on free space optical communication: a communication theory perspective. IEEE Commun Surv Tutor 16(4)
4. Gene MW, Joseph KM (1994) 50 Mbps diffuse infrared free space link using on off keying with decision feedback equalization. IEEE Photon Technol Lett 6(10)
5. Grantham P, Ka-Lim H, Thomas K, Edward Y (1999) Visible light communication for audio systems. IEEE Trans Consumer Electron 45(4)
6. Chen H et al (2016) Advances and prospects in visible light communication. J Semicond 37(1)
7. Do SK et al (2013) Simultaneous transmission of audio and video signals using visible light communications. Eurasip J Wirel Commun Netw Springer Open J
8. Ziyan J et al (2013) A digital voice transmission system based on visible light communication. Proc IC-BNMT
9. Shiliang D et al (2016) A voice transmission system based on visible light communication. In: IEEE 5th global conference on consumer electronics
10. Abdelmoujoud A, Jamal E, Ahmed H, Abdellatif A Application of Li-Fi technology in the transmission of the sound at the base of the PWM. In: 2nd international conference on electrical and information technologies ICEIT'2016
11. Mei SY et al Real time audio transmission using visible light communication. In: Proceedings of TENCON, IEEE region 10 conference, Oct 2018
12. Jawad TBT et al (2009) Development of an audio transmission system through an indoor visible light communication link. Int J Sci Res Publ 9(1)
13. Motoi K, Shinichiro H (2019) Pulse width modulation visible light communication using digital micro-mirror device projector for voice information guidance system. In: IEEE 9th annual computing and communication workshop and conference (CCWC). https://doi.org/10.1109/CCWC.2019.8666577
14. IEEE 802.15.7 VLC (2012) Modulation scheme and dimming support. IEEE communication magazine

15. Deva BK et al A review of communication-oriented optical wireless systems. Wirel Commun Netw Spr Open J
16. IEEE 802.15 Working Group for WPAN. https://www.ieee802.org/15/
17. Lee CG (2011) Visible light communication, advanced trends in wireless communications, Khatib M (ed). ISBN:978-953-307-183-1
18. Toshihiko K, Masao N (2004) Fundamental analysis for visible-light communication system using LED lights. IEEE Trans Consumer Electron 50(1)
19. Barry JR (1994) Wireless infrared communications. Kluwer Academic press, Boston, MA
20. Gfeller FR, Bapst U (1979) Wireless in-house data communication via diffuse infrared radiation. Proc IEEE 67(11):1474–1486
21. Eric G (2006) Class D audio amplifiers; What, Why, and How. Analog dialogue 40–06, June
22. Lianxi L, Yintang Y (2009) A high efficiency PWM CMOS class-D audio power amplifier. Art J Semicond
23. Zbigniew K (2007) Application of pulse modulation techniques for class D audio power amplifiers. Arch Acoust 32(03):683–706
24. Dondon PH et al (2011) Simple modeling and method for the design of a sigma delta class D power amplifier. Int J Circuits Syst Signal Process 4(5)
25. SLAA701A Application Report, Texas Instruments, October 2016-revised November 2016

Basudeb Das received the B.Tech degree in Electronic & Communication Engineering from West Bengal University of Technology, Kolkata, India in 2013 and M.E. degree in Illumination Engineering under Electrical Engineering Department, Jadavpur University, Kolkata, India in2017, where he is currently working towards the Ph.D degree in Electrical Engineering.

His research interests include Optical Wireless Communication, CCT changing Lighting System, Solar powered dynamic lighting system and electronic control of light.

Shibabrata Mukherjee received the B.Tech degree in Electrical Engineering from West Bengal University of Technology, Kolkata, India in 2013 and M.Tech degree in Illumination Tech. and Design from Jadavpur University, Kolkata, India in 2017, where he is currently working towards the Ph.D degree in Electrical Engineering.

His research interests include Non conventional energy sources, Optical communication and Environmental effect on transmission of semiconductor light sources in free space.

Saswati Mazumdar received the B.E.E., M.E., and Ph.D. degrees in electrical engineering (EE) from Jadavpur University (JU), Kolkata, India, in 1982, 1984, and 1996, respectively.

She worked first as CSIR Senior Research Fellow, then System Development Engineer in a DOE-funded Project in the Department of EE, JU. From 1987, she joined in teaching in the department of EE, JU. She acted as Director of School of Illumination Science, Engineering and Design, JU from 2006 to 2014. She is currently working as a Professor in the Department of EE, JU, Kolkata, India. She has now 32 years of experience in lighting research and teaching. She developed a modernized Illumination Engineering Laboratory in the Department of EE, JU. She has founded one full-time Master's Course on Illumination Engineering and another part-time Master's course on Illumination technology and design in JU. She has executed a large number of R&D and Consultancy Projects on illumination and allied fields. Her primary research areas are controllers of lighting systems, renewable energy-based lighting systems, smart lighting systems, LEDs and LASER communication, interior and exterior lighting design, and color control of modern lighting systems.

Forecasting of Gold Prices Using Bayesian Regularization Neural Network

Shruti Garg

Abstract Prediction of gold price has been done by many, the investors, artificial intelligence experts, machine learning experts and stock market agents, etc. Ten years gold price from January 2008 to May 2018 has been taken for prediction in this study using Bayesian neural network. Bayesian neural network is selected for prediction of price of gold because it is found suitable for time series data and does not depend on any historical feature. The forecasted price is also compared with actual price to find percentage error in predicted values. Finally, mean percentage error is calculated, which is found 1% here.

Keywords Bayesian neural network · Time series analysis · Gold price

1 Introduction

Bayesian neural network is a popular machine learning technique used for prediction of time series data. As gold prices vary according to time, the neural network is found suitable for given application. There are many different neural networks which were used for prediction in past [1]. Artificial neural network (ANN), radial basis function network (RBFN), recurrent neural network (RNN) and convolution neural network (CNN) have been used for prediction of NSE stock price in [2, 11]. Real-time stock market prediction has been done in [3]. Bayesian neural network (BNN) has been applied as predictor on different datasets in [4].

The article [5, 6, 8] and [18] implemented autoregressive integrated moving average (ARIMA) model for prediction of stock market price. The ARIMA model predicts the future values based on season and trend values present in time series data. Also, it does not use any dependent variable for prediction. Thus, this model does not give very accurate prediction.

S. Garg (✉)
Department of Computer Science, BIT, Mesra, Ranchi 835215, India
e-mail: gshruti@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_15

Prediction of gold price using artificial neural network has been done in [7], and results were compared with statistical method ARIMA model. Results of above two models have been compared based on r-square, root mean square error (RMSE) and mean absolute error (MAE). The neural network was found better in all three parameters.

The gold prices were also been predicted by logistic regression (LR) and artificial neural networks (ANN) in [9]. Both methods were applied on January 2005 to September 2016 dataset collected from various sources. Nineteen different attributes have been selected to apply the two machine learning models. Performance of ANN is found better than logistic regression model.

A naïve Bayes and ANN have been applied to predict gold price in [10]. A review on time series prediction in perspective of gold price has been presented in [11]. There are many statistical, machine learning, fuzzy, neural and genetic algorithm-based forecasting which has been from 2006–2016. Neural network and machine learning-based forecasting were found promising research area for future. As statistical forecasting methods have no learning capabilities, the demand of artificial intelligence-based predication system has increased in past few years.

Deep learning also gained popularity for prediction after 2016. In same direction, the gold prices were predicted by long short-time memory (LSTM), convolution neural network (CNN) and a combination of LSTM + CNN in [12]. The proposed models were compared by autoregressive regressive integrated moving average (ARIMA) and support vector machine(SVM). A comparison between the models LSTM + CNN and CNN + LSTM has been done. LSTM + CNN was found performing better in their work.

Many neuro-fuzzy and combination of neural and evolutionary computing systems were also proposed in the past. A BAT neural network (BAT-NN) was proposed in [13] for prediction of gold price. BAT evolutionary algorithm is metaheuristic search algorithm used for forecasting and optimization. BAT algorithm is based on ecological behavior of bats. Evolutionary algorithm along with neural network gives good accuracy in forecasting [14]. A comparison of BAT neural network has been done by artificial intelligence systems and traditional system. BAT-NN was found best among all.

A wavelet-based neural network (WNN) with ant colony optimization (ACO) was proposed in [15]. Wavelet neural network uses wavelet and sigmoid as activation function rather than only sigmoid in tradition neural network. Further, the architecture of WNN has been optimized by ACO. ACO is evolutionary optimization algorithm based on behavior of ants. The problem of overfitting has also been addressed by adjusting the parameters.

A hybrid of ARIMA and least square support vector machine (LSSVM) has been proposed for stock price prediction in [16]. A comparison of prediction ability of Levenberg–Marquardt neural network (LMNN) with Bayesian regularization network (BRNN) has been done in [17]. And Bayesian neural network was found better than LMNN.

Different experts did several statistical, artificial intelligence (AI)-based and machine learning (ML)-based modeling for prediction of verity of data. Also, there

are plenty of ML algorithms present in the literature. Neural networks were found superior then tradition statistical methods because they learn patterns from training data. The Bayesian neural network (BNN) predicts future values based on posterior probabilities. The neural network training is done based on optimization in standard network whereas BNN uses maximum likelihood estimator for weight updation.

In BNN, new weights and bias will be calculated in every execution based on following posterior probability described in [20]. Calculation of posterior probabilities helps in finding prior distribution in existing data. These prior distributions can be helpful in future analysis.

The subsequent section of thr present paper highlights functionality of Bayesian neural network, further objective of current research and experimental results which have been discussed, and lastly, the conclusion remarks have been given.

2 Bayesian Neural Network

Bayesian neural network (BNN) is a probabilistic variant multilayer perceptron (MLP) neural network that is widely used as machine learning algorithm. The architecture of BNN consists of three types of different layers. First is input layer to which input is provided, and last output layer is there which produces future values in test phases. There may be one or more hidden layers present in between input and output layer. The general architecture of BNN is shown in Fig. 1. Weights in BNN are probabilities rather than random weights which are present in ANN.

The exact learning in BNN is done same as backpropagation algorithm [21]. The weight updation procedure in BNN is given in [22].

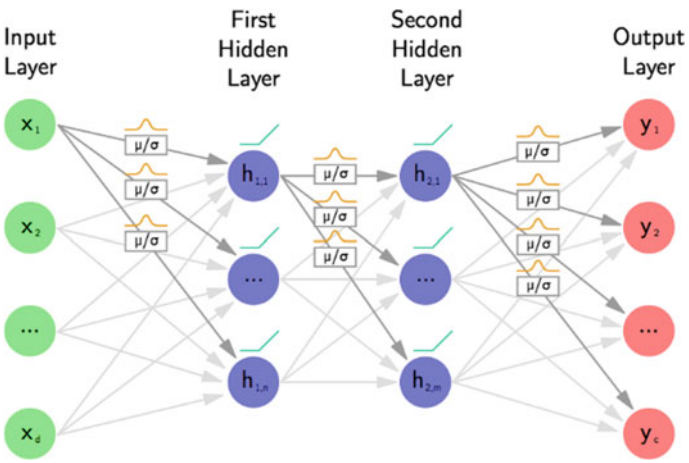


Fig. 1 Architecture of Bayesian neural network

The BNN also suffers with the problem of overfitting. The problem of overfitting can be reduced by applying regularization. The version of BNN with regularization is called Bayesian regularization neural network (BRNN).

3 Experimental Results

Research Objective: Apply Bayesian regularization neural network to predict gold price and verify its correctness using testing and validation.

Research Methodology and Approach: The history of 10 years of gold price has been taken for applying BNN. Total 2290 observations were taken from January 02, 2008 to May 16, 2018 from [19]. The gold price variation of last 5 years is shown in Fig. 2.

Empirical Validation: The gold price data has been presented to network in three sets: training, validation and testing. Gold price data used here is taken from 2008–2018. There are total 2290 observations found in which 80% were used for training, and 20% values were used for validation and testing. Then, actual prices were then compared with network output, and mean percentage error was calculated by using the formula

$$\text{Percentage error} = ((\text{Actual price} - \text{forecastedprice})/\text{Actual price}) \quad (1)$$

And finally, mean percentage error is calculated

$$\text{Mean percentage error} = \text{Percentage error}/\text{number of observations} \quad (2)$$

The total number of iterations required to train neural network is 111. After minimizing the error by 111 iteration, the mean percent error found is 0.010585. The number of hidden neuron taken in neural network is 10, and the number of delay taken is 2.

Fig. 2 Gold price variations of last 5 years

The output of BNN is shown in Fig. 3, and the error in network output and target output is shown in Fig. 4.

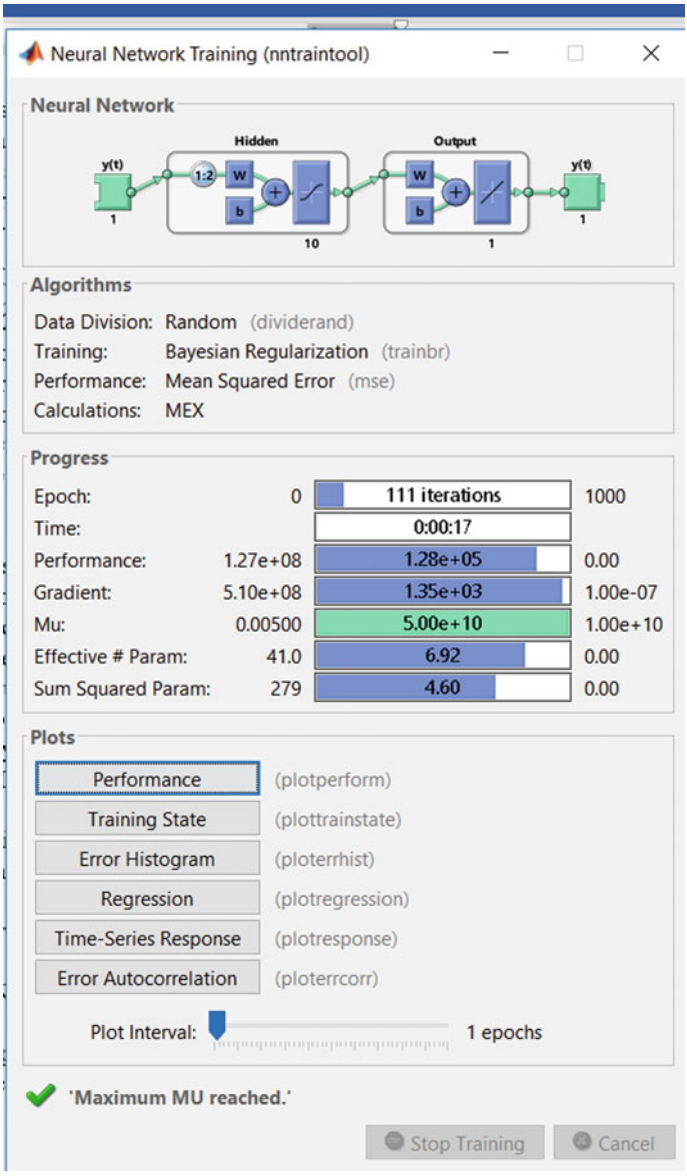


Fig. 3 Output of Bayesian neural network

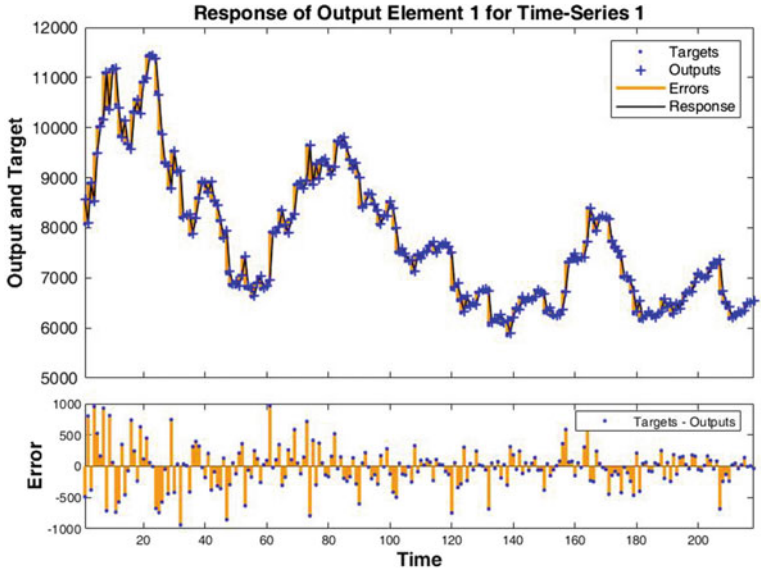


Fig. 4 Error in neural network output and target price

4 Conclusion

Many researchers in field of economics and computer science are working prediction of gold price. In the present study, BNN was used to predict the gold price, because BNN is giving very much satisfactory results in field of image processing, pattern recognition and classification, natural language processing and time series data prediction. The experimental results show that the accuracy of prediction BNN is very high, which is shown in result section, mean percentage error as 0.010585. Thus, BNN model is found suitable for prediction of gold price data.

References

1. Zhang G, Eddy Patuwo B, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. Int J Forecast 14(1):35–62
2. Hiransha M et al (2018) NSE stock market prediction using deep-learning models. Proc Comput Sci 132:1351–1362
3. Selvamuthu D, Kumar V, Mishra A (2019) Indian stock market prediction using artificial neural networks on tick data. Fin Innov 5(1):16
4. Mullachery V, Khera A, Husain A (2018) Bayesian neural networks. arXiv:1801.07710
5. Sengupta I, Chaudhury S (2018) Stock market prediction using time series analysis. https://www.papers.ssrn.com
6. Ariyo AA, Adewumi AO, Ayo CK (2014) Stock price prediction using the ARIMA model. In: 2014 UKSim-AMSS 16th international conference on computer modeling and simulation, pp 106–112

7. Mombeini H, Yazdani-Chamzini A (2015) Modeling gold price via artificial neural network. J Econ Bus Manage 3(7):699–703
8. Angadi MC, Kulkarni AP (2015) Time series data analysis for stock market prediction using data mining techniques with R. Int J Adv Res Comput Sci 6(6)
9. Osterrieder J, Chan S, Chu J, Nadarajah S (2017) A statistical analysis of cryptocurrencies, Ssrn, no, pp 1–23
10. Gandhi K, Shah N (2017) Gold price prediction using naïve bayes regression algorithm and ANN. J Open Source Devel 4(2):6–9
11. Goh WYY, Sufahani S (2018) Research review on time series forecasting of gold price movement. Int J Multidiscip Res Devel 5(5):44–49
12. He Z et al (2019) Gold price forecast based on LSTM-CNN model. In: 2019 IEEE intl conference on dependable, autonomic and secure computing, intl conference on pervasive intelligence and computing, intl conference on cloud and big data computing, intl conference on cyber science and technology congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE
13. Hafezi R, Akhavan A (2018) Forecasting gold price changes: application of an equipped artificial neural network. AUT J Model Simul 50(1):71–82
14. Ravi V, Pradeepkumar D, Deb K (2017) Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. Swarm Evol Comput 36:136–149
15. Li B (2014) Research on WNN modeling for gold price forecasting based on improved artificial bee colony algorithm. Comput Intell Neurosci
16. Xiao C, Weili X, Jiang J (2020) Stock price forecast based on combined model of ARI-MA-LS-SVM. Neural Comput Appl 1–10
17. Priya A, Garg S (2020) A comparison of prediction capabilities of bayesian regularization and Levenberg–Marquardt training algorithms for cryptocurrencies. In: Smart intelligent computing and applications. Springer, Singapore, 657–664
18. Garg S (2018) Autoregressive integrated moving average model based prediction of bitcoin close price. In: 2018 international conference on smart systems and inventive technology (ICSSIT). IEEE
19. https://www.kaggle.com/altruistdelhite04/gold-price. Accessed 23 Feb 2019
20. Miocevic M, Levy R, van de Schoot R (2020) Introduction to bayesian statistics. In: Small sample size solutions (open access): a guide for applied researchers and practitioners, vol 1
21. Leung H, Haykin S (1991) The complex backpropagation algorithm. IEEE Trans Signal Process 39(9):2101–2104
22. Blundell C et al (2015) Weight uncertainty in neural networks. arXiv:1505.05424

A New Proposed TCP for Faster Transmission of Packets

Mayank Oraon, Ravi Shankar, Kishan Manjhi, Sabina Priyadarshini, and Ramesh Narayan

Abstract Connection-oriented protocols have a big overhead of sequence numbers and acknowledgment numbers which make the protocols slow. This chapter discusses a new TCP implementation that reduces this overhead in connection-oriented protocol-TCP and therefore makes the transmission faster. A new method of congestion control has also been suggested for better performance.

Keywords TCP · Faster transmission · TCP congestion control · Connection-oriented protocol

1 Introduction

1.1 TCP—An Overview

Transmission control protocol (TCP) is a protocol in transport layer [1–5]. Transport layer transports application layer messages between the client and server sides of an application. It provides logical communication between application processes and networks. TCP provides a link-oriented service to its applications. It also provides congestion control and flow control. A connection-oriented service means that there is a handshake done, and resources and buffers are allocated on both sides of server and client before data is sent and received by both of them. TCP does a three-way handshake. First, a Syn bit set to 1 is sent and has an initial sequence number called "client_isn". This segment is called Syn segment. Then, server extracts TCP SYN segment and allocates TCP buffers and variables to the connection. It sends a Syn bit set to 1, ack (acknowledgment) field set to "client_isn + 1" and server's own first

M. Oraon (✉) · R. Shankar · K. Manjhi · S. Priyadarshini · R. Narayan
Computer Science Engineering, Ranchi 835215, India
e-mail: emayank.jr@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_16

sequence number "server_isn" (Synack segment). On receiving this acknowledgment, client also allocates buffers and variables to the connection. Client acknowledges server's connection granted segment by sending ack number "server_isn + 1". Syn bit is set to zero. Data can be sent along with this segment. After this three-way handshake, data is sent and received by both sides. TCP provides reliable data transfer using sequence numbers that uniquely identify a packet and acknowledgments that assure the sender that the packet has been received by the receiver [6].

TCP uses a combination of two protocols, namely Go-Back-N protocol and Selective repeat. In Go-Back-N protocol, the sequence numbers of the packets are in a window like 0 to 7 and after 7, the sequence number is again reset to 0 and goes till 7. It accepts in-order delivery of packets. That means, it accepts packets in correct sequence. If a packet arrives that is out-of-order, it is discarded. However, in selective repeat protocol, packet buffering is done for out-of-order packets. Individual acknowledgments are sent, so that all packets don't need to be retransmitted in case one packet is lost or received with errors [7].

1.2 TCP Segment Structure

TCP packets are called segments. Application layer messages are broken down into shorter segments by transport layer. The TCP structure is as shown in Fig. 1.

Sequence number is a 32-bit field used to identify a packet. After sequence number, there is ack field of 32 bits which is used by receiver to send acknowledgments for reliable data transfer. There is 4-bit header length followed by 6-bit flag fields. First

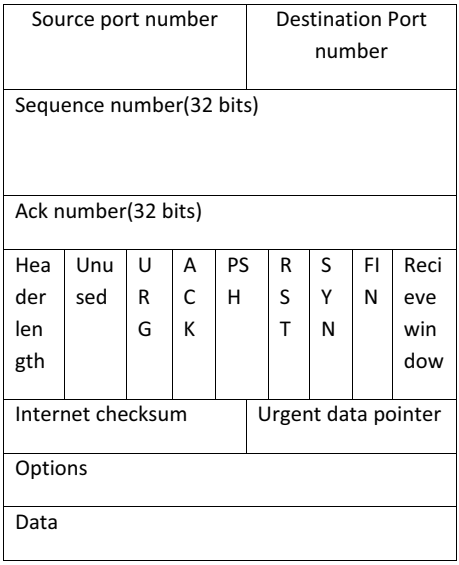


Fig. 1 TCP segment structure [1]

is URG meaning that the data is urgent. Then is ACK field which is set to 1 to indicate that ack field has valid data. RST, SYN, and FIN are used for connection set up and teardown. PSH bit indicates that receiver should pass data to upper layer immediately. Urgent data pointer is a 16-bit field to give address of last byte of urgent data. A 16-bit receive window field is used for flow control.

1.3 TCP Congestion Control

TCP congestion control is done in three ways [8–12]:

a. There is an Explicit Forward Congestion Indication (EFCI) bit present in each data cell. Router sets it to 1 to signal congestion to destination. There is a Resource Management (RM) cell that is sent by sender after every 32 data cells. It contains the EFCI bit. The destination sees the EFCI bit and sets EFCI of RM cell to 1 and sends it to sender. Thus, sender is notified of congestion.
b. Resource Management cells also contain a No Increase (NI) and Congestion Indication (CI) bit. Switches set CI to 1 to indicate severe congestion and NI bit to 1 to indicate mild congestion.
c. Resource Management cells also contain an Explicit Rate (ER) field which is a two-byte field. Congested switches can set a lower value for ER field in a passing RM cell.

The chapter is organized as follows. Section 2 contains a literature survey. Section 3 discusses the proposed TCP for faster transmission and lesser overhead. Section 4 discusses the advantage and Sect. 5 concludes the chapter.

2 Literature Survey

Advanced Research Projects Agency [13, 14] created the first network 'ARPANET' in 1969. In 1974, VG Cerf and RE Kahn developed the TCP/IP protocol. In 1980, University of Southern California came up with IP protocol. In 1985, NSFNET was developed by National Science Foundation. By 1986, there was a network architecture joining research and campuses organizations. Data was divided into small chunks of bits called datagrams. Concepts like fragmentation and reassembly came into existence. IP connected with local network protocols and took datagrams from source to next gateway or to destination. Service was provided with four parameters: Time to Live (TTL), Type of, Choices and header checksums. In early days, there were many local protocols. But in 1974, Cerf and Kahn came up with the concept of a single protocol—TCP—which was connection oriented, reliable, end-to-end protocol. TCP has communication between two sockets—sending socket and receiving socket. Socket has port number and IP address, respectively. Port numbers below 1024 are reserved. With TCP, multicasting and broadcasting are not allowed.

TCP uses a flow mechanism protocol called sliding window. This is a window scope field in TCP segment structure that requires the quantity of bytes that may be sent to receiver [15–17].

2.1 Future of TCP

TCP has limitations [18–20]:

- Wireless networks have higher bit error rates.
- TCP thinks that all packets go through the same path.
- TCP considers bandwidth as constant.
- Short-term sessions are damaging to obsolete TCP as TCP assumes all sessions have an exact number of time periods.
- Overhead of 40 bytes per packet is not efficient for larger payloads.

In TCP Tahoe [18–20], there is "slow start." At first, congestion window is one segment size. Each time an ack is received, it is increased by one. After time out, congestion window is doubled. Congestion window should be less than a threshold value says, T. When congestion window gets more than T, congestion avoidance phase begins where each RTT increments congestion window by Maximum Segment Size (MSS). Finally, when packet loss occurs, TCP Tahoe does fast retransmit. Fast retransmit means that, when Tahoe receives three duplicate acknowledgments, it resends the packet without waiting for time out. In this phase, there is maximum variation of window size.

In TCP Reno [21–24], Fast Recovery is an additional algorithm. Recovery starts from the second phase. Minimum window size in recovery process of TCP Tahoe can be overcome by starting recovery with second phase. Therefore, in TCP Reno, there is less variation in window size. In Fast Recovery, after receiving three acks, threshold T is set to half the current window; lost section is retransmitted; congestion window is fixed to congestion window +3; and, after receiving every ack, congestion window is increased by one.

Reno is efficient for single packet drop but not for multiple packet drops.

SACK TCP or TCP selective acknowledge [21, 25] contains a variable called pipe which shows number of outstanding packets. If number of packets in route is smaller than congestion window, then sender can send or retransmit packets. Pipe preserves acks from previous SACK options. Therefore, TCP SACK gets maximum window size as time increases when compared to TCP Tahoe and Reno.

In Vegas [25], an early and opportune choice to retransmit a dropped fragment is made. In TCP, there is a clock used to time full circle that ticks each 500 ms. Break check is additionally done when this "coarse grain" clock ticks. Thus, contrast between time when bundle is lost and time when it is detesting is any longer than would normally be appropriate. Along these lines, Reno thought of fast retransmit and fast recuperation instruments.

In TCP Vegas, when a copy ACK is obtained, it checks whether distinction between current time and when the fragment was sent (recorded as timestamp) is more noteworthy than break esteem. In the event that indeed, Vegas retransmits the section without hanging tight for three copy affirmations. At the point when a non-copy ack is obtained l in later retransmission, Vegas checks if time interim of the then sent section is bigger than break esteem. On the off chance that truly, Vegas retransmits the portion. This will get whatever new fragment that may have been misplaced past to the retransmission without sitting tight for a copy ack. The goal of this new strategy is to identify lost bundles despite the fact that there might be no second or third copy ack. This technique is exceptionally fruitful at accomplishing this objective as it added lessens the quantity of coarse grained breaks in Reno significantly. Additionally, Vegas possibly diminishes the clog window whenever retransmitted fragment was recently sent after the last abatement. Any misfortunes that occurred before the last window decline don't infer that the system is clogged for the present blockage window estimate, thus don't suggest that it ought to be diminished once more. This change is required on the grounds that Vegas recognizes misfortunes much sooner than Reno. Reno is responsive. In any case, Vegas is proactive.

In New Reno [26], during the moderate beginning stage, the sender builds the clog window win by 1/win with every affirmation got. In the wake of getting three copy affirmations, sender quickly retransmits. Sender at that point sets edge T to win/2; parts its success and actuates quick recuperation calculation. In quick recuperation, it retransmits on accepting one copy ack. After getting an incomplete ack, the sender retransmits the main unacknowledged parcel. At the point when an ack recognizes all bundles transmitted before initiation of quick retransmit, sender leaves quick recuperation and sets clog window to slow beginning edge T. On account of different bundles released from a solitary window of information, the primary novel data accessible to transmitter comes when sender gets an affirmation for retransmitted bundle. If there is a single packet drop, acknowledgment will acknowledge all packets transmitted. When there are multiple packet drops, the acknowledgment will acknowledge only some packets before fast retransmit. This packet is called a partial ack.

TCP Hybla [27] evaluates congestion window by removing the performance dependence on Round-Trip Time. Hybla has a larger average 'win' than TCP standard. TCP Hybla is used for satellite communication.

In TCP BIC [28], at the point when parcel misfortune happens, BIC decreases its window by a multiplicative feature (beta). The scope when the misfortune happened is set to max and dimension after decrease is fixed to min. There is a situation among min and max which is taken as smax. On the off chance that separation among min and midpoint among min and max is bigger than smax, smax is taken as the new window size. On the off chance that bundle misfortune happens, at that point smax is taken as new max else it is taken as new min. This goes on till window increase turns out to be not exactly a consistent smin. At the point when this occurs, window is set to current max. In the event that window develops past max, balance window dimension must be bigger than existing max and new max must be initiate. This is called max examining. The window at first develops gradually to locate the new

max; if not discovered, for example, parcel misfortunes happen, and at that point it changes to a quicker increment by changing to added substance increment where window size is increased by an enormous fixed augmentation.

BIC's evolution function can be also destructive for TCP. A new window growth function—cubic function—can be used. Origin is set as max, and after window decreasing, window grows very fast and as it reaches max, and it slows down its evolution. At max, its growth falls to zero. Then, window develops gradually accelerating its development as it transports away from wmax. Its development amount accelerates much more slowly than BIC's. It is TCP friendly because of slow growth.

DCTCP resolves TCP problem in the data center when it does not meet demand of apps. DCTCP provides high burst tolerance by combining large buffer headroom and destructive marking. It provides low latency by having small buffer occupancies. It also has high throughput as it employs FCN averaging by smooth rate adjustments and low variance. It comes with features like simplicity while converting to TCP and a single-switch parameter. It has its roots on mechanisms already accessible in silicon.

High bandwidth problem occurring in the TCP can be given a fix by implementation of high-speed TCP. High-speed TCP fine-tunes the added substance increment and multiplicative-decline parameter α and β so that for huge estimations of cwnd, the pace of increment of cwnd among misfortunes is significantly more quick and the cwnd decline at misfortune occasions is unimportant. This licenses organized utilization of all the accessible data transfer capacity for enormous transmission capacity postpone item comparably, when the size of cwnd is in run when TCP Reno works proficiently. High-speed TCP'S data is just reasonably faster than TCP Reno's so the two rivial moderately, unassumingly.

TCP Westwood represents the endeavor to utilize the RTF-observing techniques of TCP Vegas to determine the high-transfer speed issue, review that there is issue while separating among longest and non-congestive misfortunes. TCP westwood can likewise be considered as an improvement of TCP Reno's cwnd = cwnd/2 procedure, which is a more prominent drop than would normally be appropriate if the line limit at the bottleneck switch is not exactly the travel limit. It keeps on being a type of misfortune based blockage control.

Fast TCP is thoroughly connected to TCP Vegas, the plan is to retain the stable queue application attribute of TCP Vegas for as many cases as possible ad to also give competition to TCP Reno with its improved performance. We have observed that in TCP Vegas, RTT min at even lesser stable intervals (e.g., 20 ms) cwnd is amended via the resulting average:

$$\text{Cwnd new } = (1-\gamma) * \text{cwnd } + \gamma * ((\text{RTT no load/RTT}) * \text{ cwnd } + \alpha)$$

where γ is a constant in the middle of 0 and 1 ascertaining how "volatile" the cwnd inform is ($\gamma \simeq 1$ is the most volatile) and α is a fixed constant, number of packets the sender attempts to store in the restricted access queue.

Compound TCP adds a defer-based segment to TCP Reno. To this termination, CTCP enhances TCP Reno's cwnd by giving through a postpone-based component to the window dimension called dwnd, the absolute window dimension is then winsize = cwnd + dwnd (of course, winsize isn't acceptable to surpass the recipient's window dimension). Each RTT presently increments by 1/winsize.

HTCP shares similarity with high-speed TCP in the matter of growth of cwnd; has differences with high-speed TCP when it comes to cwnd incrementation, being determined by the slipped by time since the past misfortune occasion, and not by the size of cwnd. The limit for quickened cwnd development is normally fixed to be 1 s after the latest misfortune occasion. Utilizing RTTs of 50 ms, that signifies 20 RTTs, recommending that when CWNDmin is under 20 then HTCP carries on especially like TCP Reno.

TCP Cubic solves the issue of cost-productive TCP transport when bandwidth * delay is huge. TCP Cubic permits quick window extension; anyway, it likewise makes endeavors to hinder the flourishing of cwnd forcefully as cwnd is going to arrive at the current systems administration roof, and vouches for reasonable treatment of other TCP associations. Some portion of TCP Cubic's key breakdown uncovers that to accomplish, that is for hindering window development work as the last system roof is drawn closer and afterward find the fast increment, if this roof is outperformed without misfortunes. This inward then curved conduct mimics the diagram of the cubic polynomial cwnd = t^3, consequently the name.

TCP veno merges the ideas given by the congestion detection mechanism of TCP vegas into TCP reno to:

1. differentiate connection development in congestive or non-congestive state so that the window size can be altered soundly.
2. augment additive increase phase to better make use of available bandwidth and lower the happening of congestion loss. It is a refinement of reno's md as: It declares random loss when loss occur when not in congestive state otherwise it declares congestion loss. It is a refinement of reno's as: It tries to stay at the ideal transmission rate for as long as possible.

If the features of the network were known early, TCP would have been aimed in a better way which gives the idea of NATCP.

NATCP [19] utilizes an out-of-band criticism from the system to the servers found close by the input from the system, which incorporates the limit of cell get to connect and the base RTT of the system, manages the servers to alter their sending rates. NATCP replaces the customary TCP conspire at the sender. NATCP beats the best in class TCP conspires by in any event accomplishing 2× higher Power.

Binary Increase Congestion Control [20] is an congestion control algorithm which used for the transmission control protocol.

BIC optimized for the high-speed network with high latency generally known as long-fat-network. Advantage of BIC is correction of severely underutilized bandwidth.

BIC finds the maximum cwnd by searching.

1. binary search
2. additive inverse
3. slow start.

So when networks fails, the BIC uses multiplicative decrease in current cwnd.

3 Proposed TCP

TCP is a connection-oriented protocol. But because of being a connection-oriented protocol, TCP has a big overhead of sending 32-bit sequence number and 32-bit acknowledgment number in each packet. The proposed TCP packet structure aims at reducing this overhead and therefore makes the transmission faster.

According to the proposed TCP, for every 32 packets, an acknowledgment-sequence is sent. Before sending any packet, a 32-bit acknowledgment-sequence packet is sent. This field contains 32 bits. Each bit corresponds to one bit carried by the following 32 TCP segments. Each TCP packet sent thereafter has one bit acknowledgment field. For example, if the 32-bit acknowledgment-sequence sent before sending packets, 10, 111, 100, 001, 100, 100, 000, 111, 010, 101, 010, then, the first TCP packet sent thereafter has its acknowledgment bit set to 1, second packet has acknowledgment bit set to 0, third packet has acknowledgment bit set to 1, fourth packet has acknowledgment bit set to 1, and so on till 32 packets. The receiver extracts the acknowledgment bit from all the 32 packets when they are received and then compares it with the 32-bit acknowledgment-sequence sent before sending the 32 packets. If the bit patterns match, then it sends a positive acknowledgment to sender; otherwise, it sends the sequence numbers of packets not received. For example, if the third packet received has an acknowledgment field set to 0, then the third bit in the above-mentioned acknowledgment-sequence does not match with the acknowledgment bit received in the third packet. So, the receiver makes out that the third packet is either not delivered (lost) or not an in-order delivery. It checks the sequence number of the packet and sees that the packet is not the third packet. If the packet is an out of order delivery, then it is placed in its correct position in the 32-bit pattern; otherwise, the sequence number of the third packet is sent to sender to resend the packet. In this way, the overhead of 32-bit acknowledgment number in each packet is reduced to 1 bit in each packet.

The sequence number in the first packet is 32 bits in length. But the subsequent 32 packets are sent with only the last 16 bits of the sequence number incremented by 1 for each subsequent packet. There is an “attached bit” (1 bit in length) in each of the 32 packets set to 1 signaling that the packets are attached to the first packet, and they have a sequence number which consists of the first 16 bits of sequence number of the first packet and last 16 bits of the sequence number explicitly specified in the 16-bit sequence number field. After 32 packets, again, a 32-bit sequence number packet is sent and so on. So, all the attached packets have a 1-bit acknowledgment field, 1-bit

attached field and 16-bit sequence number field. The total overhead is 18 bits per packet instead of 64 bits per packet. The transmission is, therefore, faster.

The first 32 segments sent by sender and receiver looks in the Fig. 2.

The next 31 sender- and receiver-side segments sent have a structure as given in Fig. 3.

After 32 packets, again one segment is sent whose structure is as shown in Fig. 2.

For TCP congestion control, a Congestion Indication bit is sent from sender to each router on the way to destination. Each router sets the Congestion Indication bit to 1 if there is congestion and 0 otherwise and sends it back to sender. If a Congestion Indication bit is 1, the number of packets being sent is reduced by one-fourth of the total number of packets sent by the sender. If N is the total number of packets and M is the new number of packets, then

$$M = N - (1/4 * N) = N - N/4$$

If Congestion Indication bit is set to 0, then there is no curtailment in the number of packets being sent. For example, if there are three routers in the way from sender to receiver, and one router sends a 0 congestion indication bit and two routers send a 1 in congestion indication bit, then, the number of packets sent is reduced by 1/4th plus 1/4th of the total number of packets. If N is the total number of packets and M be the new number of packets, then

$$S = N - (1/4 * N) = N - N/4$$
$$M = S - (1/4 * S) = S - S/4$$

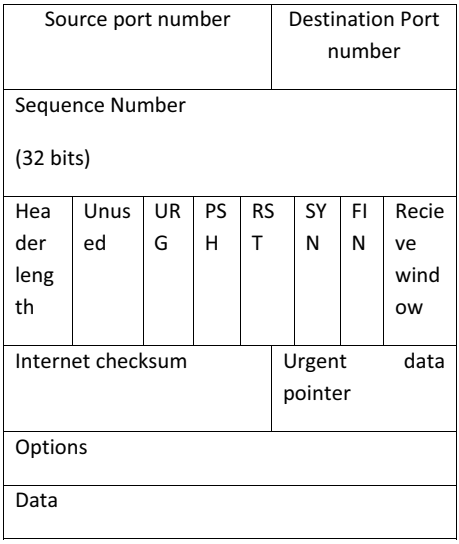


Fig. 2 TCP segment structure for first segment in every 32 packets

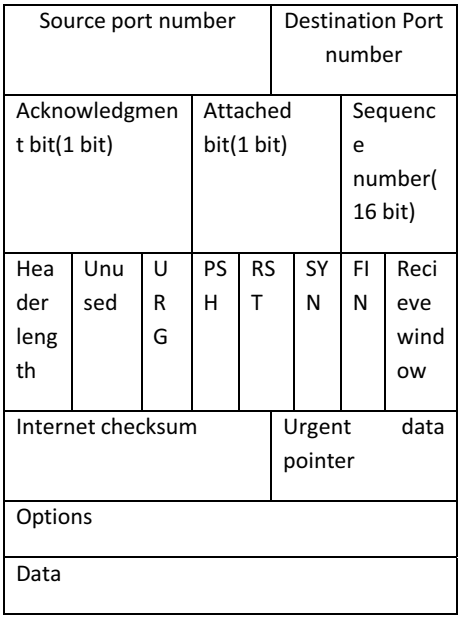


Fig. 3 TCP segment structure for rest 32 packets

where *S* is the intermediate total number of packets. In this way, congestion can be reduced. This value has been obtained empirically.

4 Advantage

The advantage of this proposed TCP is that the transmission would be faster as the overhead of each packet would be less. So, there is an advantage of both connection-oriented protocol and less overhead per packet.

5 Conclusion

Faster communication is a need these days. A new TCP has been proposed in this chapter which aims at reducing the length of the packets being sent. The number of bits in sequence number field and acknowledgment field together is reduced from 64 to 18 bits. This improves the speed of transmission, and therefore, the communication can be done faster. Also, congestion can be reduced by curtailing one-fourth of the total number of packets being sent for every router that says that there is congestion.

References

1. Floyd S (2001) A report on recent developments in TCP congestion control. IEEE Commun Mag 39(4):84–90
2. Afanasyev A, Tilley N, Reiher P, Kleinrock L (2010) Host-to-host congestion control for TCP. IEEE Commun Surv Tutor 12(3):304–342
3. Priyadarshi R, Soni SK, Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. Lect Notes Electric Eng. https://doi.org/10.1007/978-981-13-0776-8_27
4. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
5. Kumar A (1998) Comparative performance analysis of versions of TCP in a local network with a lossy link. IEEE/ACM Trans Netw 6(4):485–498
6. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
7. Rawat P, Chauhan S, Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circuits Syst Comput. https://doi.org/10.1142/S0218126620502047
8. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
9. Ho C-Y, Chen Y-C, Chan Y-C, Ho C-Y (2008) Fast retransmit and fast recovery schemes of transport protocols: a survey and taxonomy. Comput Netw 52(6):1308–1327
10. Randheer Soni SK, Kumar S, Priyadarshi R (2020) Energy-aware clustering in wireless sensor networks. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_39
11. Kumar S, Soni SK, Randheer, Priyadarshi R (2020) Performance analysis of novel energy aware routing in wireless sensor network. In: Lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_44
12. Priyadarshi R, Singh L, Randheer, Singh A (2018) A novel HEED protocol for wireless sensor networks. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474286
13. Jin C, Wei DX, Low SH (2004) FAST TCP: motivation, architecture, algorithms, performance. In: IEEE INFOCOM 2004, vol 4. IEEE, pp 2490–2501
14. Priyadarshi R, Singh L, Singh A, Thakur A (2018) SEEN: stable energy efficient network for wireless sensor network. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474228
15. Priyadarshi R, Yadav S, Bilyan D (2019) Performance and comparison analysis of MIEEP routing protocol over adapted LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung P-A, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 237–245. https://doi.org/10.1007/978-981-13-6295-8_20
16. Yu X, Qiao C, Liu Y (2004, March). TCP implementations and false time out detection in OBS networks. In IEEE INFOCOM 2004, vol 2. IEEE, pp 774–784
17. Priyadarshi R, Yadav S, Bilyan D (2019) Performance analysis of adapted selection based protocol over LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung P-A, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 247–256. https://doi.org/10.1007/978-981-13-6295-8_21
18. Priyadarshi R, Gupta B, Anurag A (2020a) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. J Supercomput. https://doi.org/10.1007/s11227-020-03166-5
19. Eggert L, Heidemann J, Touch J (2000) Effects of ensemble-TCP. ACM SIGCOMM Comput Commun Rev 30(1):15–29
20. Priyadarshi R, Gupta B, Anurag A (2020b) Wireless sensor networks deployment: a result oriented analysis. Wirel Pers Commun. https://doi.org/10.1007/s11277-020-07255-9

21. Feng W-C, Kandlur DD, Saha D, Shin KG (1999) Understanding and improving TCP performance over networks with minimum rate guarantees. IEEE/ACM Trans Netw 7(2):173–187
22. Priyadarshi R, Gupta B (2019) Coverage area enhancement in wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-019-04674-y
23. Bohacek S, Hespanha JP, Lee J, Lim C, Obraczka K (2006) A new TCP for persistent packet reordering. IEEE/ACM Trans Netw 14(2):369–382
24. Anurag A, Priyadarshi R, Goel A, Gupta B (2020) 2-D coverage optimization in WSN using a novel variant of particle swarm optimisation. In: 2020 7th international conference on signal processing and integrated networks, SPIN 2020. https://doi.org/10.1109/SPIN48934.2020.9070978
25. Priyadarshi R, Tripathi H, Bhardwaj A, Thakur A (2018) Performance metric analysis of modified LEACH routing protocol in wireless sensor network. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i1.5.9146
26. Feng W-C, Kandlur DD, Saha D, Shin KG (1999) Understanding and improving TCP performance over networks with minimum rate guarantees. IEEE/ACM Trans Netw 7(2):173–187
27. Priyadarshi R, Singh L, Sharma I, Kumar S (2018) Energy efficient leach routing in wireless sensor network. Int J Pure Appl Math 118(20):135–142
28. Priyadarshi R, Singh L, Kumar S, Sharma I (2018) A hexagonal network division approach for reducing energy hole issue in WSN. Eur J Pure Appl Math 118
29. Fall K, Floyd S (1996) Simulation-based comparisons of Tahoe, Reno and SACK TCP. ACM SIGCOMM Comput Commun Rev 26(3):5–21
30. Hagag S, El-Sayed A (2012) Enhanced TCP westwood congestion avoidance mechanism (TCP WestwoodNew). Int J Comput Appl 45(5):21–29

Design and Analysis of Koch Snowflake Geometry with Enclosing Ring Multiband Patch Antenna Covering S and L Band Applications

Nukala Srinivasa Rao

Abstract This chapter discusses the advancement in the performance of the Koch snowflake fractal antenna's behavior when it is enclosed in a circle. This design using the fractal geometry properties results in an antenna that resonates at multiple frequencies that are 2.141 GHz, 3.352 GHz, 4.260 GHz, 4.622 GHz, 5.161 GHz, 5.462 GHz and 6.910 GHz and the bandwidth at the resonant frequencies are 115.6 MHz, 60.2 MHz, 102.1 MHz, 139.2 MHz, 115.6 MHz, 102 MHz and 140.1 MHz, respectively. On comparing, this modified design of fractal antenna provides better results in S11 parameter than basic Koch snowflake design.

Keywords Component · Formatting · Style · Styling · Insert (keywords)

1 Introduction

In this day and age of wireless communication, ongoing improvements in wireless communication industry keep on deriving prerequisite of little, perfect and moderate microstrip patch antennas. Present-day telecom systems require antennas with more extensive frequency ranges and minute measurements than expected. It embarked research on antennas in various angles, one of them by utilizing fractal-shaped antenna elements. As an effort to enhance the technology of modern communications system, various approaches are being studied by the researchers to create novel antennas. The paper contains fractal geometry as well as left-handed, meta-material in order to achieve an antenna design suitable for several wireless applications.

The second-generation antenna is a microstrip, which is a metallic patch which uses a process. For example, lithography in which patterns are printed on this grounded dielectric substrate while fabricating the PCB's or IC's (Fig. 1).

It is the most successful and revolutionary design for antennas ever. The success of this antenna is from its advantage of having lightweight, low profile and low

N. S. Rao (✉)
ECE, Matusri Enginnering College, Hyderabad, Telangana, India
e-mail: srinivasa.nookala@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_17

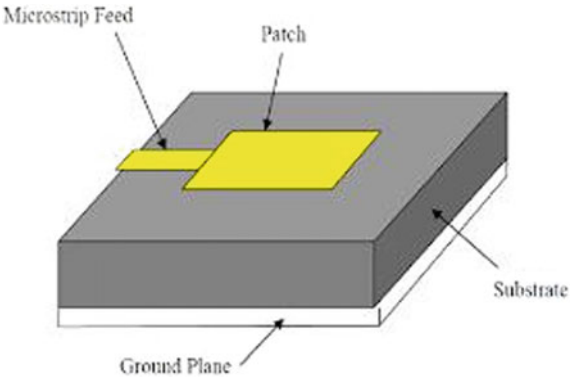


Fig. 1 Basic microstrip patch antenna

cost of fabrication, mechanical robustness, versatility and many more. However, it has some disadvantages also, the disadvantages include narrow bandwidth and low efficiencies.

Clouds, mountains, plant leaves and coastlines are the inspiration for fractal geometries [1–5]. It is essential to design antenna as compressed as achievable for some application. Fractal antenna has entered the view of many as a very promising solution. There are various types of microstrip antennas that are shown in Fig. 2.

Fractal antenna [2, 4] is the best suitable radiating structure. In modern technology, fractal antenna theory exists as a new area. Fractal geometry has a space-filling property which is self-similar [3]. These pattern looks complex but because of their self symmetry and having simple geometry. The basic geometry patterns are Koch and Hilbert snowflakes [6–8]. In order to provide feeding mechanism to an antenna, there are various techniques of feeding are used like microstrip line [9], coaxial probe [10], coplanar waveguide [11], etc. Each feeding mechanism is used for various applications.

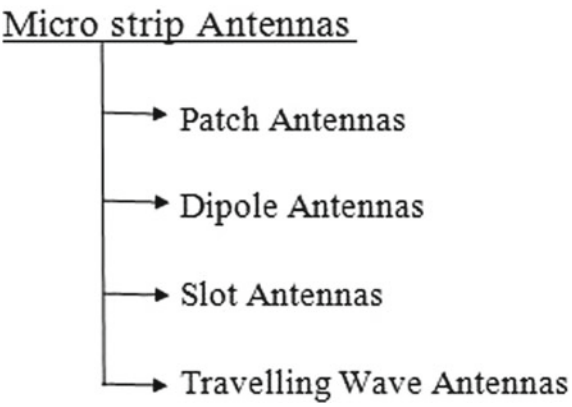


Fig. 2 Types of microstrip antennas

2 Fractal Geometry

There are various types of geometry used in fractal antenna that are discussed here.

Sirepinski gasket: it is a simple fractal that has fixed set with an overall shape of equilateral triangles which are iteratively sub-divided into small equilateral triangles (Fig 3).

Koch snowflake: The origin for Koch snow flake is an equilateral triangle. Then, partition the line segment into three parts of equilength. Next, draw the equilateral triangle by considering drawn line segments as its base and points to be outward. Afterward, remove the base of triangle made from first line segment. The Koch curve originally explained by Koch is constructed with only one side of original triangle. So three Koch curves make a single Koch snow flake [8, 14, 15].

Sirepinski carpet: In this a shape, it is sub-divided into small copies by removing one or more copies.

Hilbert curve: It is a continuous space filling curve which fills the square. The starting state is situated on the left. He designed a curve by connecting the center points of four sub-squares and again which leads to making up of a large square.

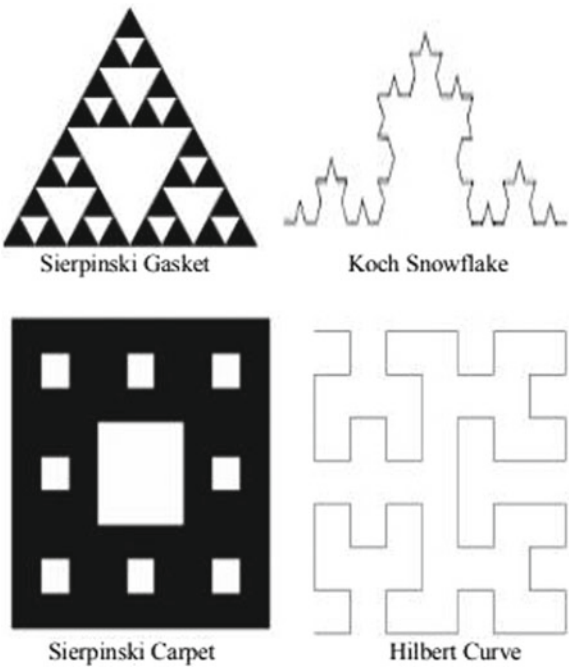


Fig. 3 Different kinds of geometry used in fractal antenna [12, 13]

3 Proposed Method

The fractal antenna either in communications or military and marine applications has been an emerging good research-oriented topic in the upcoming years. It is important for a single one device in order to access the additional services like GSM, WCDMA, WI-MAX, Wi-Fi, UMTS, 802.11a, Bluetooth, and 802.11b or satellite communication bands. For facilitate this, the antenna not only work on single frequency band but also it should work with multiple bands. Fractal antennas will work on multiband frequencies. Development of communication technologies, reduction of size and range of frequencies broadening are becoming very important design considerations for practical design applications of fractal antenna. The limitation of microstrip patch antenna such as less power handling capability, very low gain, and so fractal antenna may help to improve high gain and bandwidth of antenna. The ultimate objective of this paper is to design a compact patch antenna using novel fractal geometry and optimizing the proposed antenna to cover various bands by covering the satellite communication bands like L, S, C, X,KU, K, Ka, etc.

4 Implementation

Figure 4 shows in detailed dimension of fundamental equilateral triangle patch antenna with the overall size of the substrate are 60 mm $\times$ 60 mm. The one side of the equilateral triangle is $a = 33$ mm.

The design specifications are given in Table 1.

Second iteration of Koch snowflake geometry (Fig. 5).

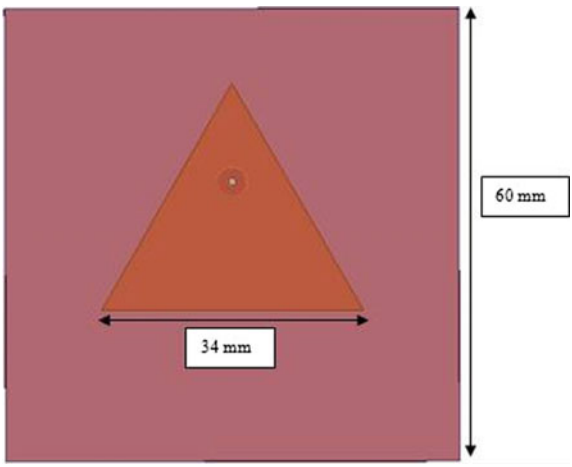


Fig. 4 Fundamental equilateral triangular patch designed by using HFSS

Table 1 Design specification of general patch antenna

Type of the substrate	FR4
Dielectric constant (ε_r)	4.40
Tangent loss (δ)	0.020
Thickness of the substrate (h)	1.60 mm

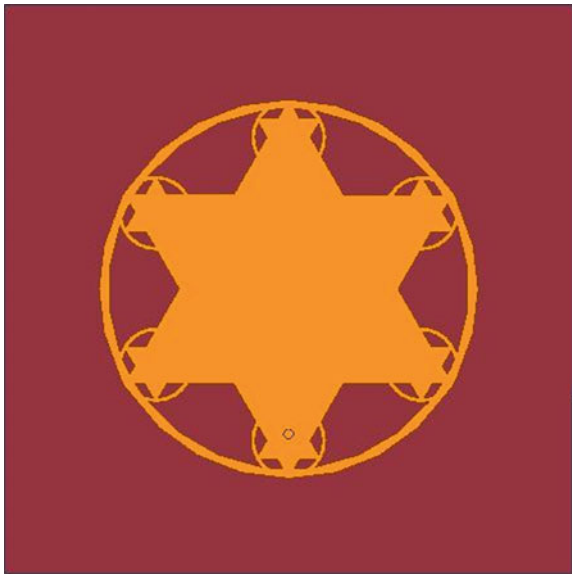


Fig. 5 Proposed factual antenna second iteration [16]

Gain:

The gain plot in Fig. 6 shows gain values at different bandwidths. The figures show the gain at resonant frequencies from 2.14 to 6.91 GHz. From the 3D diagram, we can observe the relation between the resonant frequency and the gain that is observed.

Radiation Pattern:

From 2D plot view of radiation pattern as shown in Fig. 7, it can be seen that at resonant frequencies radiation pattern obtained is omnidirectional.

Radiation Pattern:

From 2D plot view of radiation pattern as shown in Fig. 8, it can be seen that at resonant frequencies radiation pattern obtained is omnidirectional.

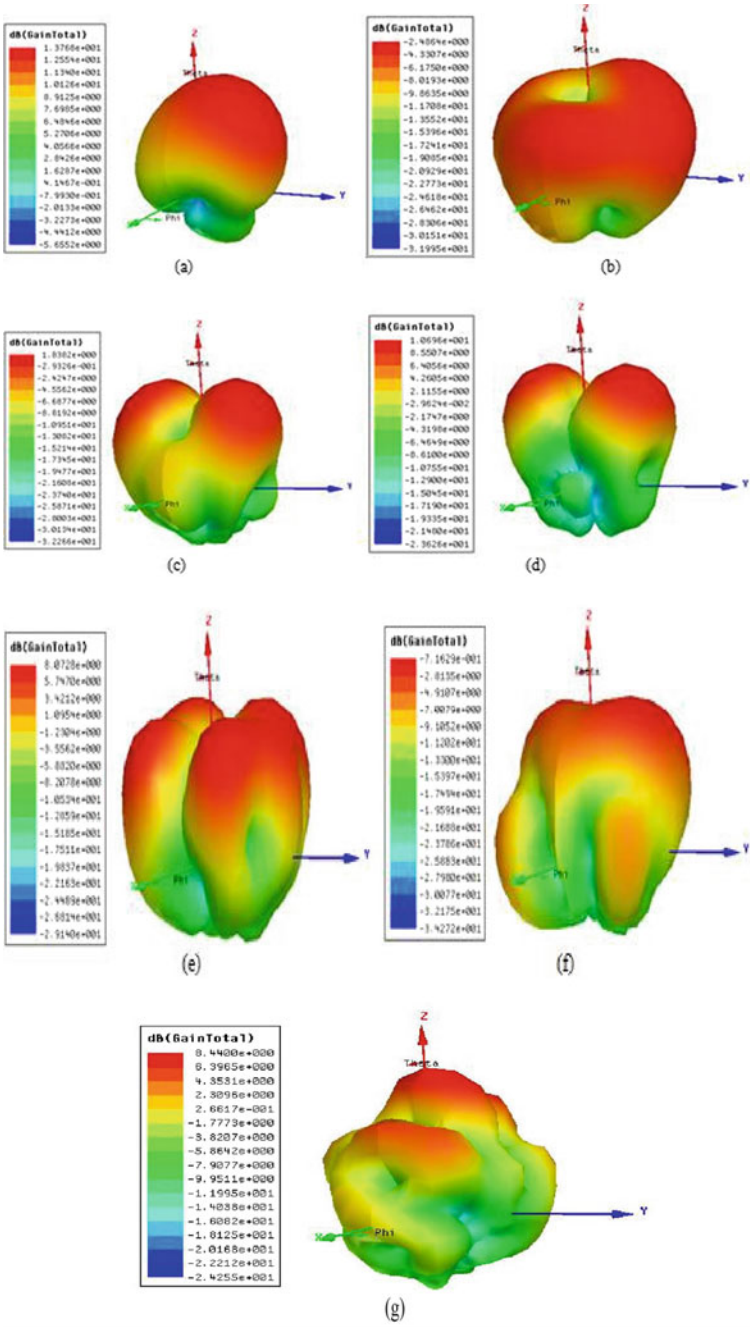


Fig. 6 Simulated gain of triangular patch antenna at **a** 2.141 GHz, **b** 3.350 GHz, **c** 4.261 GHz, **d** 4.621 GHz, **e** 5.160 GHz, **f** 5.461 GHz, **g** 6.91 GHz (3D view)

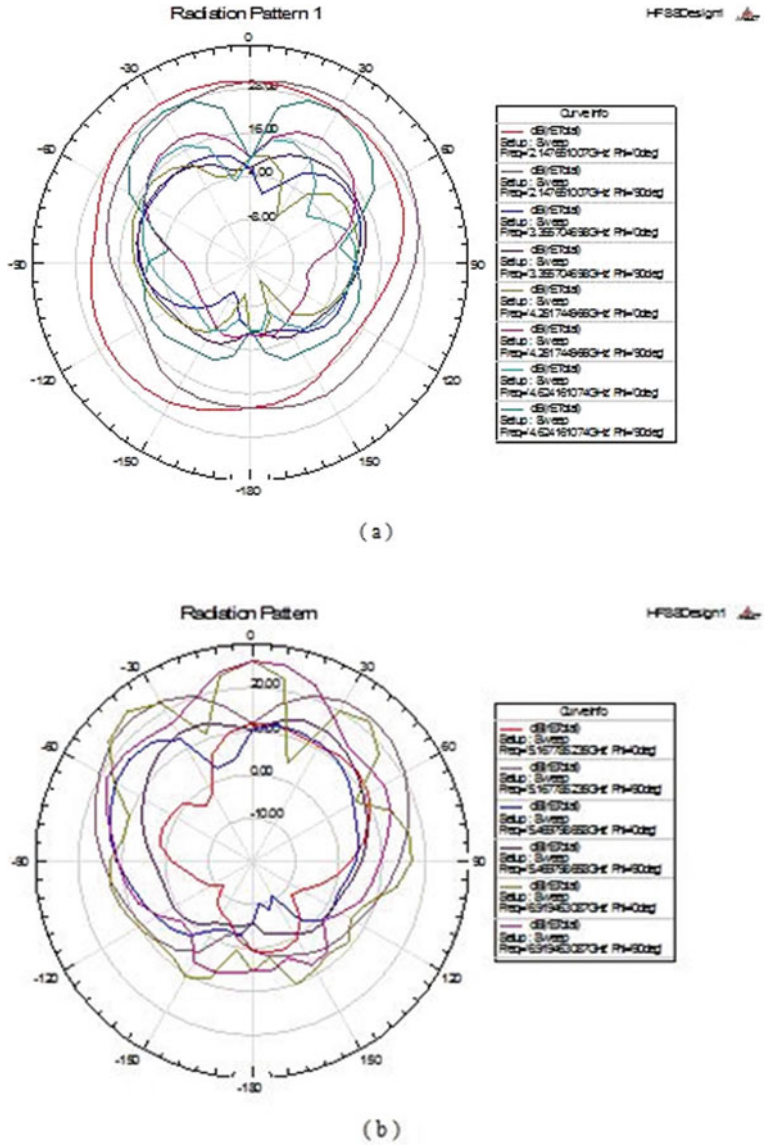


Fig. 7 Radiation pattern of the designed antenna at **a** 2.14 GHz, 3.35 GHz, 4.26 GHz, 4.62 GHz. **b** 5.16 GHz, 5.46 GHz, 6.91 GHz (2D view) for phi = 0°, 90°

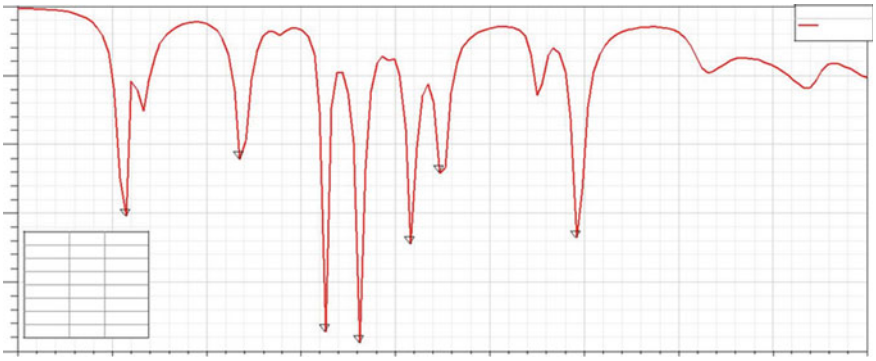


Fig. 8 Simulated return loss S11 of proposed fractal antenna (in dB)

Comparison of results.

References	Aim	Bandwidth (GHz)	VSWR and return loss
[17]	An improved compact and multiband fractal antenna	7.6% improvement at fL = 30.91 GHz and fH = 33.41 GHz	−29 to −21
[18]	A design of quad-band hexagonal antenna with fractal slots using inset feeding technique	750 MHz at 9.47 GHz frequency band	1.32 −12.99 to −20
[19]	Quad-band antenna for Wi-MAX applications	2–6 GHz	1.3 −13 to −18
Proposed	Design and analysis of Koch snowflake geometry with enclosing ring multiband patch antenna covering S and L band applications	2.16 GHz, 3.39 GHz, 4.26 GHz, 4.62 GHz, 5.20 GHz, 5.50 GHz 6.95 GHz, 9.20 GHz	1.3–1.4 Overall gain 7.95 dB 14.90 dB, −10.94 dB, −39.29 dB, −30 dB, −12.79 dB, −14 dB,−13.47 dB, −14.51 dB

5 Results

By applying first iteration using FR4 substrate material, antenna resonated at 4.440 GHz, 4.810 GHz, 6.550 GHz, 7.341 GHz and 8.240 GHz with the return loss achieved is −24.360 dB, −13.051 dB, −13.490 dB, −19.650 dB and −13.081 dB, respectively.

This antenna has a maximum gain of 9.201 dB with bandwidth of approximately 602 MHz. Analysis was done in terms of gain, bandwidth and return loss by applying coaxial feed. For application point of view, this antenna can be best use for WLAN, Wi-MAX, S band, C band, X band, and Ku band application. This antenna can also be useful for RADAR, satellite and military applications.

6 Conclusion

This chapter describes the multiband and fractal patch antenna by applying fractal geometry and ring structure. The proposed design was based on the snowflakes fractal design. The proposed geometry was simulated in HFSS using FR4 as the substrate. The zeroth and first iterations of the design were compared minutely in terms of return loss, bandwidth, VSWR and gain. The designs are showing multiband performance ranging between 1 and 10.1 GHz. There has been a good improvement in gain in successive iteration. In terms of size also, the patch area was reduced when compared to basic rectangular patch using FR4.

References

1. Werner DH, Haup RL, Werner PL (1999) Fractal antenna engineering, "the theory and design of fractal antenna arrays. IEEE Antennas Propag Mag 37–58
2. Lakshmana Kumar VN, Satyanarayanana M, Sridevi PV, Parakash MS (2014) Microstrip fractal linear array for multiband applications. In: IEEE international conference on advanced communication and computing technologies, pp 1818–1822
3. Sinha SN, Jain M (2007) A self-affine fractal multiband antenna. IEEE Antenna Wirel Propag Lett 6:110–112
4. Cohen NL (1995) Fractal antenna part 1. Commun Quart Summer 5–23
5. Baker JM, Iskander MF (2015) Electrically small fractal antennas. In: IEEE international symposium on antennas and propagation & USNC/URSI national radio science meeting, pp 1242–1243
6. Killi S, Kumar KTPS, Naresh Kumar D, Sahithi P (2016) Design and analysis of pentagonal patch antenna with H-tree fractal slots for S-band and WI-Max applications. Int J Eng Sci Comput 6(4):4537–4540
7. Singh J, Sharma N A design of quad-band hexagonal antenna with fractal slots using inset feeding technique. Int J Comput Appl 8:43–46
8. Kaur N, Sivia JS, Kaur M (2015) Design of modified Sierpinski Gasket fractal antenna for C and X-band applications. In: 2015 IEEE 3rd international conference on MOOCs, innovation and technology in education (MITE), pp 248–250
9. Jeemon BK, Shambavi K, Alex ZC A Multifractal Antenna for WLAN and WiMAX application. In: Proceedings of 2013 IEEE conference on information and communication technologies (ICT 2013)
10. Chauhan S, Deegwal JK, Gera R (2013) A design of triangular sierpinski gasket fractal antenn. IJECT 4(4)
11. Sundaram A, Maddela M, Ramadoss R (2007) Koch fractal folded-slotantenna characteristics. IEEE Antenna Wirel Propag Lett 6
12. Tiwari A, Rattan M, Gupta I (2014) Review on: fractal antenna design geometries and its applications. Int J Eng Comput Sci 3(9):8270–8275
13. Oraizi H, Hedayati S (2014) Microstrip multiband dipole antennas using the combination of sierpinski, Hilbert and Giuseppe peano fractals. In: Microwave symposium, 2014 14th mediterranean, pp 1–4
14. Kaur S, Sivia JS (2016) Analysis and design rectangular patch with half rectangular fractal techniques. In: Elsevier international conference on computational modeling and security, pp 386–392
15. Arun, Mittal A (2014) Multiband fractal microstrip patch antenna for wireless applications. Int J Adv Res Electron Commun Eng (IJARECE) 3(9)

16. More P, Khobragade SV A hexagonal-shaped fractal antenna array for multiband applications., emerging trends in computer science and information technology-2012(ETCSIT2012). Proceedings published in International Journal of Computer Applications® (IJCA)
17. Jena MR, Mangaraj BB, Pathak R (2014) An improved compact & multiband fractal antenna using the koch curve geometry. Wirel Mobile Technol 2(1):1–6
18. Singh J, Sharma N (2016) A design of quad-band hexagonal antenna with fractal slots using inset feeding technique. Int J Comput Appl (0975–8887) 150(8)
19. Krishna KS, Mohan N, Purushothaman K, Mohammed Abbas A (2018) Quad-band antenna for wi-max applications. Int J Pure Appl Math 119(15):785–791
20. Kaur N, Rani S (2016) Analysis and design of rectangular patch antenna with triangular fractal techniques. Int J Appl Sci Technol 6(2)

Phase Portrait Analysis of Continuous Stirred Tank Reactor (CSTR)

Ankita Ranjan and Sarbani Chakraborty

Abstract The behavioral analysis of continuous stirred tank reactor (CSTR) has been presented in this work. The mathematical model has been proposed using mass balance and energy balance equations. Linearization of CSTR has been carried out using Taylor series. The stable and unstable regions of CSTR have been depicted through the phase portrait. The prime objective of this work is to present the variation of concentration and temperature with change in initial conditions. For various initial conditions, the behavior of the system has been analyzed. It has been observed that for a certain range of temperature, the system is found to be unstable under any initial conditions. This finding is the novelty of the work. Simulations have been carried out in MATLAB.

Keywords Continuous stirred tank reactor (CSTR) · Phase portrait · Equilibrium point · Linearization

1 Introduction

CSTR is being used widely in process industries. Control of CSTR is a bit complex, as it is associated with multiple number of interacting parameters and variables. The behavior is highly sensitive toward small variations in the initial conditions [1]. This behavioral study of CSTR has been done to analyze the region of instability, thus locating the points where control needs to be applied. The concentration and temperature of the CSTR have been considered for the analysis purpose, thus giving prerequisites for control of CSTR. Phase portrait has been employed for examining the areas of stability at the equilibrium points. The stability of the system is of

A. Ranjan (✉) · S. Chakraborty
EEE Department, BIT, Mesra Ranchi 835215, India
e-mail: ankita1901@gmail.com

S. Chakraborty
e-mail: schakraborty@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_18

practical importance and has been established with the help of energy balance and mass balance equations [2].

The works related to the simulation and modeling of CSTR have been studied, and rigorous literature survey has been done. Ray [3] proposed a detailed study in order to control the error and define the trajectory of the CSTR.

Phase portrait analysis is useful for behavioral analysis of second-order nonlinear systems also. Analysis of CSTR has been done using phase portrait to check the instability in the dynamics. Analysis has also been carried out proposed by Berezowski [4] for cascade reactors consisting of 15 reactors in series comprising of self-induced oscillations. Further, Berezowski [5] proposed parametric continuation method to fabricate diagrams of steady states generated by tank reactor. The criteria for framing emergency starting regimes have been determined through parametric analysis in CSTR. With a proper amalgam of thermophysical and kinetic parameters, the system goes from transient state to steady state within acceptable range of dynamic overshoot, which has been proposed by Bykov et al. [6]. Phase equilibrium has been calculated using Redlich–Kwong–Soave equation of state. Mathematical modeling of two-phase flow in CSTR has been illustrated. Steady states have been analyzed under isothermal and adiabatic conditions in CSTR, for the hydrogenation of benzene, which has been proposed by Yermakova and Anikeev [7]. The multiplicity behavior of CSTR has been reviewed. Brooks [8] revealed unique bifurcation diagrams considering different parameters. Input multiplicity analysis has been carried out in non-isothermal CSTR for hydrolysis of acetic anhydride. The results have been analyzed by Hashim and Thomas [9] using the time concentration plot taking the experimental data. Thornhill [10] proposed simulation of continuous heat reactor. Experimental data including the disturbances have been analyzed for fault detection and diagnosis.

Nonlinear identification model has been applied to exothermic CSTR data. This model has been approximated by Hammerstein model with noise model, which has been proposed by Aljamaan et al. [11]. Simulation of CSTR has been done using Range–Kutta's method. The results have been obtained for different control inputs. The results obtained from simulation, proposed by Vojtesek and Dostal [12], are beneficial for security reasons. Malinen et al. [13] proposed a homotopy-based strategy for finding multiple steady-state solutions of CSTR. A homotopy path has been tracked on which all the steady-state solution lies. Nguyen et al. [14] proposed multiplicity behavior analysis in CSTR for polystyrene production through system theory tools and heat balance concept. Bifurcation diagrams enabled the visualization of multiplicity behavior. Alankar et al. [15] proposed nonlinear system identification method leading to polynomial nonlinear state-space models.

Discrete adaptive control method has been proposed for Weiner nonlinear systems by Yuan et al. [16]. Recursive least squares algorithm has been used for updating uncertain parameters. Oechsler Bruno et al. [17] examined the micromixing effects on bifurcation behavior of CSTR through Interaction-by-Exchange-with-the-Mean model. Trollberg and Jacobsen [18] proposed process dynamics from which steady-state properties have been evaluated. Extremum points in the input–output map

have been calculated. Phase-lock loop control has been carried out to decrease the frequency when nearer to steady-state conditions.

Kulikov and Kulikova [19] proposed state estimation technique for continuous-discrete nonlinear stochastic systems through Kalman filtering. Jiang-Bo et al. [20] proposed global robust output tracking control through dynamic switching mechanism for a type of cascade uncertain nonlinear systems. Output feedback tracking control has been introduced with reduced-order observer. Du and Johansen [21] presented control-relevant nonlinearity measure (CRNM) to measure the degree of nonlinear system. Closed-loop simulations have been carried out which revealed that the CRNM-based integrated multi-model control approach can get controller bank and schedule the local controllers.

However, in all these review papers, no such range of temperature has been identified, in which the dynamics become unstable. Comparative analyses of the previous works are shown in Table 4. The chapter has been organized as follows. The mathematical modeling using nonlinear differential equations and linearization of the system has been done in Sect. 2. The state-space modeling and phase portrait analysis under different initial conditions have been proposed in Sect. 3. Results and conclusions of the simulation have been concluded in Sect. 4 and Sect. 5, respectively.

2 Mathematical Model and Linearization of CSTR

In this work, an exothermic reaction has been considered, in which chemical species A reacts to form chemical species B. A is the ethylene oxide which reacts with water to form ethylene glycol, B. Here, the intermediate species water is not considered into account. The heat of the reaction is removed by a coolant medium that flows around the reactor. The fluid inside the reactor is fed via the overhead tank through inlet pipes, through which the feed concentration is monitored via flow rate. The fluid inside the reactor is perfectly mixed, and the product is obtained through the exit valve. The reaction rate has been assumed to be of first order which follows Arrhenius relation. A block diagram of CSTR is illustrated in Fig. 1.

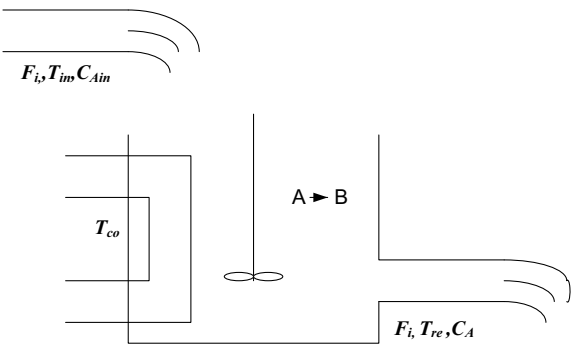


Fig. 1 Continuous stirred tank reactor

The concentration $C_A(t)$ of the participating chemical and the tank temperature $T_{re}(t)$ has been equivalently spread in the reactor and in the outlet flow. The stream rate F_i across the reactor has been considered unchanged, and therefore, the liquid volume V_{liq} becomes constant. Using material balance and energy balance equations, Eqs. (1) and (2) have been obtained. Implementing the strategies mentioned in [2], CSTR has been modeled by the following nonlinear differential equations.

$$f_1(C,T) = \frac{\mathrm{d}T_{re}(t)}{\mathrm{d}t} = \frac{Fi}{V_{\text{liq}}}(T_{\text{in}}(t) - T_{re}(t)) + \frac{UA}{V_{\text{liq}}C_h\rho} (T_{co}(t) - T_{re}(t)) + \frac{-\Delta Hr}{\rho C_h} \cdot k \cdot (T_{re}(t)) \cdot C_A(t) \tag{1}$$

$$f_2(C,T) = \frac{\mathrm{d}CA(t)}{\mathrm{d}t} = \frac{F_i}{V_{\text{liq}}}(CA_{\text{in}}(t) - CA(t)) - k(T_{re}(t)) \cdot CA(t) \tag{2}$$

where

'ρ'	density of the liquid in kmol/m^3,
'$T_{\text{in}}(t)$'	temperature of the inlet flow in K,
'F_i'	flow rate in m^3/min,
'$Vliq$'	volume of the liquid in m^3,
'C_h'	heat capacity in J/K,
'ΔHr'	heat of reaction of A in KJ/mol,
'A'	heat transfer area in m^2,
'U'	total heat transfer coefficient in W/m^{2}K,
'$T_{co}(t)$'	temperature of the coolant in K and
'$C_{Ain}(t)$'	inlet flow's concentration in kmol/m^3.

The reaction rate constant 'k' which is the function of temperature '$T_{re}(t)$' is governed by Arrhenius law stated by the following equation,

$$k(T_{re}(t)) = ko \exp\left(-\frac{E}{RT_{re}(t)}\right) \tag{3}$$

where 'k'$_o$ is the frequency factor, 'E' is the activation energy, 'R' is the ideal gas constant. 'C'$_{Ani}(t)$ and 'T'$_{\text{in}}(t)$ have been assumed constants.

The numeral values of the constants, initial conditions and parameters used in the CSTR for carrying out this reaction where chemical species A reacts to form

Table 1 Numeral values of constants and initial conditions

Constants	Values (units)	Initial conditions	Values (units)
F_i/V_{liq}	1 (1/min)	$T(0)$	311.266 (K)
C_{Ain}	10 (kmol/m^3)	$C_A(0)$	8.57 (kmol/m^3)
T_{in}	298.2 (K)	$T_{co}(0)$	297.977 (K)

Table 2 Numerical values of Parameters

Parameters	Values (units)
$E\backslash R$	5963.242 (K)
k_0	14825*3600 (1/min)
$\Delta H/(\rho Ch)$	5960/500 (Km^3/kmol)
$UA/(\rho Vl_{iq}C_h)$	150/500 (1/min)

chemical product B are mentioned in Table 1 [22]. The nonlinear CSTR model has been linearized at equilibrium values (Table 2).

The reactor has been modeled based on its energy balance and material balance equations denoted by equations. Dynamics are nonlinear which have been linearized using Taylor series. The general representation of Taylor series has been given by the following equation [1]:

$$f(x) = f(xs) + \partial f/\partial x(x - xs) + 1/2(\partial^2 f/\partial x^2) \\ (x - xs)^2 + \text{higher order terms} \tag{4}$$

where, 'x_s' is the operating point. Applying this Taylor series and linearizing the CSTR model around operating points (C_{Ao}, T_o), we get,

$$\exp\left(-\frac{E}{RT}\right)CA = \exp\left(-\frac{E}{RTo}\right)CAo + \frac{E}{RTO^2}CAo \\ (T - TO) + \exp\left(\frac{-E}{RTo}\right)(CA - CAo) \tag{5}$$

Applying the linearized equation and converting Eqs. (1), (2) into a deviation variable form, we obtain the following [23] equations:

$$\frac{\mathrm{d}(C_A - C_{AO})}{\mathrm{d}t} = -\left[\frac{Fi}{V} + k_o \exp\left(-\frac{E}{RT_o}\right)\right](C_A - C_{AO}) \\ - k_o \frac{E}{RT_o^2} \exp\left(-\frac{E}{RT_o}\right)C_{AO}(T - T_0) + \frac{F_i}{V}(C_i - C_{io}) \tag{6}$$

$$\frac{\mathrm{d}(T - To)}{\mathrm{d}t} = \frac{-\Delta Hr}{\rho C}k_o \exp\left(-\frac{E}{RT_o}\right)(C_A - C_{AO}) + \\ \left[\frac{-\Delta Hr}{\rho C}k_o \cdot \frac{E}{RT_O^2} \exp\left(-\frac{E}{RT_o}\right)C_{AO} - \frac{UA}{VC\rho} - \frac{Fi}{V}\right] \\ (T - To) + \frac{UA}{VC\rho}(T_C - T_{CO}) + \frac{Fi}{V}(T_i - T_{io}) \tag{7}$$

where $(C_A - C_{AO})$ and $(T - T_o)$ are the deviation state variables and C_o, C_{io}, T_o, T_{io}, T_{co} are the operating conditions, where subscript *o, i, c* denotes the initial conditions, inlet feed and coolant, respectively.

After Taylor series approximation, the state matrices have been obtained putting the parameter constants and initial conditions of CSTR mentioned in Table 1.

3 State Space Modeling and Phase Portrait Analysis of CSTR

After doing the Taylor series approximation, the state space modeling of the CTSR has been done. Considering Eqs. (6) and (7), Eqs. (8) and (9) have been obtained where the states are C' and T' input is T_{co} and the output is T'; A, B and C matrices, namely state matrix, input matrix and output matrix, respectively.

$$\begin{bmatrix} \frac{dC'}{dt} \\ \frac{dT'}{dt} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} C' \\ T' \end{bmatrix} + \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} T_{co} \tag{8}$$

$$y = \begin{bmatrix} C_{11} & C_{12} \end{bmatrix} \begin{bmatrix} C' \\ T' \end{bmatrix} \tag{9}$$

Here, elements of the matrix A are as follows,

$A_{11} = -\left(\frac{F_i}{V} + k_o \exp(-E/RT_o)\right)$,

$A_{12} = (-k_o \frac{E}{RT_o^2} \exp(-E/RT_o) * C_o)$,

$A_{21} = -\Delta H/\rho C_p k_o \exp(-E/RT_o)$,

$$A_{22} = [-\Delta H/\rho C_p k_o (\frac{E}{RT^2}) \exp(-E/RT_o) C_o + (-UA/\rho VC - F_i/V)]$$

$B_{11} = 0$,

$B_{12} = UA/V\rho C$,

$C_{11} = 0$,

$C_{12} = 1$.

Substituting the first initial condition and the constant parameters in the (8) and (9) from the Table 1, we get,

$$A = \begin{bmatrix} -1.1669777 & -0.088016993 \\ -1.98841098 & -2.3491625 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0.3 \end{bmatrix}, C = [0 \quad 1] \tag{10}$$

By substituting the constant values and the initial conditions in the above equations, the following transfer function has been obtained from the state space matrices. As we have considered single input single output, it is a SISO model.

$$G(S) = \frac{N(s)}{D(s)} \tag{11}$$

where $N(s)$ = numerator of the transfer function $G(s)$ and $D(s)$ is the denominator of the transfer function $G(s)$, which has been represented in the Eq. (12)

$$G(s) = \frac{0.3000s + 0.3501}{s^2 + 3.5161s + 2.5664} \tag{12}$$

Phase portrait has been considered to explore the region of stability and instability in the dynamics of CSTR. The trajectories of the state space of CSTR have been illustrated at the equilibrium points. Equilibrium points have been obtained tending the function to zero. Functions have been denoted by Eqs. (1) and (2), respectively, as $f_1(C, T)$ and $f_2(C, T)$.

Taking different initial conditions, eight sets of equilibrium points have been obtained as, $P_0 = [0.8609\ 4.83197]$. $P_1 = [-3.992\ 9.635]$, $P_2 = [2.2303\ -9.22260]$, $P_3 = [10.888\ -7.343]$, $P_4 = [3.345\ 18.245]$, $P_5 = [-5.512\ 25.327]$, $P_6 = [-5.1956\ 30.997]$, $P_7 = [-6.1410\ 45.623]$. The Jacobian matrix [13] of the CSTR has been used to determine the eigenvalues, from which eigenvectors have been calculated, and thus, the unstable region has been observed.

The generalized Jacobian matrix has been given in the form:

$$J(C, T) = \begin{bmatrix} \partial f_1/\partial C & \partial f_1/\partial T \\ \partial f_2/\partial C & \partial f_2/\partial T \end{bmatrix} \tag{13}$$

where C and T are the variables concentration and temperature, respectively. At each equilibrium points, and the following eigenvalues have been obtained, $E_0 = [-1.0339\ -2.48223]$, $E_1 = [-1.0552\ -1.3170]$, $E_2 = [-1.0488\ -3.721783]$, $E_3 = [-2.41931\ -0.3828]$, $E_4 = [0.03376\ -3.226808]$, $E_5 = [-3.4696\ 0.0965]$, $E_6 = [-1.094463\ -6.538836]$, $E_7 = [-1.1456715\ -9.2384]$. The eigenvectors corresponding to eigenvalues E_0, E_1, E_2, E_3, E_4, E_5, E_6, E_7 are Q_{01}, Q_{02}, Q_{11}, Q_{12}, Q_{21}, Q_{22}, Q_{31}, Q_{32}, Q_{41}, Q_{42}, Q_{51}, Q_{52}, Q_{61}, Q_{62}, Q_{71}, Q_{72}, respectively, mentioned in Table 3. There are two sets of eigenvectors corresponding to each set of eigenvalues. The eigenvalues having negative values are considered stable, while with positive values are considered unstable, and one with both positive value negative value is considered as unstable saddle point. These eigenvalues concluded that the system is unstable and has a saddle point at E_4 and E_5 equilibrium point, while stable at all other equilibrium points. The phase portrait at these equilibrium points has been obtained by MATLAB simulation giving the same results as expected theoretically.

4 Results

The phase portraits for different eigenvalues of the system have been simulated in MATLAB using the ode45 solver. The phase portrait for the eigenvalues has been observed to be same as expected. The phase portrait obtained at the negative eigenvalues is converging toward the origin as shown in Figs. 2, 3, 4, 5, 8 and 9,

Table 3 Initial conditions with its eigenvalues

S.no.	Initial conditions	Eigenvalues and eigenvectors	Results
1	T(0) = 311.26 K, C(0) = 8.57 kmol/m^3, Tco(0) = 297.98 K	$E_0 = -[1.0339\ -2.48223]$, $Q_{01} = [1\ -1.511]^T$, $Q_{02} = [1\ 14.9431]^T$	Stable
2	T(0) = 318.25, C(0) = 10 kmol/m^3, Tco(0) = 290.95 K	$E_1 = [-1.0552\ -1.3170]$, $Q_{11} = [-0.28559\ 1]^T$, $Q_{12} = [0.01975\ 1]^T$	Stable
3	T(0) = 325.75 K, C(0) = 6.85 kmol/m^3, Tco(0) = 305.24 K	$E_2 = [-1.0488\ -3.721783]$, $Q_{21} = [1\ -2.31]^T$, $Q_{22} = [1\ 13.3961]^T$	Stable
4	T(0) = 330.25 K, C(0) = 6.25 kmol/m^3, Tco(0) = 285.25 K	$E_3 = [-2.41931\ -0.3828]$, $Q_{31} = [1\ 0.18707]^T$, $Q_{32} = [-0.15329\ 1]^T$	Stable
5	T(0) = 340.25 K, C(0) = 5.15 kmol/m^3, Tco(0) = 296.79 K	$E_4 = [0.0337.-3.22680]$, $Q_{41} = [1\ -7.97285]^T$, $Q_{42} = [1\ 5.51]^T$	Unstable saddle
6	T(0) = 345.25 K, C(0) = 4.25kmol/m^3, Tco(0) = 292.52 K	$E_5 = [-3.4969\ 0.0965]$, $Q_{51} = [0.1674\ 1]^T$, $Q_{52} = [0.17992\ 1]^T$	Unstable saddle
7	T(0) = 350.25 K, C(0) = 3.55kmol/m^3, Tco(0) = 290.54 K	$E_6 = [-1.094463\ -6.5388]$, $Q_{61} = [1\ -5.4824]^T$, $Q_{62} = [1\ 12.58729]^T$	Stable
8	T(0) = 370.13 K, C(0) = 2kmol/m^3, Tco(0) = 305.03 K	$E_7 = [-1.145671\ -9.2384]$, $Q_{71} = [1\ 1.2489]^T$, $Q_{72} = [1\ 12.370487]^T$	Stable

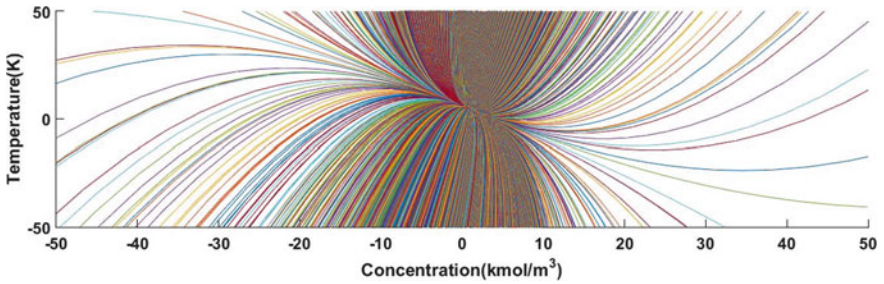


Fig. 2 Phase portrait for the eigenvalue E_0

while obtained at eigenvalues having both positive and negative has unstable saddle point shown in Figs. 6 and 7. The trajectory of the unstable saddle point starts at infinite distant away, moves toward, but never converges to the equilibrium point.

To analyze the trajectories of the nonlinear differential equations, the time series for the concentration of CSTR has also been obtained as shown in Fig. 10. Concentration of the reactant has been observed to be decreasing with the time, which has

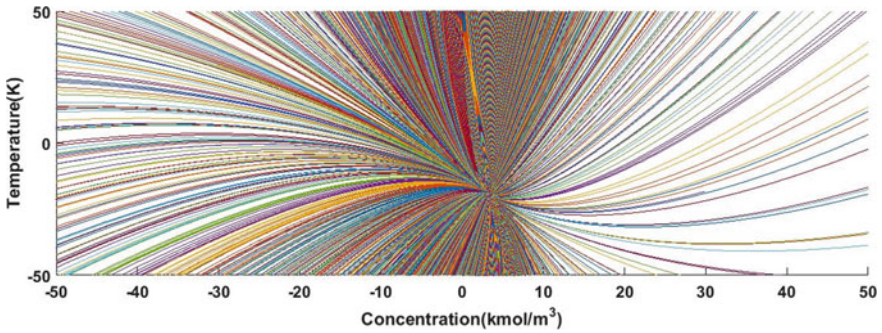


Fig. 3 Phase portrait for the eigenvalue E_1

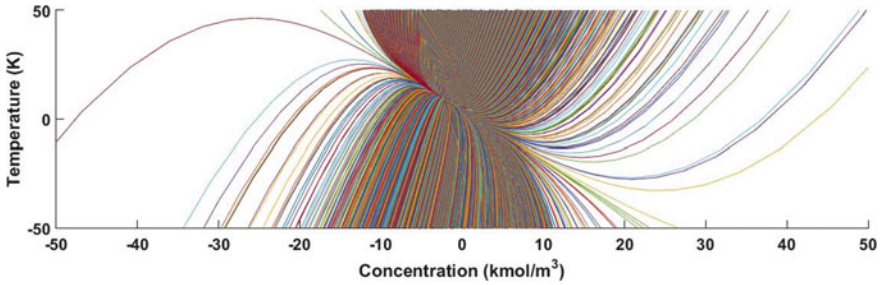


Fig. 4 Phase portrait for the eigenvalue E_2

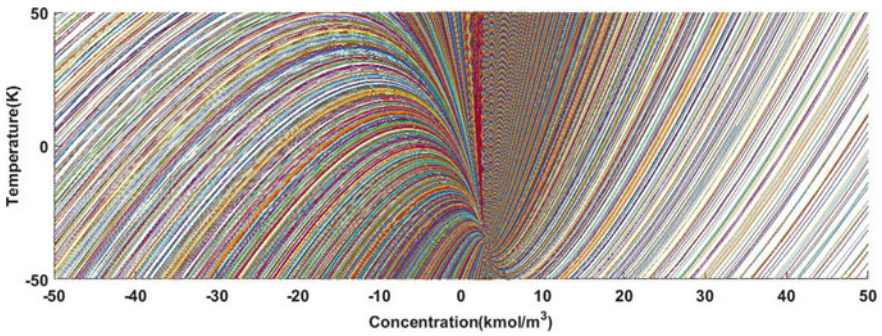


Fig. 5 Phase portrait for the eigenvalue E_3

been observed to be valid. During the course of the reaction, the concentration of the reactant decreases with the rise in temperature to yield better output.

Figure 11 displays vectors associated with Eqs. (6), (7) at the equilibrium points. The phase portrait of eigenvectors is depicted in Figs. 11, 12, 13, 14, 15, 16, 17 and 18 in red.

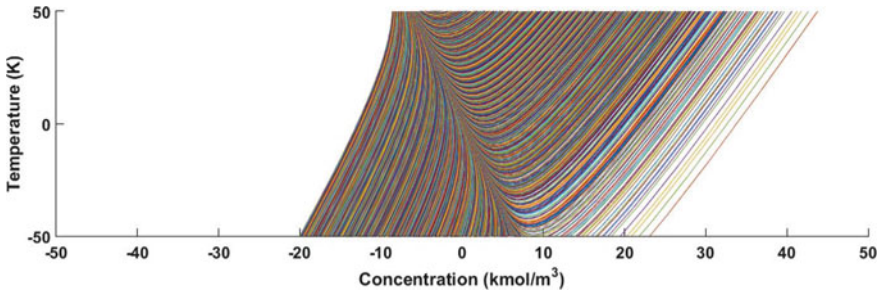


Fig. 6 Phase portrait for the eigenvalue E_4

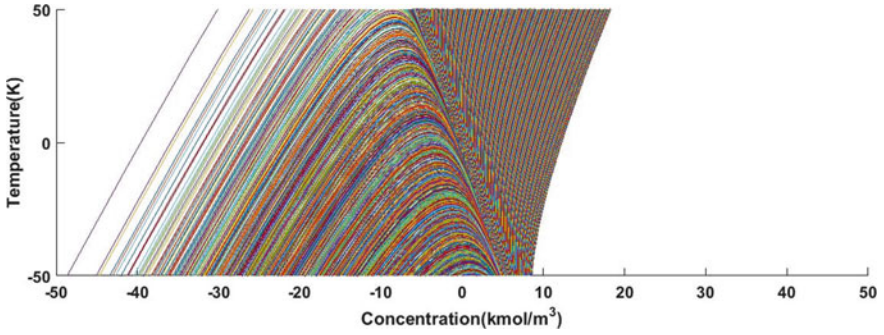


Fig. 7 Phase portrait for the eigenvalue E_5

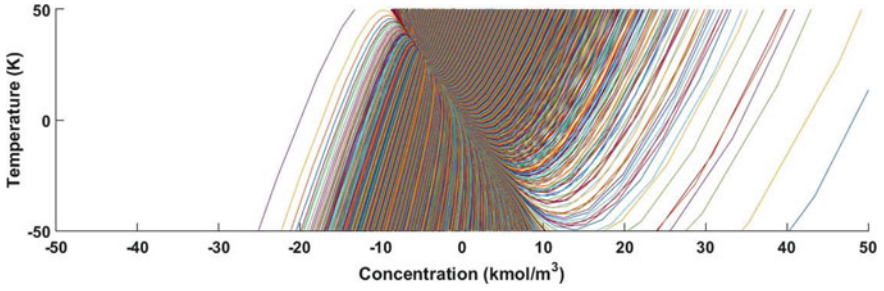


Fig. 8 Phase portrait for the eigenvalue E_6

Separate eigenvectors have been plotted, namely E_0, E_1, E_2, E_3, E_4, E_5, E_6, E_7 at equilibrium points. The vector arrow direction and the length represent the corresponding values in the Eqs. (6), (7). These eigenvectors also define the separatices that determine the characteristic behavior of state trajectories. Plotting slope marks give a qualitative feel for phase plane behavior analysis. Figures 11, 12, 13, 14, 15, 16, 17 and 18 depict the phase portrait for the eigenvectors $V1$, $V2$ corresponding to

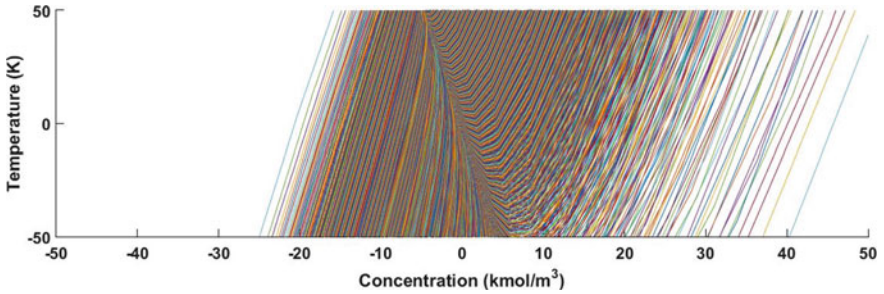


Fig. 9 Phase portrait for the eigenvalue E_7

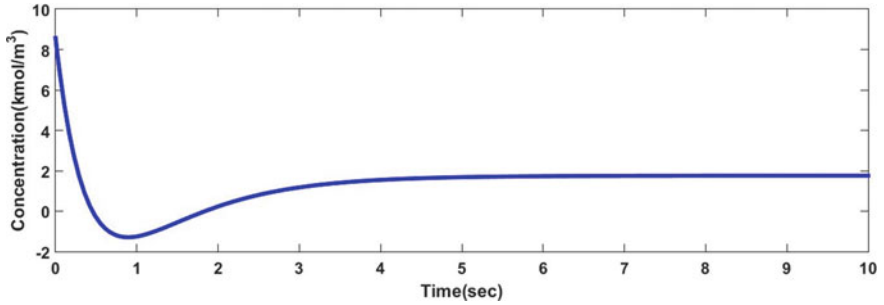


Fig. 10 Time series plot for concentration of the reactant

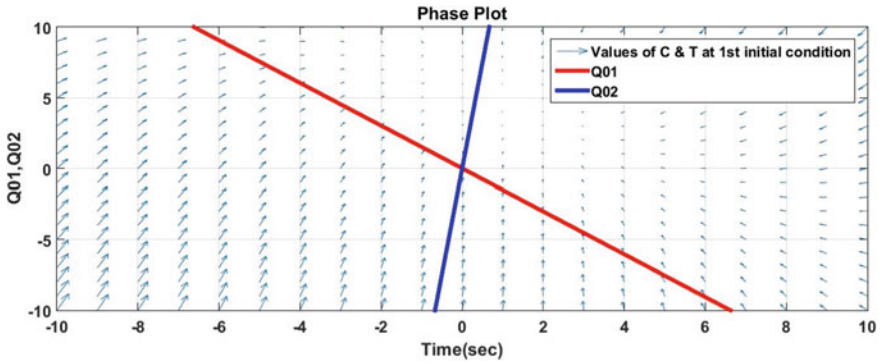


Fig. 11 Phase portrait of eigenvectors corresponding to eigenvalues E_0

the eigenvalues E_0, E_1, E_2, E_3, E_4, E_5, E_6, E_7, respectively. Also, for better pictorial representation, three-dimensional phase portrait has been obtained. It has been obtained to depict the behavior of the system variables, to show how the variables which vary with respect to time in Fig. 19. This gave an idea about the trajectory pattern, which the variables of CSTR will follow.

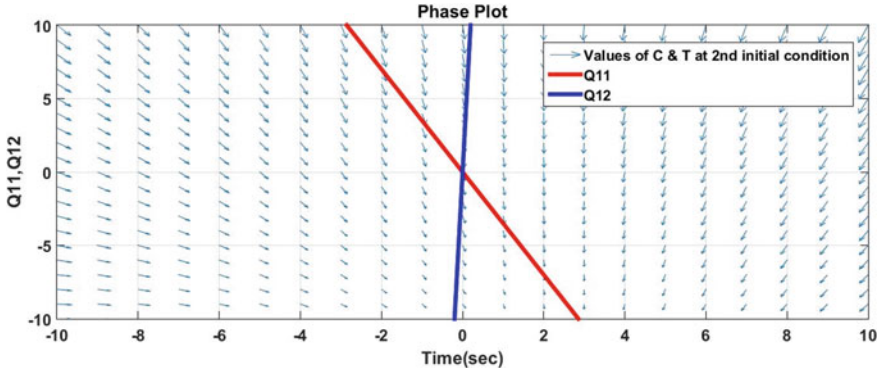


Fig. 12 Phase portrait of eigenvectors corresponding to eigenvalues E_1

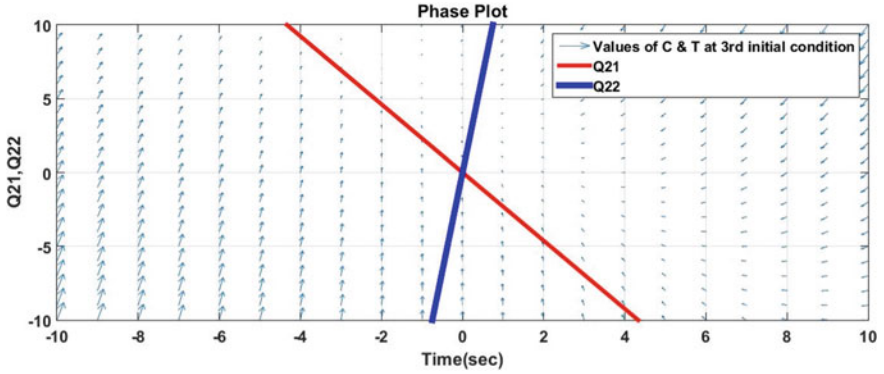


Fig. 13 Phase portrait of eigenvectors corresponding to eigenvalues E_2

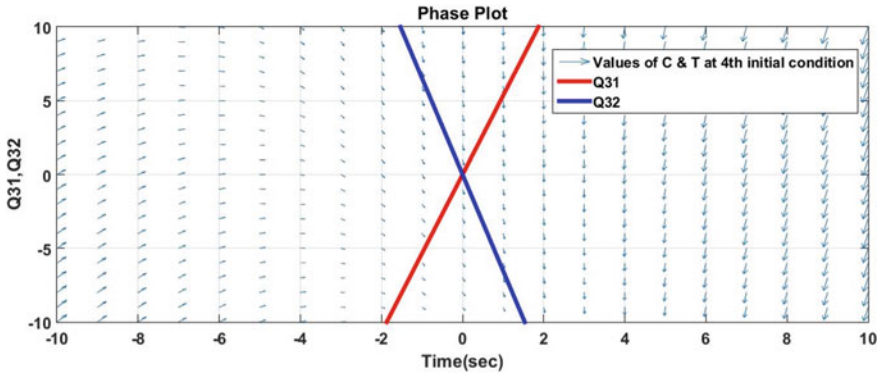


Fig. 14 Phase portrait of eigenvectors corresponding to eigenvalues E_3

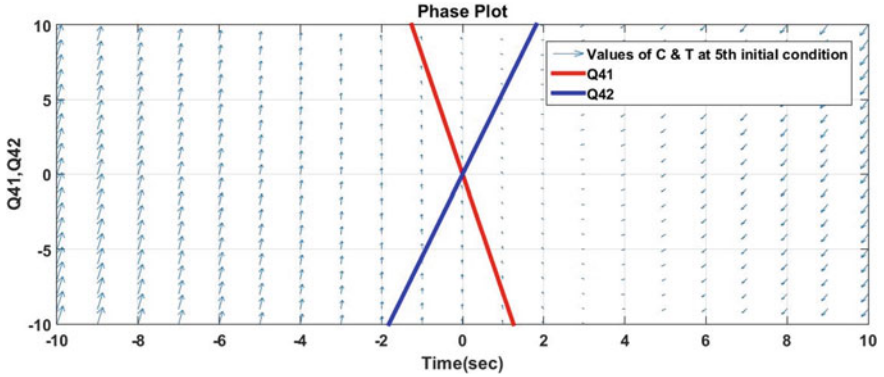


Fig. 15 Phase portrait of eigenvectors corresponding to eigenvalues E_4

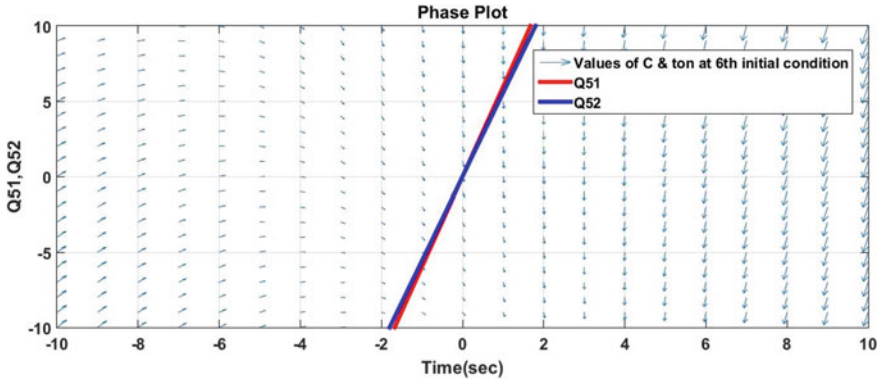


Fig. 16 Phase portrait of eigenvectors corresponding to eigenvalues E_5

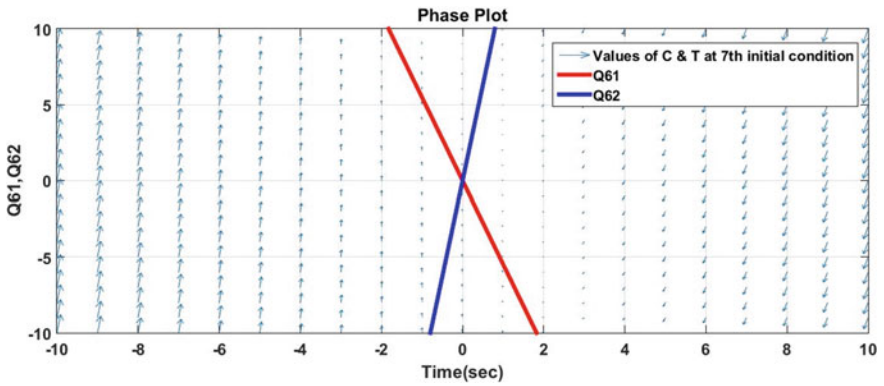


Fig. 17 Phase portrait of eigenvectors corresponding to eigenvalues E_6

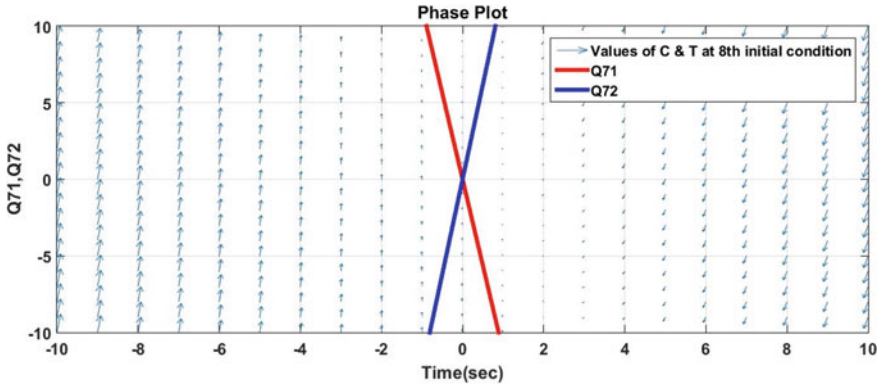


Fig. 18 Phase portrait of eigenvectors corresponding to eigenvalues E_7

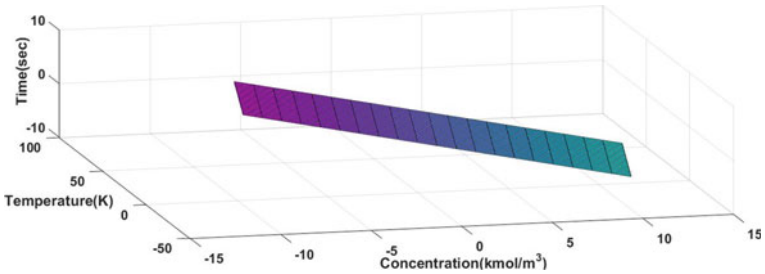


Fig. 19 Three-dimensional phase portrait of CSTR parameters

Comparative analysis of the previous papers related to phase portrait is done in Table 4.

5 Conclusion

The study of dynamic behavior of CSTR comprised the stable and unstable graphical analysis. The phase portrait of CSTR has been validated with simulation results at the equilibrium points. The results of stable and unstable regions of operations are explained in Table 4. The reactor has been observed to be unstable in the certain range of operation, from 335 to 345 K. In this range of temperature, under any initial conditions, the system remains unstable.

Three-dimensional plot has also been obtained. This is necessary in order to get a prerequisite of temperature levels which could be reached, making the system unstable. Also, ensure whether the reaction in the reactor is taking place efficiently, whether the concentration is decreasing with course of time and giving the desired product. Analyzing the concentration and temperature is therefore important, which

Table 4 Comparative analysis

Paper	Type of CSTR	Dynamics	Range of operation
[4]	Adiabatic	Unstable and stable limit cycles were observed in phase portrait	No range of stable or unstable operation has been mentioned
[6]	Exothermic	Phase portrait analysis for behavior in emergency conditions	No range of stable or unstable operation has been mentioned
[7]	Isothermal and adiabatic	Phase portrait analysis resulted in multiple steady-state solutions	Stable and unstable branches in the limit cycle have been mentioned. No range of unstable operation has been mentioned
[8]	Adiabatic	Bifurcation diagrams have been used for safe region of operation	No range of stable or unstable operation has been mentioned
In this paper	Non-isothermal	Phase portrait for eigenvalues has been done	Unstable range of temperature has been mentioned

has been done in this work, so that appropriate control could be implemented to make the system stable.

To keep the CSTR unaffected, under any initial conditions, control strategy could be applied. Future works in this direction could be designing control strategies in order to make the system follow the desired trajectory and control the coolant temperature so that the system does not get overheated and also laying emphasis on obtaining desired product with minimal cost of production.

References

1. Bequette BW (2003) Process control : modeling, design, and simulation. Prentice Hall PTR
2. Seborg DE (2011) Process dynamics and control. Wiley
3. Ray WH (1974) On the dynamic behavior of continuous stirred tank reactors. Chem Eng Sci 29:967–985
4. Berezowski M (2017) Limit cycles that do not comprise steady states of chemical reactors. Appl Math Comput 312:129–133
5. Berezowski M (2010) The application of the parametric continuation method for determining steady state diagrams in chemical engineering. Chem Eng Sci 65(19):5411–5414
6. Bykov VI, Serafimov LA, Tsybenova SB (2015) Emergency starting regimes of a continuous stirred tank reactor. Theor Found Chem Eng 49(4):380–388
7. Yermakova A, Anikeev VI (2005) Analysis of the CSTR with two-phase flow under phase equilibrium conditions. Chem Eng Sci 60:3199–3206
8. Brooks K (2013) Steady-state multiplicity in an adiabatic continuous stirred tank reactor with vapor recycle. AIChE J 59(2):553–559

9. Hashim MA, Thomas MT (2010) Input multiplicity analysis in a non-isothermal CSTR for acid-catalyzed hydrolysis of acetic anhydride. Chem Eng Technol Ind Chem Plant Equip Process Eng Biotechnol 33(3):499–507
10. Thornhill NF (2008) A continuous stirred tank heater simulation model with applications. J Process Control 18(3–4):347–360
11. Aljamaan I, Westwick D, Foley M (2015) Non-linear closed-loop identification of CSTR in the presence of a non-stationary disturbances. IFAC-PapersOnLine 48(28):1029–1034
12. Vojtesek J, Dostal P (2015) Using simulink in simulation of dynamic behaviour of nonlinear process. In: Computer science on-line conference. Springer, pp 285–298
13. Malinen I, Kangas J, Ahola J, Tanskanen J (2016) A new homotopy-based strategy for the robust determination of all the feasible solutions for CSTR systems. Period Polytech Chem Eng 60(1):8–23
14. Nguyen ST (2017) Analysis of the steady-state multiplicity behavior for polystyrene production in the CSTR. Chem Product Process Model 12(4):1–12
15. Alanqar A, Durand H (2015) On identification of well-conditioned nonlinear systems: application to economic model predictive control of nonlinear processes. AIChE J 61(10):3353–3373
16. Yuan P, Zhang B, Mao Z (2017) A self-tuning control method for Wiener nonlinear systems and its application to process control problems. Chin J Chem Eng 25(2):193–201
17. Oechsler BF, Melo PA, Pinto JC (2015) Micromixing effects on the dynamic behavior of continuous stirred tank reactors. Appl Math Model 40(7–8):1–17
18. Trollberg O, Jacobsen EW (2015) Connecting optimality to process dynamics. IFAC-Papers OnLine 48(8):170–175
19. Kulikov GY, Kulikova MV (2019) Numerical robustness of extended Kalman filtering based state estimation in ill-conditioned continuous-discrete nonlinear stochastic chemical systems. Int J Robust Nonlinear Control 29(5):1377–1395
20. Yu J, Zhao Y, Wu Y (2018) Global robust output tracking control for a class of uncertain cascaded nonlinear systems. Automatica 93:274–281
21. Du J, Arne T (2017) Control-relevant nonlinearity measure and integrated multi-model control. J Process Control 57:127–139
22. Gain-Scheduled Control of a Chemical Reactor—MATLAB and Simulink Example—MathWorks India. https://in.mathworks.com/help/control/examples/gain-scheduled-control-of-a-chemical-reactor.html. Accessed on 16 June 2019
23. Banu US, Uma G (2008) Fuzzy gain scheduled CSTR with GA based PID. Chem Eng Commun 195(10):1213–1226

Machine Learning-Based Implementation of Image Corner Detection Using SVM Algorithm for Biomedical Applications

Santosh M. Herur, S. S. Kerur, and H. C. Hadimani

Abstract Support vector machine approach in machine vision with the help of OpenCV simulation tool is used for corner detection. Path of the maximum gray-level changes meant for every edge-pixel is calculated in the picture, this edge-pixel is represented by four-dimensional feature vectors. It is made up of count of the edge pixels in window center and has four directions since their maximum gray-level direction change. This feature vector and support vector are used for designing of support vector machine. For corner detection, it represents critical points in a classification. This algorithm is straight forward with less computational complexity. It has machine learning capability which gives good results.

Keywords Machine learning · Corner detection · SVM algorithm · Biomedical application

1 Introduction

Here, we have implemented and studied corner detection using SVM algorithm for biomedical application using Raspberry-Pi which is the tiny single-board computer with Linux OS running on it. Machine learning is a technique which is like adding brain to machines so that it can do task by itself. It is employed to train machines to handle information with additional efficiency. Every now and then, after viewing information, the pattern cannot be interpreted. In such situation, the machine learning can be applied with computer vision. With large quantity of datasets present, demand

S. M. Herur (✉) · H. C. Hadimani
Department of E&CE, GMIT, Davangere, India
e-mail: santoshmherur@gmail.com

H. C. Hadimani
e-mail: hchadimani9@gmail.com

S. S. Kerur
Department of E&CE, SDMCET, Dharwad, India
e-mail: kerurss@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_19

of machine learning is high. Several fields such as medicine, military, etc., apply a machine learning in order to extract the significant information [1]. As computer power grown over the years, there is lot of demand in employing highly developed algorithms for facilitating our daily use of medical images for enhancing information which we gain from those [2].

Image is the function of two variables $f[x, y]$, which is a intensity 'f' at the [x, y] location. This image is further processed using computer for the image enhancement, segmentation, restoration, description, coding, recognition, reconstruction and transformation which is known as image processing [3]. Corner play an important role on image processing [4]. Image corners are nothing but image location which has larger intensity variations occurring in more than one direction. Corner detection in image is "the approach employed within a computer vision system to extract some sort of the features and understand the contents of image." Image corner detection is often advantageous in the motion detection; image registration; video tracking; image-mosaicing; panorama stitching; 3D-modeling; object recognition. Support vector machine (SVM) belongs to the family of generalized linear classifiers. SVMs shall be defined as the system employing hypothesis space of the linear functions in an high-dimensional feature space, which is trained by means of learning algorithm from a theory of optimization which implements a learning bias derived from the theory of statistical learning [5, 6]. Biomedical is an application of principles of engineering and design concepts to biology and medicine for healthcare purpose (e.g., diagnostic/therapeutic) [7]. The image is captured through Raspberry-Pi camera and corner detection algorithm is tested. This concept can be used in real-time biomedical applications. Due to the small size and the less weight, it is easy to use in design [8].

2 Support Vector Machine Image Corner Detection Algorithm

Image corner is the region in an image with huge changes in the intensity in every direction. Such attempt for locating the corners had implemented by [9]. A small variation is made in Harris corner detector by [10]. In this case, a SVM is used as a classifier [11]. In the Harris corner detector, scoring function was expressed as:

$$R = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 \tag{1}$$

Alternative to Eq. (1), Shi-Tomasi has proposed:

$$R = \min(\lambda_1, \lambda_2) \tag{2}$$

If this is greater than the threshold value, then it will be treated as a corner. If given labeled training data, then algorithmic rule of SVM output a best possible hyper-plane that categorizes some new examples.

Hyper-plane is obtained finest by considering the subsequent straightforward example:

For the intended linearly distinguishable set of two-dimensional points that pertains to 1 among 2 classes, let us locate the separating line.

In Fig. 1, we can notice that there are many lines that provide way out to problem. We may spontaneously term a measure for estimation of worth of lines, i.e., line is not good if it passes too much nearer to points since it happens to be sensitive to noise and may not be accurately generalized. Hence, the target must be to locate the line which passes almost from each and every point. To classify a breast mammogram image as normal or abnormal, improved classification algorithm was developed [12].

Support vectors usually are nothing but subset of data point which can maximize the minimum margin [13]. In that case, the procedure of the SVM algorithmic rule is on the basis of finding hyper-plane that provides the biggest minimum distance to training example [14]. Two times of this distance gains an important tag of margin inside the theory of SVM. Hence, the best partitioning hyper-plane maximize margin of training information as depicted in Fig. 2.

We compute the optimal hyper-plane as follows:

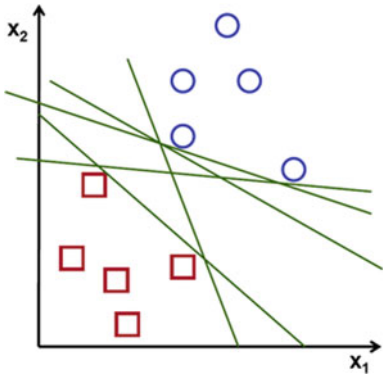


Fig. 1 Two set of data

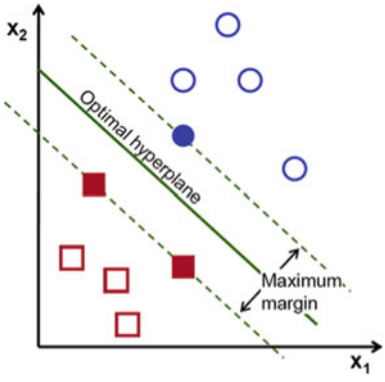


Fig. 2 Linearly separable data

Let us initiate notation that is used in defining formally a hyper-plane:

$$f(x) = \beta_0 + \beta^T \chi \tag{3}$$

where β is *the* weight vector and β_0 is the bias.

The best hyper-plane is effective representation in various ways by the scaling of β and β_0. Among every possible representations, the one selected is

$$\left|\beta_0 + \beta^T \chi\right| = 1 \tag{4}$$

where χ symbolize the training examples nearest to the hyper-plane and are usually termed as support vectors.

At this moment, let us use the end result of geometry which provides distance between point 'χ' and hyper-plane (β, β_0)

$$\text{distance} = \frac{\left|\beta_0 + \beta^T \chi\right|}{\|\beta\|} \tag{5}$$

But for support vector, the numerator = 1 and, therefore,

$$\text{distance}_{\text{support} - \text{vectors}} = \frac{\left|\beta_0 + \beta^T \chi\right|}{\|\beta\|} = \frac{1}{\|\beta\|} \tag{6}$$

Remember that margin which was introduced in preceding part is denoted here by 'M' and is two times of $\text{distance}_{\text{support-vectors}}$:

$$M = \frac{2}{\|\beta\|} \tag{7}$$

Ultimately, maximizing the margin 'M' is almost nothing but minimizing function '$L(\beta)$' which is subjected to a few constraint. The requirements for hyper-plane are modeled by constraints in order to categorize properly all 'χ_i'.

Formally,

$\min\limits_{\beta,\beta_0} L(\beta) = \frac{1}{2}\|\beta\|^2$ is subjected to the

$$y_i(\beta^T \chi_i + \beta_0) \geq 1 \quad \forall i, \tag{8}$$

where y_i represent all the labels of training examples.

Image corners are nothing but the principal local features in image [15]. To decide whether the candidate point is a corner, the region in the neighborhood of the corner is analyzed. Surrounding pixels are analyzed to check if they are brighter or darker than pixel being analyzed. This is done by checking if all the pixels are greater than

pixel being tested value + threshold value or smaller than pixel being tested value—threshold value. Global gradient thresholds auto-estimation can be done for image restoration [16].

The value of neighboring pixels is chosen. If point p is a corner, then few pixels must be brighter than pixel being tested value + threshold or darker than pixel being tested value—threshold value. When both conditions are not satisfied then the point cannot be a corner. Thus, this test can be used to reject corners. If all these criteria are satisfied, then we test the entire line segment. Given a pixel point p and image matrix, we need to quickly identify the location of the different neighboring points.

The location of different points relative to point p are (0, −3), (1, −3), (2, −2), (3, −1), (3, 0), (3, 1), (2, 2), (1, 3), (0, 3), (−1, 3), (−2, 2), (−3, 1), (−3, 0), (−3, −1), (−2, −2), (−1, −3). These relative co-ordinates from the center pixels are stored in a vector and can be used to obtain corner. Let k be the neighboring pixel and we want to test if $k > p + t$ or $p - k < -t$ and $k < p - t$ or $p - k > t$. If $k - p < -t$, we will label pixel as 1 and if $k - p > t$, we will label pixel as 2. The range of values taken by $k - p$ need to be determined. Here, k can take values from 0 to 255 and p can take values from 0 to 255. Thus, the minimum values if −255 and the maximum value 255. Thus, we construct a lookup table of size 512.For values of $k - p$ from −255 to 255 which satisfy the condition $k - p < -t$ value 1 is assigned to corresponding element of look up table. For values of $k - p$ which do not satisfy both the condition are assigned value 0 the lookup table. Consider that values of k and p are given. Thus, the task is to compute the index $k - p$ and look up the corresponding values from the lookup table. Thus, given pixel $p(x, y)$ we need to derive all the pixel using relative offset and look up their corresponding values from the look up table. We have the pointer to location of pixel $p(x, y)$. Let x and y offsets be denoted by (dx, dy). Then, location of the offset pixel is given by ptr + dy * cols + dx.

Thus, a simple check of pixel labels is performed to reject invalid corners. The aim is to check if contiguous pixel with labels either 1 or 2 is observed. This is performed by process of elimination. The following pair of pixels is checked. If both of elements of any of pairs are zero then it cannot be a corner. If all the tests are positive we need to check if we obtain a continuous sequence of 1 or 2. If we obtain a continuous sequence, then corner is detected. This will provide a list of corners. Typically, a corner filtering operations is performed after corner detection. The aim is to retain the significant corner which as spaced at minimum distance from each other to avoid clustering of corners at same location. In the present implementation, we have no measure to indicate a strong or weak corner. In the earlier corner detector, the eigenvalues of matrix indicated strength of corner pixels. In generic terms, we can say that a corner response function is required for evaluate the strength of their corner. Thus, whenever a corner is detected, we need to compute a measure which indicates the strength of the corner. The response function used in the present application is total of the absolute difference of pixels in a contiguous arc and middle pixel.

$$R(x, y) = \sum P \in arc|I(p) - I(x, y)| \tag{9}$$

The algorithmic rule starts with real set '*S*' as the root node. On every iterations of the algorithm, it iterates through each unused attribute of the set '*S*' and calculates entropy 'H(*S*)' (or information gain 'IG(*S*)') of that attribute. It will then select the feature which has lowest entropy (or largest information gain) value. Set '*S*' is then partitioned by the particular feature to yield subset of data. Algorithmic rule continues to persist on each and every subset by consideration of only the attributes which were never chosen earlier.

Entropy 'H(*S*)' is a measure of the amount of uncertainty in dataset '*S*'.

$$H(S) = \sum\nolimits_{x \in X} -p(x) \log_2 p(x) \tag{10}$$

where

S Current data set for which entropy is being calculated that varies at each and every step of algorithm,
X Set of classes in '*S*'.
$p(x)$ Proportion of the number of elements in class '*X*' to the number of elements in set '*S*'.

When H(*S*) = 0, the set '*S*' is said to be perfectly classified (i.e., all the elements in '*S*' are belonging to same class).

In algorithm, entropy will be calculated for each remaining feature. The attribute or feature with lowest entropy is used to partition the set '*S*' on this iteration. A stable quantity has Entropy = 0, since its distribution is known perfectly. In contrast, the uniformly distributed random variable maximizes entropy. So, at this stage of tree, greater the entropy at node, the less data known to have the classification of data; and hence, greater the ability to improve the classification.

The information gain 'IG(*S*, *A*)' is a measure of difference in entropy from the before to after the set '*S*' is partitioned onto an attribute '*A*'.

$$IG(S, A) = H(S) - \sum\nolimits_{t \in T} p(t) H(t) = H(S) - H(S|A) \tag{11}$$

where

'H(*S*)' is entropy of set '*S*',
'*T*' is subsets produced from partitioning set '*S*' by attribute '*A*',
'$p(t)$' is proportion of the no. of element in subset '*t*' to the no. of element in set '*S*', and
'H(*t*)' is entropy of subset '*t*'.

In algorithm, we can calculate information gain (instead of entropy) for each of the remaining attribute. The feature/attribute with highest information gain 'IG(*S*, *A*)' is used to partition set '*S*' on this iteration.

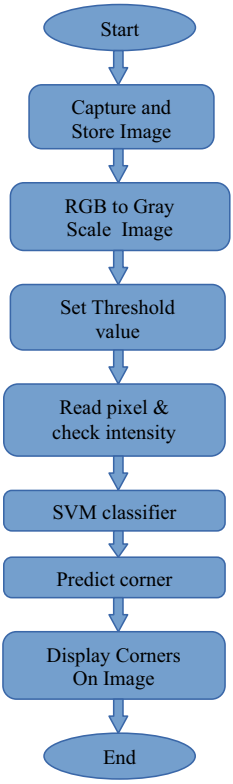


Fig. 3 Flowchart for SVM image corner detection

3 SVM Corner Detection Flowchart

The work flow of our support vector machine-based image corner detection algorithm is shown in Fig. 3.

4 Results and Discussion

A special attention is paid to the features of the SVM which provides a better accuracy of classification and nonlinear filters based on SVM is also possible [17, 18]. The traditional edge detection techniques are discussed in [19]. The image corners detected using the SVM algorithm and other corner detectors are shown below for some of the following biomedical images. In order to enhance and extract useful information from MR-Angiogram or scanned images and any such biomedical images, image processing can be applied on these images. Following are few applications where image processing is applied to given input images. Different corner detection techniques are applied on these images and results are shown in Figs. 4 and 5.

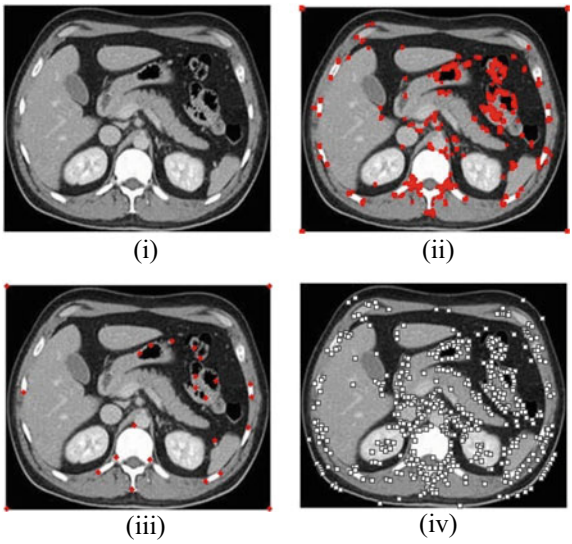


Fig. 4 Abdominal CT-scan **i** input image, **ii** Harris corner detection, **iii** Shi-Thomasi corner detection, **iv** SVM corner detection

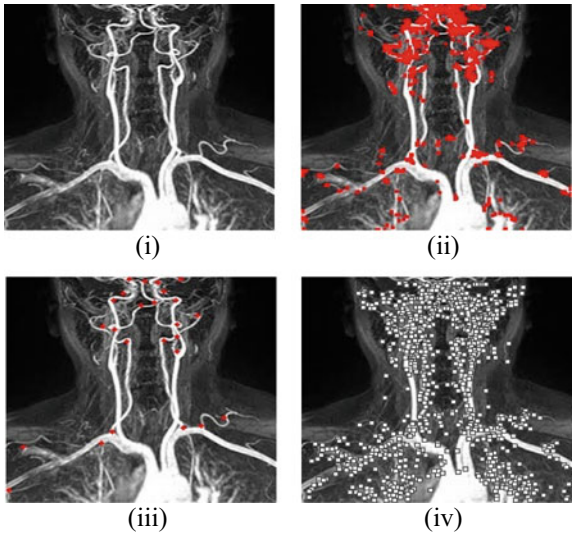


Fig. 5 MR-angiogram **i** input image, **ii** Harris corner detection, **iii** Shi-Thomasi corner detection, **iv** SVM corner detection

A. *Abdominal CT-scan*

B. *MR-Angiogram*

In Figs. 4 and 5, for the input image (i), the respective corner detections such as Harris corner detection, Shi-Thomasi corner detection and SVM corner detection are depicted as (ii), (iii) and (iv), respectively. We can observe the improvement

Table 1 Performance analysis of detected corner count for abdominal CT-scan

Method	Total corner count
Harris corner detection	86
Good features corner detection	29
SVM corner detection	335

Table 2 Performance analysis of detected corner count for MR-angiogram

Method	Total corner count
Harris corner detection	122
Good features corner detection	29
SVM corner detection	281

in total number of corners detected through machine learning-based SVM corner detection.

Tables 1 and 2 show that the performance of proposed SVM corner detection approach is robust in detecting corners with large number of detected corner count in both abdominal CT-scan and MR-angiogram, respectively, performed in OpenCV simulation tool.

A deep learning-based method for moving objects detection and tracking is also possible as presented in [20].

5 Conclusion

An algorithm of SVM intended for image corner detection is presented in this research paper. It is fully on the basis of computation of path of maximum gray-level modification meant for all border pixels and after that, representing the border-pixel by 4D feature vector comprised by amount of the pixels in the window centered with respect to this which has all feasible four paths of the maximum gray-level modification. SVM is developed by means of this feature and support vectors. Computing of complex differential geometric operators is not involved in our developed algorithmic rule. Developed algorithm has internal learning ability. Further, it can be refined making use of some different nonlinear SVMs and additional complex feature vectors.

Acknowledgements The authors wish to thank the anonymous reviewers for their feedback.

References

1. Richert W, Coelho LP (2013) Building machine learning systems with Python. Packt Publishing, p 7. www.packtpub.com
2. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC (2014) Machine learning in medical imaging. IEEE Signal Process Mag
3. Easton RL Jr (2010) Fundamentals of digital image processing, pp 1–4
4. Jiao N, Kang W, Xiang Y, You H (2017) A novel and fast corner detection method for sar imagery. Int Arch Photogram Remote Sens Spatial Inf Sci XLII-2/W7:605–608
5. Taheri M (2017) Enhanced breast cancer classification with automatic thresholding using support vector machine and Harris corner detection. Theses Dissertations 1718:8–15
6. Shah PM (2012) Face detection from images using support vector machine. Master's projects, vol 321
7. Biological and medical physics, biomedical engineering, Springer
8. Dubal S, Yadav M, Singh V, Uniyal V, Singh M (2016) Smart aero-amphibian surveillance system. In: International conference & workshop on electronics & telecommunication engineering (ICWET 2016)
9. Harris C, Stephens M (1988) A combined corner and edge detector, Plessey Research Roke Manor, United Kingdom © The Plessey Company plc
10. Shi J, Tomasi C (1994) Good features to track. In: IEEE conference on computer vision and pattern recognition (CVPR94) Seattle
11. Sneha PK (2016) Identification of leaf by using Canny edge detection and SVM classifier. IJAREEIE 5(4):3360–3363
12. Kim HI, Shin S, Wang W, Jeon SI (2013) SVM-based Harris Corner detection for breast mammogram image normal/abnormal classification, RACS'13
13. Shinde AS, Desai VV (2018) Relative investigation of machine learning algorithms for performance analysis on brain MR images. In: 8th international conference on advances in computing and communication (ICACC-2018), sciencedirect procedia computer science, vol 143, Elsevier B. V., pp 552–562
14. Kaur M, Kaur G (2015) Detection and categorization of objects using svm classifier and canny edge detection. Int J Electric Electron Eng 2(2):1–4
15. Herur SM, Kerur SS (2018) A technical review on image corner detection in machine vision. Int J Pure Appl Math 120(6):757–765
16. Bhat N, Eranna U, Singh MK (2018) Robust global gradient thresholds estimation in anisotropic diffusion for image restoration using DE, © Springer Nature Singapore Pte Ltd., pp. 677–691
17. Latha BM, Veena KN, Usha Devi MB (2017) A novel biometric identification system using DWT-HJDLBP feature extraction and SVM classifier. J Adv Res Dyn Control Syst 9(Sp-17)
18. Marquez DA, Paredes JL, Garcia-Gabin W (2007) Nonlinear filters based on support vector machines. In: Conference paper in acoustics, speech, and signal processing
19. Pooja AS, Smitha Vas P (2018) Edge detection using deep learning. Int Res J Eng Technol 5(7):1247–1252
20. Ye DH, Li J, Chen Q, Wachs J, Bouman C (2018) Deep learning for moving object detection and tracking from a single camera in unmanned aerial vehicles (UAVs). Imag Multimed Anal Web Mobile World

An Implementation of Nine-Level Hybrid Cascade Multi-level Inverter (HCMLI) Using IPD-Topology for Harmonic Reduction

Vishal Rathore, K. B. Yadav, and Spandan Dhamudia

Abstract Nowadays, multi-level inverters (MLI) with different topologies are used for various industrial applications. For average and large power applications, three commonly used topologies of MLI are flying capacitor, diode clamped and cascaded. This chapter implements the nine-level asymmetric cascaded multi-level inverter with IM for various kinds of modulation techniques in MATLAB/ Simulink. Increase in number of inverter levels, the response has more number of staircase steps available to achieve the required wave shape. As more number of steps is added to the output, waveform will lead to decrease in harmonics with the increase in levels, also the voltage across the devices connected in cascade increases.

Keywords Multi-level inverter · PWM · Topologies · Induction motor

1 Introduction

In past decades, MLI is used in various applications of large power switching as in traction motors, VAR compensators, HVDC transmission, renewable energy systems and HV laboratories [1].

The major benefit of using MLI over its two-level counterparts is its stepped output voltage which results in lowering the requirement of a transformer at the voltage distribution level result of which the cost of transmission system employed for large power practical applications. The common topologies of MLI are classified as capacitor clamped, diode camped and cascaded [2]. Cascaded MLI is more common

V. Rathore (✉) · K. B. Yadav · S. Dhamudia
Department of Electrical Engineering, NIT Jamshedpur, Jamshedpur, India
e-mail: 2018rsee009@nitjsr.ac.in

K. B. Yadav
e-mail: kbyadav.ee@nitjsr.ac.in

S. Dhamudia
e-mail: spandandhamudia@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_20

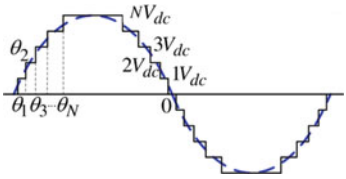


Fig. 1 Typical N-level staircase waveform of HCMLI

among the various type of topologies because of its flexible structure and higher output voltage which results in higher power levels and large reliability. Cascaded MLI uses the number of 1-φ bridge inverters that have ability of providing voltages at medium levels by using lower voltage components [3–7]. Generally, there must be four to nine inverters that are connected in cascade to achieve the required output voltage. Same topology and control technique has been used for each unit of converters used in inverter. If a fault occurs in any one of the units, it can be easily and quickly replace [8, 9]. Moreover, the faulty unit without disturbing the continuous availability of the load can be removing with a suitable control technique.

Generally, MLI uses voltages from a various DC supply. This supply can be separated in HCMLI structures or by connecting internally in clamped diode structure [10–14]. In various MLI, the DC supply used in the circuits to balance the voltage levels of the supply. The desired staircase waveform can be obtained with an appropriate switching based on the voltage levels such a typical N-level stair shape wave with different switching angles is shown in Fig. 1.

2 Proposed Methodology

The proposed work is implemented with the nine-level asymmetric HCMLI with an induction motor for IPD-PWM techniques in Simulink. The efficiency of HCMLI structure increases in terms of total harmonic distortion (THD) that is shown in later section of the proposed work. The response of HCMLI consists of large number of stairs that produce a stair shape wave that moves toward the required sinusoidal waveform. As large number of steps are added to output waveform will lead to decrease in harmonics also voltage across the cascade devices increase as the number of levels [15–17].

A. *Block Diagram of Proposed Work*

Basic block diagram of the proposed work is shown in Fig. 2 which comprises of three DC sources of 100, 200 and 100 V. By increasing levels of the voltage at the output required number of DC sources may be reduced [18]. Twelve semiconductor switches are used for inverter configuration. Generally, IGBT is used for switching purpose because of high power carrying capability, less switching and conduction losses. Finally, single phase asynchronous split phase induction motor (ASPIM) is used as a load [11, 19].

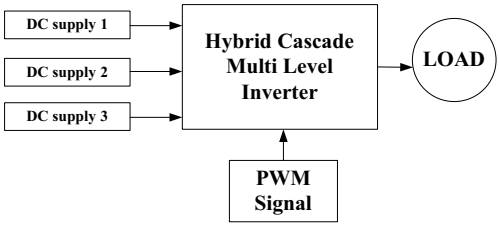


Fig. 2 Block diagram of proposed nine-levels HCMLI

3 Schematic Description of the System

B. *Simulation Diagram*

The simulation model of nine-level HCMLI is shown in Fig. 3 and is simulated by using MATLAB/ Simulink environment. It is basically an H-bridge cascaded asymmetrical MLI.

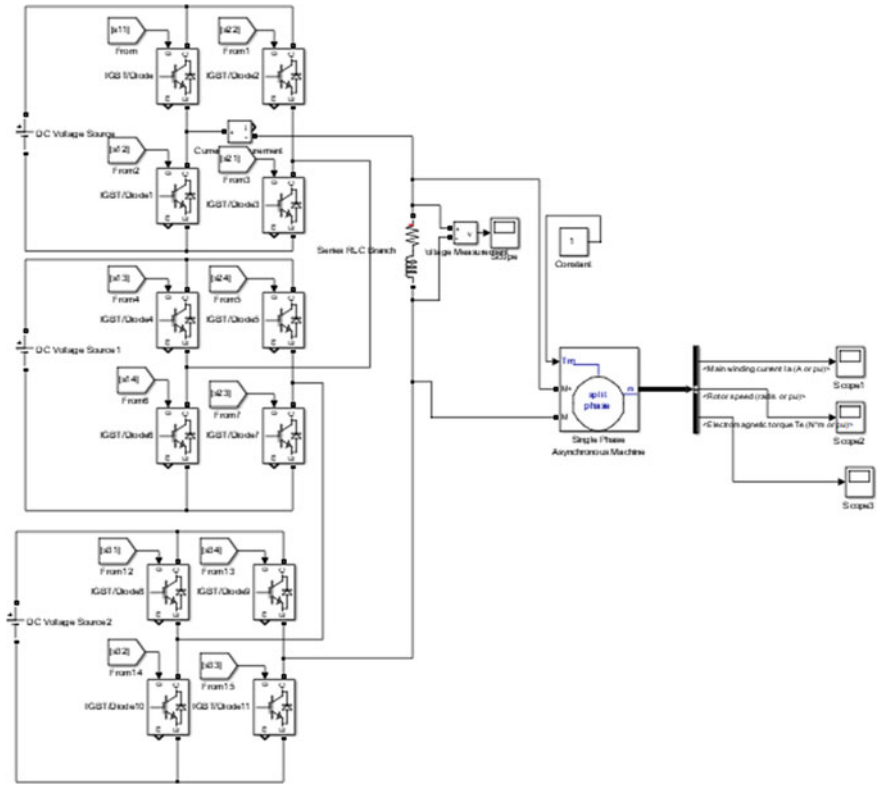


Fig. 3 Simulink model of nine-level HCMLI with IM

Table 1 Switching table for nine-level asymmetrical HCMLI

Voltages	0	100	−100	200	−200	300	−300	400	−400
s_1	0	1	0	1	0	1	0	1	0
s_2	0	1	0	0	1	1	0	1	0
s_3	0	0	1	0	0	0	1	0	1
s_4	0	0	1	1	1	0	1	0	1
s_5	0	0	0	1	0	1	0	1	0
s_6	0	1	1	1	0	1	0	1	0
s_7	0	0	0	0	1	0	1	0	1
s_8	0	1	1	0	1	0	1	0	1
s_9	0	0	0	0	0	0	0	1	0
s_{10}	0	1	1	1	1	1	1	1	0
s_{11}	0	0	0	0	0	0	0	0	1
s_{12}	0	1	1	1	1	1	1	0	1

Here, asymmetrical DC supply used is as an input to the inverter, and asymmetrical source defines that it has different values of DC sources used in an inverter. For a nine-level HCMLI, here, DC voltages are 100, 200 and 100 V.

C. *Working and Analysis*

The working of this MLI is depending on the switching of semiconductor switches (IGBTs) ON and OFF as to obtain the required voltage. The switching is done in such a ways to obtain staircase output voltage which is nearly equal to sinusoidal wave. For different angles of switching, the circuit behaves differently producing different output phase voltage waveforms. In this topology, we have generated nine voltage levels as 0, 100, 200, 300 and 400 V. The circuit working for each level is described in Fig. 4

The detail operation of new topology can also be understand by analyzing Table 1. Here, 0 means switches are OFF and 1 means switches are ON.

4 Simulation Result

This section shows the simulation result of output voltage of nine-level asymmetrical hybrid cascaded multi-level inverters (HCMLI) using IPD-topology and THD profile of a required voltage. HCMLI with the proposed topology results in lowest THD profile without the use of any type of combinational circuit of inductor and capacitor for smooth current waveform and without the use of any type of filter. The proposed topology has less number of switching devices due to which the cost of firing circuit reduces and lesser switching increases the performances of the circuit with reduce harmonics. Figs. 5 and Fig. 6 shows the output phase voltage waveform and their

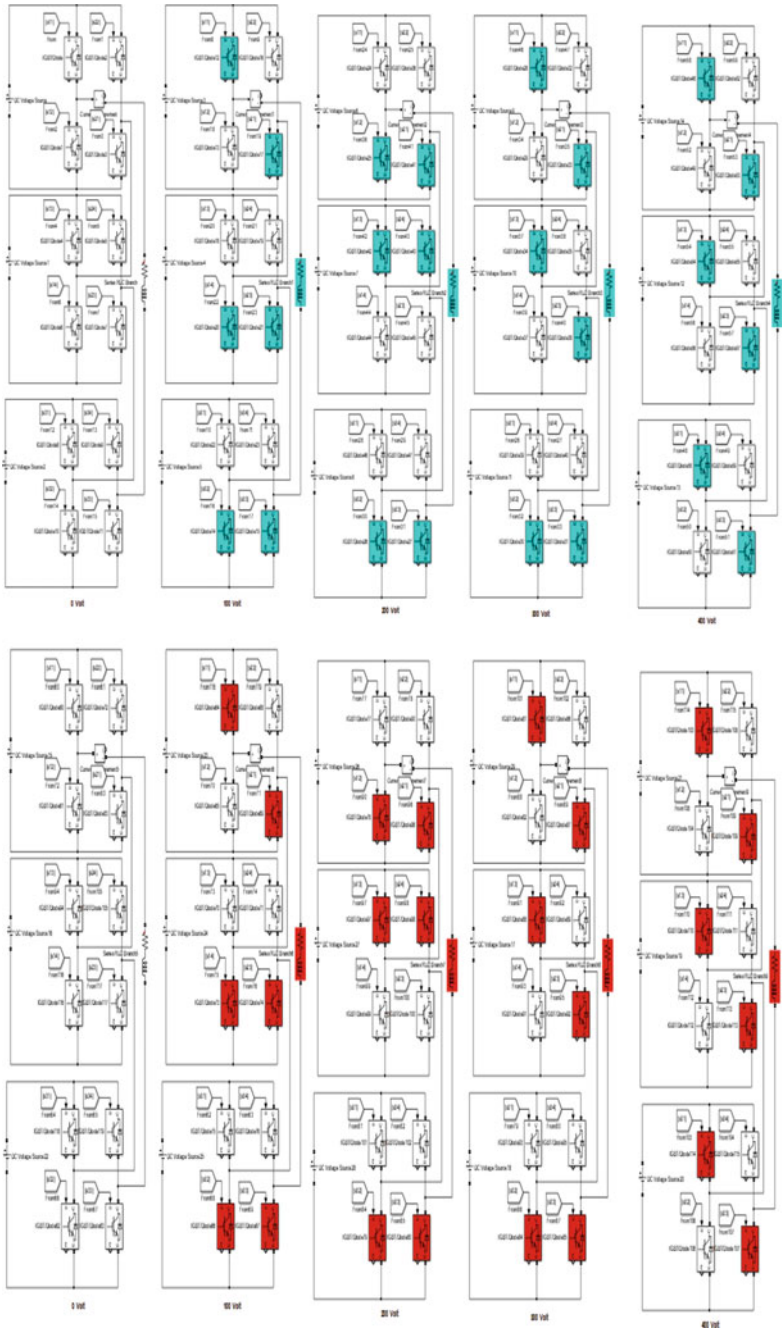


Fig. 4 Different stages of output voltage level

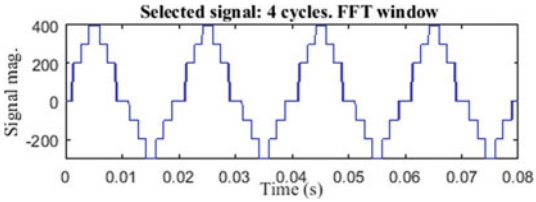


Fig. 5 Waveform of output phase voltage for asymmetric HCMLI using IPD-topology

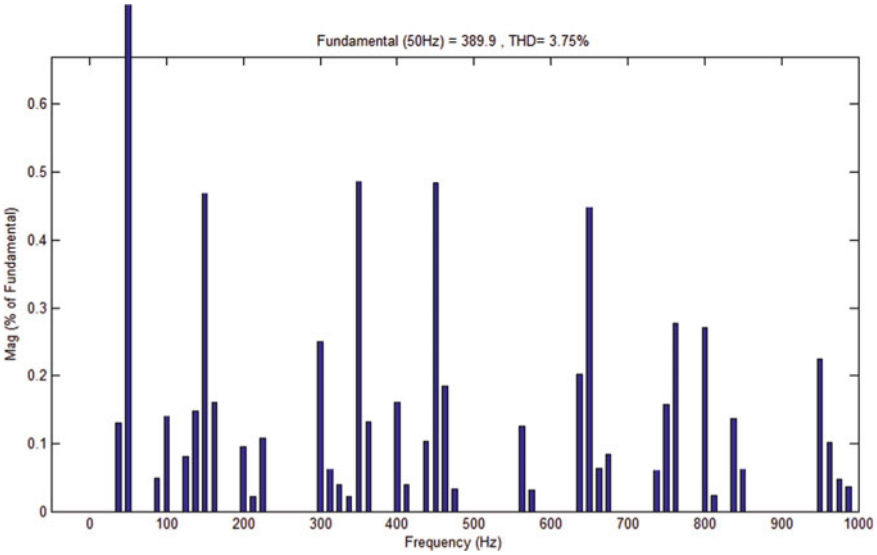


Fig. 6 FFT analysis of output voltage of asymmetric HCMLI using IPD-topology

harmonic content using IPD-topology and Fig.7 shows the output of induction motor using nine-level HCMLI. Table 2 shows the parameters of nine level inverter.

5 Conclusion

In this work, the multicarrier PWM-based modulation techniques for a nine-level HCMLI have been presented. Performance factor like THD of the required voltage of hybrid cascade multi-level inverter (HCMLI) have been presented and analyzed. The THD present in the response of asymmetrical HCMLI is studied by using IPD-technique. By applying IPD-PWM switching, the quality of the output is increased with reduce harmonics so the proposed topology of HCMLI is only adopted for harmonic reduction applications. The various advantages of HCMLI make it compatible for high power application like in traction drives, VAR compensators, HVDC transmission, renewable energy systems and high voltage laboratories, etc.

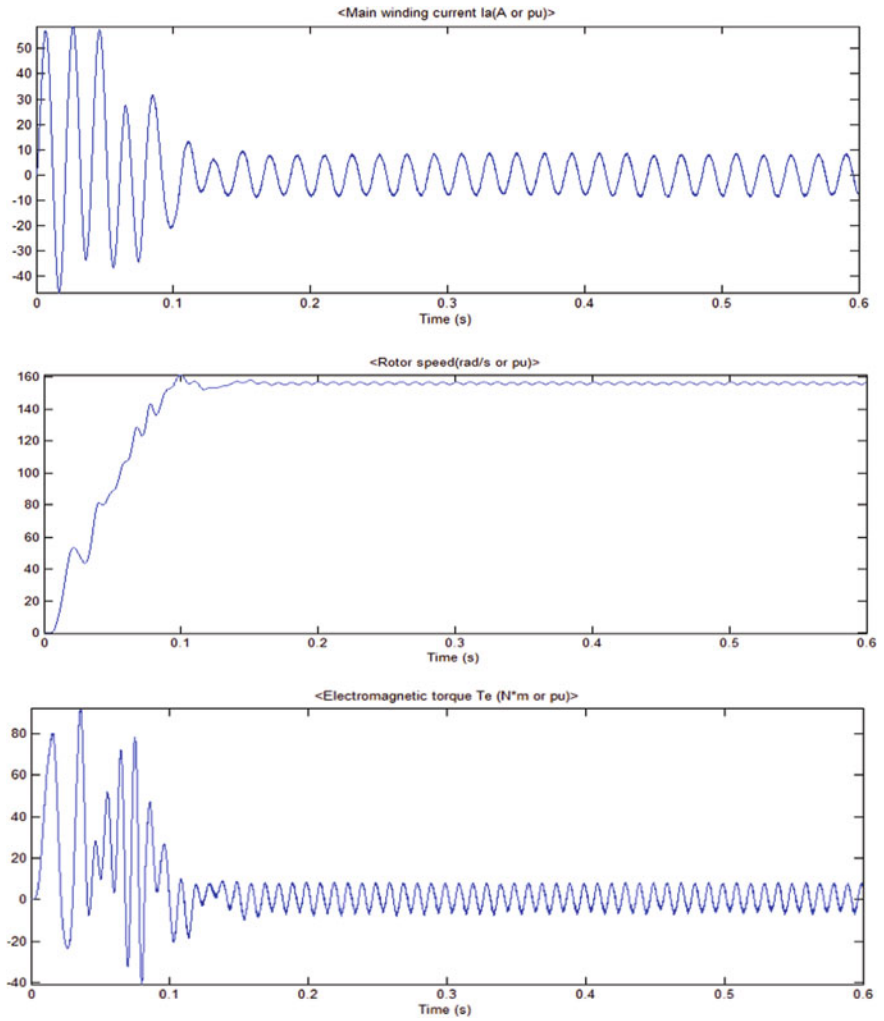


Fig. 7 Output of induction motor using nine-level HCMLI using IPD-topology

Table 2 System parameters for nine-level inverter

Parameters	Value
Frequency	50 Hz
Carrier frequency	3 K Hz
R_L	1 Ω
L_L	1 mH
DC supply	100, 200, 100 V

References

1. Wang J, Ahmadi D (2010) A precise and practical harmonic elimination method for multilevel inverters. IEEE Trans Ind Appl 46(2):857–865
2. Rodriguez J, Lai J-S, Peng FZ (2002) Multilevel inverters: a survey of topologies, controls, and applications. IEEE Trans Ind Electron 49(4):724–738
3. Malinowski M, Gopakumar K, Rodriguez K, Perez MA (2010) A survey on cascaded multilevel inverters. IEEE Trans Ind Electron 57(7):2197–2206
4. Wells JR, Nee BM, Chapman PL, Krein PT (2005) Selective harmonic control: a general problem formulation and selected solutions. IEEE Trans Power Electron 20(6):1337–1345
5. Kumar J, Das B, Agarwal P (2009) Harmonic reduction technique for a cascade multilevel inverter. Int J Recent Trends Eng 1(3)
6. Dahidah MSA, Agelidis VG (2008) Selective harmonic elimination PWM control for cascaded multilevel voltage source converters: a generalized formula. IEEE Trans Power Electron 23(4):1620–1630
7. Yousefpoor N, Fathi SH, Farokhnia N, Abyaneh HA (2012) THD minimization applied directly on the line-to-line voltage of multilevel inverters. IEEE Trans Ind Electron 59(1):373–380
8. Zhong Du, Tolbert LM, Chiasson JN (2006) Active harmonic elimination for multilevel converters. IEEE Trans Power Electron 21(2):459–469
9. Li Li, Czarkowski D, Liu Y, Pillay P (2000) Multilevel selective harmonic elimination PWM technique in series connected voltage inverters. IEEE Trans Ind Appl 36(1):160–170
10. Fei W, Ruan X, Bin Wu (2009) A generalized formulation of quarter-wave symmetry SHE-PWM problems for multilevel inverters. IEEE Trans Power Electron 24(7):1758–1766
11. Rathore V, Dubey M (2012) Speed control of asynchronous motor using space vector PWM technique. Int J Electric Eng Technol (IJEET) 3(3):222–233
12. Mahalakshmi R, Sindhu Thampatty KC (2015) Grid connected multilevel inverter for renewable energy applications, Elsevier, SMART GRID Technologies, Aug 6–8
13. Hemici K, Zegaoui A, Aissa Bokhtache A, Mahmoudi MO, Aillerie M (2015) Three-phases flying-capacitor multilevel inverter with proportional natural PWM control. In: International conference on technologies and materials for renewable energy, environment and sustainability, TMREES-15, Elsevier, energy procedia, vol 74
14. Abu-Rub H, Holtz J, Rodriguez J, Baoming G (2010) Medium-voltage multilevel converters—state of the art, challenges, and requirements in industrial applications. IEEE Trans Ind Electron 57(8)
15. Seyezhai R, Mathur BL, Jaibunisha A (2011) Design and implementation of silicon carbide based hybrid cascaded multilevel inverter using THIPDPWM technique. Eur J Sci Res 49(2): 225–233. ISSN 1450-216X
16. Mohan D, Kurub SB (2012) A comparative analysis of multi carrier SPWM control strategies using fifteen level cascaded H-bridge multilevel inverter. Int J Comput Appl (0975–8887) 41(21)
17. Gomathi C, Navyanagath, Purnima SV, Veerakumar S (2013) Comparsion of PWM methods for multi level inverter. IJAREEE 2(12)
18. Thombre N, Rawat RS, Rana P, Umashankar S (2014) A novel topology of multilevel inverter with reduced number of switches and DC sources. Int J Power Electron Drive Syst (IJPEDS) 5(1):56–62
19. Umashankar S, Sreedevi TS, Nithya VG, Vijayakumar D (2013) A new 7-level symmetric multilevel inverter with minimum number of switches. Hindawi Publishing Corporation, ISRN Electronics

Applying Soft Computing Techniques for Software Project Effort Estimation Modelling

Sudhir Sharma and Shripal Vijayvargiya

Abstract Software project development phase is a crucial one; under this, calculating the accurate effort and cost is challenging and tedious task, and for that, one has to undergo several hit and trial software practices. It is the need of every software organization to complete their software project development within the stipulated schedule and budget effectively; to accomplish this, traditional approaches are quite insufficient to model their current needs. Overestimation and underestimation of software development effort may cause financial implications in the form of cost of staffing, resources and the budget of the software project. Thus, there is always a constant need of a suitable and defined model of soft computing which can combined to estimate accurately and predict the effort or cost in time. In presented article, the soft computing approaches or machine learning (ML) approaches, e.g., linear regression (LR), support vector regression (SVR), multilayer perceptron or artificial neural networks (ANNs), ensemble learning viz. decision tree (DT), random forest, etc., are used and compared for evaluating the accurate effort estimation model, and various error evaluation metrics are computed and documented. For this, we have taken a prominent dataset into consideration which is Chinese dataset (499) standard project instances. After training and testing the datasets, it was obtained that the mean absolute error (MAE), magnitude of error relative to the estimate (MER), median of MRE (MdMRE), mean magnitude of relative error (MMRE) and PRED (25) for decision tree and random forest were significantly better than the existing listed techniques. Thus, these techniques could be taken into consideration for enhancing

S. Sharma (✉) · S. Vijayvargiya
Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Jaipur Campus, Jaipur 302017, India
e-mail: sudhirsharma@birlainstitute.co.in

S. Vijayvargiya
e-mail: shripalvijay@rediffmail.com

S. Sharma
Department of Computer Science and Engineering, Birla Institute of Applied Sciences, Bhimtal 263136, India

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_21

and improving the estimation accuracy of traditional effort estimation models while calculating effort, so that software organization meets their budget accordingly.

Keywords Soft computing · ANNs · Decision tree (DT) · Random forest · COCOMO · Software effort estimation

1 Introduction

In the software project development process, the project planning, scheduling and budgeting are the initial requirements that have to be taken in order to develop the software end product. Effort estimation plays an important and crucial role in IT and software industries to exactly determine how much resources are required to analyze the overall cost and budget. It has been seen that costing of PC equipment has fundamentally diminished in contrast with software, which is ceaselessly expanding [1]. On the other part, the key element for product development is the person-months (or effort in terms of person-months). The cost prediction models initially process the required effort to finish the software product undertaking that can be further evaluated over into cost viz. rupees. The present estimation models debilitate software team or developers by over-assessed spending plan or under-evaluated spending plan coming about into a total project disappointment and/or failure. Different models of effort estimation are often accessible in the marketplace. The effort of software project development is computed with the traditional COCOMO model in person-months. In the original version of COCOMO model, development effort is directly related to lines of code (loc) or software size [2]. In COCOMO, there are 15 effort multipliers which affect the software development effort and cost directly or indirectly. When we increase the value of effort multipliers, it causes subsequent reduction in development effort. Subsequently, if decreasing the value of effort multipliers, it shows increment in development effort; hence, resources can be consumed efficiently and effectively by estimating the software development effort precisely [3–5].

There are several cost estimation techniques proposed till date, and they are grouped into three major categories: (i) expert judgment, it is based on discussion with people who have the best hands-on live project experience and understand the project requirements; (ii) parametric models or algorithmic models, to predict software project cost or effort that uses mathematical formula which lies on the estimates of software size in terms of loc or function point, the no of employees, and additional process and product factors [6, 7]. By probing the costs and attributes of completed software projects, models can be developed using the formula which fits best, to replicate actual experience, and (iii) non-algorithmic or non-parametric models, viz., artificial neural networks (ANNs), evolutionary computation (EC) or genetic algorithms (GAs) and fuzzy logic (FL) [6, 8]. Alternatively, out of all software effort estimation approaches, constructive cost model (COCOMO) is widely known algorithmic model for ease of effort estimation for software project at different levels, in person-months [9]. Soft computing techniques can model multifaceted nonlinear

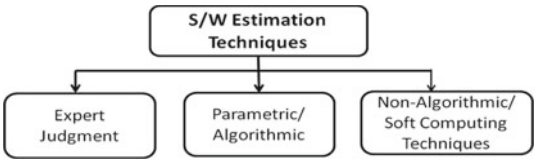


Fig. 1 Classification of software project effort estimation

relationships and approximate any quantifiable function. They are particularly useful in problems where there is a complex relationship [8, 10]. Many researchers have tried to investigate the possibilities of applying soft computing techniques for software effort estimation since, they are mainly useful in situations where there is an intricate association between input and output (as shown in the Fig. 1).

In this paper, we are applying soft computing, e.g., a bunch of machine learning techniques to compute the estimated effort and compare it with desired or actual effort. The traditional or algorithmic approaches are inadequate to handle dataset and their nature of dominance. Here, the five machine learning (ML) and one regression technique are used for estimating the effort for actual dataset viz., Chinese (499). These techniques have been applied in different fields over the years and show significant results with the minimum or no error. Further, to validate the computed, we have calculated various evaluation metrics for performance measure, i.e., MMRE, MdMRE, MAE, R^2 and PRED (25), which shows the error percentage, hence validating the accuracy of our estimation in comparison with actual effort.

Further, the paper is arranged into following six areas such as: Section 2 gives the review of the literature related to techniques of effort and cost estimation using soft computing techniques. Section 3 discusses various effort estimation and soft computing techniques. Section 4 describes*background and planning of research work*. Section 5 illustrates the dataset description and evaluation matrices for software cost estimation. Section 6 shows the experimental results and analysis. Section 7 summarizes the *conclusion* part of the research work along with future directions.

2 Literature Survey

The software development effort estimation-based survey presents information to demonstrate the significance of latest work done in the said area. It discusses significant procedures and activities related to software project planning and development mainly in term of manpower, effort or cost and time or schedule assessment. A number of effort estimation models have been proposed and created in the course of the most recent three decades. But, people are still using traditional or classical techniques like COCOMO model, Putnam SLIM, Halstead Function Point [11], etc., which are not adequate as well as relevant anymore according to the present IT and software industry benchmark and needs. Software cost is directly related to software quality and productivity. On the other hand, unrealistically low-cost estimates frequently lead to poor product quality, assessment and low project productivity [6].

The current literature showcases some selected and relevant articles/research papers that are explored and reviewed from the reputed journals and sources.

Putnam developed an early model known as SLIM, the Software Lifecycle Management [12]. Finnie et al. concluded that conclusiveness of ANNs depends largely on the dataset on which they are trained and is limited to suitable to dataset which is available [10]. Hearst et al. located the SVM calculation at the crossing point of learning hypothesis and practice: It contains a huge class of neural nets, radial basis function (RBF) nets and polynomial classifiers as extraordinary cases. However, it is sufficiently basic to be dissected numerically, in light of the fact that it very well may be appeared to compare to a direct strategy in a high-dimensional component space nonlinearly identified with input space [13].

The primary analyst Barry Boehm considered the cost estimation on a budgetary point of view and thought of a cost or effort estimation model, COCOMO 81 in the year 1981, in the wake of researching a huge arrangement of information in the 1970s. Precise effort estimation can give amazing help to planning choices in financial decision-making process in any software organization. COCOMO, SLIM and Albrecht's function point approaches assess the overall functionality of the system, where all are based on linear regression technique through gathering the data from the past projects as the major input to their models. Numerous algorithmic methods are deliberated as the most popular methods [12]. Huang et al. revealed in their examination a recently presented learning technique dependent on statistical learning hypothesis, support vector machines, together with a much of the time utilized elite strategy, back-spread neural systems, to the issue of credit score forecast. They utilized two dataset collections for Taiwan financial organizations and US business banks for test. The outcomes demonstrated that support vector machines gain precision similar to that of back-propagation neural systems [14].

I. Attarzadeh and Siew Hock Ow proposed a novel neuro-fuzzy constructive cost model on the NASA dataset, based on COCOMO II and fuzzy logic. Since their model had learning ability and good interpretability, therefore while maintaining the merits of the COCOMO model, it performed better than ordinary COCOMO II and they got results closer to actual effort [15]. C. S. Yadav et al. tuned the parameters of COCOMO II model to estimate the effort using GAs. Also, COCOMO II was modified by tuning additional parameters to estimate the software project effort more realistically on NASA projects datasets [2].

Malhotra and Kaur detailed that use case point analysis beats different models with the least weighted normal (MMRE) of 39.11%, contrasted with 90.38% for function point analysis and 284.61% for COCOMO model [16]. Goyal and Parashar explored a neural network model with different network architectures and parameters viz. SG1–SG6 out of which models SG5 prepared utilizing COCOMO dataset and tried for KEMERER dataset given 228.70 MMRE value and 317.65 SDMRE value which is better than recorded exactness of COCOMO that is 284.61 MMRE value. Consequently, they reasoned that no model is versatile to the most recent software advancement approaches, coming about into the undertaking disappointments [5].

3 Software Effort Estimation Approaches

This section elaborates the overview of effort estimation approaches and models with soft computing techniques, which will be further used in the research work. Although different approaches and standards are as of now accessible in the software industry, AI or machine learning is one of the developing methodologies right now [5]. The main model which is used to calculate the effort is the COCOMO model. Out of all effort/cost estimation approaches, the COCOMO is most commonly adopted traditional cost modeling technique, known for its truthfulness thus approximating the required person-months as an effort of software project at varied levels. Many researchers have tried to investigate ANNs for finding the effort [10, 15, 17]. ANNs can handle intricate nonlinear associations and estimated a few quantifiable functions. In contrast, there are number of such strategies accessible for assessing the product exertion and cost. Estimation procedure is for the most part distributed into two models viz. parametric and non-parametric models. The two models are produced for carrying out the exact estimation [18].

3.1 The Process of Software Effort and Cost Estimation

The software effort and cost estimation is the piece of software advancement life cycle, where the process of software development is split into number of phases or levels. The cost estimation relies on how we associate the guesstimates of effort and calendar time by the project behaviors. Alternatively, estimation is the activity of searching an approximate evaluation, which determines the amount of effort, time, resources and money required to develop a software project or product [12, 19]. Generally, the assessment or estimation is based on the following parameters:

- Past statistics or familiarity,
- Reachable facts, information and documents,
- Certain assumptions and
- Accepted threats.

The software project estimation is categorized into the following steps:

(i) Size evaluation of the product to be developed,
(ii) Estimation of effort in person-months or hours,
(iii) Time or duration estimation in calendar months,
(iv) Project cost estimation in accepted legal tender [10].

Estimation is not a one-time task; it is made according to requirement of the project. It can take place during the project beginning, development of the project. In addition, the goal of software project should be thoroughly understood before actual process starts. Thus, it might be helpful in historical project data [20].

Historical perspective and vital input for making quantitative approximation are provided by project metrics. To gain credibility, initial commitment about planning

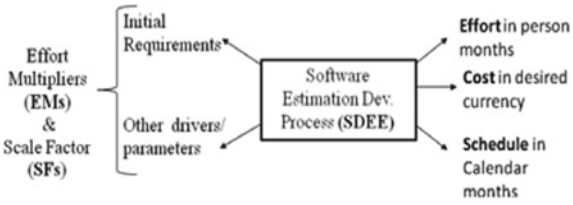


Fig. 2 Classical view of software development effort estimation process (SDEE)

is required by the software development team. They require at least two estimation techniques to reach to the final estimate and merge the significant values [10]. Figure 2 shows the classical view of software development effort estimation.

The best known extensively used formal estimation models are Boehm's COCOMO-I and II, function point analysis, Putnam's SLIM [12], expert judgment, analogy system, Parkinson's law and price to win strategy. Additionally, the software effort estimation process is way more typical than that displayed in Fig. 2. There is interdependency between many input parameters of effort estimation model viz. effort multipliers, scale factors and/or output, all of which are directly and indirectly relevant to the software estimation process.

3.2 Constructive Cost Model (COCOMO)

The constructive cost estimation model is widely known and written as COCOMO, an algorithmic effort and cost estimation model invented by Barry Boehm in 1981 [1]. Basically, this model computes the effort and duration for a software development project based on inputs related to size of the software and effort multipliers that influence its efficiency. COCOMO was developed from the analysis of *63* real-life software projects from the period of 1964–1979 by an *American Company TRW Systems Inc*. It follows a series of cost estimation models, viz., *basic*, *intermediate* and *detailed* sub-models, which reflects the complexity of the software project. Further, sub-models have three development modes defined as organic, semi-detached and embedded [6]. It was the most cited and credible cost estimation model in comparison with all the other conventional models. Moreover, COCOMO version II is the improved edition of the previous COCOMO version I and is tuned to the software development practices of the current needs [18].

At first, three development modes were characterized, and then, alignment was made utilizing the first 56 undertaking project databases to show signs of improvement. Barely, any more activities were added to the original database coming about into the well-known 63 undertaking project databases. COCOMO depends on experimentally determined connections among different cost drivers [12].

The basic COCOMO model equation in the form:

$$\text{Effort (E)} = \text{a. (SLOC or KDSI)}^{\text{b}} \text{ in } \left[\text{person} - \text{months}\right]$$

$$\text{Development Time (D)} = \text{c. (E)}^{\text{d}} \text{ in [calendar } - \text{ months]}$$

where SLOC or KDSI means the estimated number of delivered lines (in thousands) or Kilo delivered source of instructions of code for project and the coefficients a, b, c and d are equation constants that are dependent upon the three modes of software development projects. Cocomo II model uses the line of code as the size metrics and uses five scale factors and 17 effort multipliers for the final calculation of effort or cost. The five scale factors are rated on a *six-point scale from very low to extra high* (i.e., from 0 to 5 points) [1].

The cost drivers which fine-tune the preliminary estimates and generate multiplier in the post-architecture model are divided in the following sections [1]:

(a) Product Attributes: The attributes are concerned with required features for developing software product.
(b) Computer Attributes: These are the restriction enforced on the software by the hardware platform.
(c) Personal Attributes: the multipliers, which obtain into account the ability and familiarity of the persons employed in product development.
(d) Project Attributes: The features are related to precise attributes of product development.

4 Research Background and Planning

In this section, we are providing the brief detailing of what soft computing methods or machine learning techniques we have used to compute the software development effort estimation. There are total six techniques were used, in which one is regression and five are machine learning methods viz. support vector machines, artificial neural network (or logistic regression), decision tree, random forest, bagging and boosting, etc; each technique has its own advantages and disadvantages. Also, these methods have seen the wide interest and great scope of applicability over the years and can be applied in other area too. On the other hand, the logistic regression is a statistical model which uses a logistic function to model a binary dependent variable. In regression analysis, logistic regression estimates the parameters of a logistic model (binary regression) [21].

4.1 Linear and Logistic Regression

Regression or linear regression is a method of modeling a desired or expected value based on independent prediction. This method defines how one independent variable (X) is related with another dependent variable (Y) in terms of their relationship. Furthermore, regression equation discovers a appropriate line which reduces the addition of the squares of the perpendicular distances of the points from the certain

line. Mainly, it is the method finding the interdependence value between variable Y and variables X [5]. On the other hand, logistic regression is an additional technique which has the origin from machine learning and the field of statistics. But, when the dependent variable is dichotomous (binary), it works.

4.2 Multilayer Perceptron (MLP)

A MLP is a multilayered feed-forward ANN model that has one input layer, at least one hidden layer, and one yield or output layer. Every neuron of the input layer speaks to an info vector. In the event that a system is just made out of an input layer and a output layer (no hidden layer), at that point, the name of the system becomes perceptron. By and large, for a MLP, a nonlinear activation function is utilized in the neurons of the hidden layer. On the differentiation, a linear activation function is normally utilized in the output layer. The quantity of the neurons in the hidden layer differs depending on the quantity of input neurons and the sort of the training algorithm utilized. One of the well-known training algorithms is the back-propagation algorithm which is a sort of gradient decent learning algorithm [8, 15].

4.3 Support Vector Machine (SVM)

Support vector machines are probably one of the most extensively and well-known machine learning algorithms. It is an algorithm for classification problems similar to logistic regression (LR). The purpose of SVM algorithm is to find the hyperplane that has the highest margin in an N-dimensional space ('N' is number of features) that definitely organizes the data points. The purpose of SVM is to separate the data into different categories by building a hyperplane [15]. There are different classes available for SVM viz. linear kernel SVM, polynomial kernel SVM, radial kernel SVM, etc. Nonlinear restrictions can be extended using kernel trick. SVM support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. The margin of the classifier is maximized using these support vectors. The position of the hyperplane is changed by deleting the support vectors. These are the points that help us build our SVM. The dashed line in Fig. 3 shows the distance between separating line and closest vectors to the line [5]. In our research, we use SVM for continuous variable for estimating effort and cost.

4.4 Artificial Neural Networks (ANNs)

The structure of ANN comprises a system of interconnected nodes known as neurons or processing components. It consists of three layers, for example, the output layer,

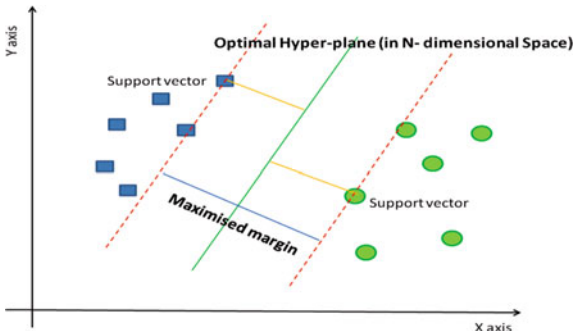


Fig. 3 *N*-dimensional hyperplane for support vector machine

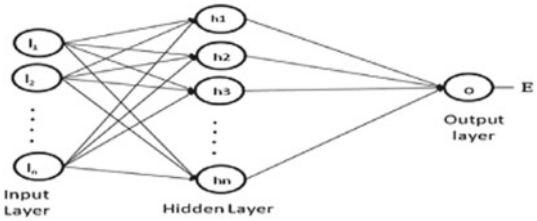


Fig. 4 Multilayer perceptron architecture (ANNs)

hidden layer and input layer. The first layer has input neurons, which passes information through connection called as weights to the second layer of neurons and afterward again through more weights to the last layer of output neurons. Extra multifaceted frameworks have more than one hidden layers [16]. ANN is a widely known algorithm for training or learning and known as error back-propagation algorithm. The error is calculated by subtracting predicted output from the desired output, and this error is again fed back to the configured network which minimizes and adapts the tunning of parameters in a desired manner. The whole procedure is iterated till the desired result is matched [15, 22] (Fig. 4) shows the architecture of ANNs.

4.5 Decision Tree (DT)

DTs are used for classification and regression; they are essential building blocks for random forest. They are required to analyze and forecast model through mislaid importance and categorical attributes. It is a data mining technique which continually splits the previous dataset by depth-first greedy technique or breadth-first technique till all data items fit in to a desired class. It has an architecture which consists of nodes (i.e., root, internal and leaf node) and arcs. The hierarchical formation organizes unidentified datasets. Using impurity measure technique, a decision of best split is reconstructed at every internal node of the tree [5, 15].

4.6 Bagging

Bagging is a part of ensemble machine learning or classification algorithm which is a very powerful ensemble learning method. Bootstrap aggregation is a general procedure which can be used to reduce the variance for those algorithms that has high variance as well as helps to avoid over-fitting. We can say that bagging is a bootstrap aggregation which repetitively sample (through substitution) from a dataset in accordance to a uniform probability distribution (where each sample has similar unique dataset) [21]. In view of the fact that the sampling is implemented with substitution strategy, some cases may emerge several times in the identical training set, whereas others might be replaced from the training set. Usually, a bootstrap aggregation dataset sample (D_i) holds roughly 63% of real training data. Since each dataset sample has a probability of $1 - (1 - {}^1/_N)^N$ of being chosen in every dataset sample D_i where N is satisfactorily large, then probability converges to $(1 - {}^1/_e) = 0.632$ [5].

4.7 Random Forest

Random forests are form of supervised learning algorithm. It tends to be utilized for both classification and regression. It is likewise the most adaptable and simple to utilize method. A forest is contained trees. It is said that the more trees it has, the more hearty a timberland is. Fundamentally, random forest or random decision forests belong to ensemble learning approach for classification, regression and different tasks that works by developing a huge number of choice trees at training time and yielding the class that is the mod of the classes or mean forecast of the individual trees [22]. It additionally gives a really decent marker of the feature significance and extraction. Random forests have an assortment of applications, for example, recommendation search engines, image classification and feature selection. It may be utilized to characterize worthy credit candidates, distinguish deceitful activities and predict diseases [2].

5 Dataset Description and Model Evaluation Metrics

5.1 Dataset Collection and Analysis

To perform the implementation, we need benchmark industry used dataset, which can be utilized for effort estimation modelling. These datasets contains raw data that comes in arff file format and this data is not suitable for application process; thus, one need to preprocess the entire dataset before using it for actual implementation. There

Table 1 List of promise repository datasets and their features or attributes [24]

Dataset	NASA	COCOMO81	MAXWELL	NASA93	CHINA
# Project instances	18	63	62	93	499
#Attributes/features	3	17	27	23	19

are many software repositories out of which Promise and ISBSG dataset repositories contain an abundance of data for effort estimation and prediction [2, 23]. In this research study, several datasets such as COCOMO81, COCOMO_NASA2 (NASA93), and CHINA and NASA are analyzed and studied. Each dataset has varying number of attributes and features which are different in nature. The dataset which is named as COCOMO81 is taken only for assessment purposes (Table 1) tabulates the datasets description of promise repository.

The China dataset has total 19 features or attributes; out of which one is dependent and rest are independent attributes. A few of the independent attributes were discarded since they were not relevant in effort prediction, consequently making the model to a great extent simpler and proficient. Furthermore, we have used cross-validation; leave-one-out (LOO) technique, which is avilable in the MATLAB to decrease the number of independent variables. After applying feature selection or pca, all 19 attributes minimized into 10 attributes (i.e., one dependent and nine independent attributes). The dependent attribute is effort and independent attributes are Enquiry, Interface, Output, Added, PDR_AFP, PDR_UFP, NPDR_AFP, NPDU_UFP and Resource. Every independent attribute corresponds to FP method [1, 5].

In this research article, we have performed experimental/test by keeping dominant independent variables viz. our test holds nine independent and one dependent variables of China dataset.

5.2 Model Evaluation Matrices

Here, we have taken a number of model evaluation matrices or performance indices to assess the estimation accuracy and capabilities. Among all, the most universal used performance criterion is MMRE and PRED (l). They will be applied to measure and compare the results produced by ML techniques.

(a) *Mean magnitude of relative error (MMRE)*
It is the mean of all relative error from 1 to n.

$$\mathrm{MRE} = \frac{\mathrm{act}_i - \mathrm{est}_i}{\mathrm{act}_i}$$

$$\mathrm{MMRE}(\%) = \frac{1}{n}\sum \mathrm{MRE} * 100$$

where the act_i is the *actual effort*, est_i is the *estimated effort* and '*N*' is the *number of cases* [19].

(b) *Prediction level—Pred* (*l*)
Prediction level is calculated through relative error and represented as the proportion of data points with error MRE less then equal to level l to the total no. of data points. Thus, it is the accuracy measure not the error measure and higher the value of Pred(l), the better the accuracy score [15, 25]. In our results, we have considered pred (25) for accuracy measure. Moreover, the other evaluation matrices are also used for getting the more precise and unique results.

$$\text{Pred}(l) = d/n$$

where '*d*' is the no. of projects instance whose value of MRE is $\leq$ l and '*n*' is the total project instances taken for implementation.

6 Experimental Result and Analysis

This section presents the research techniques applied to obtain the experimental results and discusses the model results. The analysis undertaken in this paper is the dataset used in experiment which is from COCOMO Chinese database, and these datasets are publicly available, which consist of 499 (dataset-I) project instances. The datasets are required to calculate approximately the accuracy of software estimation. We have divided the whole datasets into two parts i.e., training and testing or validation part in 8:2. That is, 80% dataset is used for train the system and the rest (or 20%) of the dataset is used to validating the accuracy or prediction of the model. Altogether, six ML methods are applied out of which five belongs to soft computing and one technique is based on regression to estimate the results. Throughout the research text, we have used soft computing and ML techniques interchangeably since both the techniques are part of each other.

6.1 Result Outcome

In dataset which has total 499 project instances or PIDs split into the training set consists of 399 project instances and test/validation set consist of 100 project instances selected randomly. The effort is calculated and implemented on Python/MATLAB software tool. Table 2 summarizes the test and training results for randomly selected project instance ids (PIDs):

The error estimation measure like MMRE, MER, MAE and RMSE having lower values is quite better result in comparison with the previous results shown in the comparison taxonomy. On the other hand, high correlation coefficient and Pred (25) are taken to be finest in comparison with all other models [5, 18]. According to our

Table 2 Effort estimation results for various ML techniques

Error evaluation metrics	SVR	Decision tree (DT)	Linear regression	Bagging	Logistic regression	Random forest
Mean magnitude of relative error (**MMRE**) %	**60.22**	**0.39**	**49.02**	**57.97**	**15.37**	**15.78**
Mean absolute error (**MAE**) %	**5.70**	0	**3.85**	**4.04**	**1.24**	**0.98**
Root mean squared error (**RMSE**) %	7.91	0.01	7.47	8.21	3.55	1.82
Magnitude of error relative (**MER**) %	**54.44**	**0.37**	**72.14**	**48.37**	**14.30**	**13.82**
Median of MRE (**MdMRE**) %	**38.94**	**0.00**	**29.60**	**42.47**	**8.85**	**9.13**
Correlation coefficient ($\mathbf{R^2}$)	0.87	1.00	0.82	1.00	0.98	1.00
Prediction level—**PRED (25)** %	**43**	**100**	**46**	**30**	**92**	**84**

table, the results of random forest and logistic regression are found to be best with the corresponding MMRE values which are 15.37 and 15.78%, MAE values are 0.98 and 1.24% and PRED(25) values are 92 and 84%, respectively. Hence, these two methods are significantly effective in estimating the software effort accurately. Also, the results of bagging technique are quite competitive with the random forest and logistic regression (Figs. 5, 6 and 7).

The above plots shows the estimation results for effort and MRE. From the plots, it is clearly visible that for DT, random forest and logistic regression, the predicted results are quite significant in comparison with other applied ML techniques. During

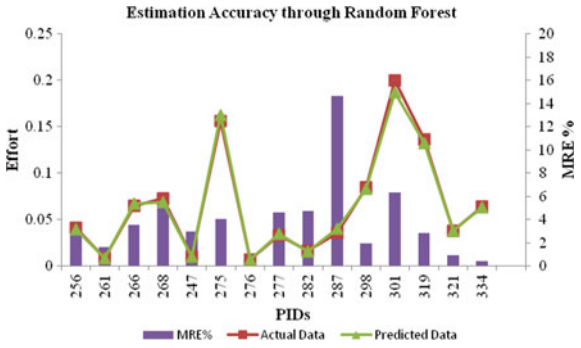


Fig. 5 Estimation plot for random forest

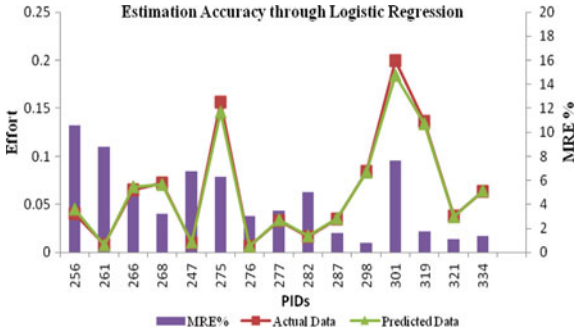


Fig. 6 Estimation plot for logistic regression

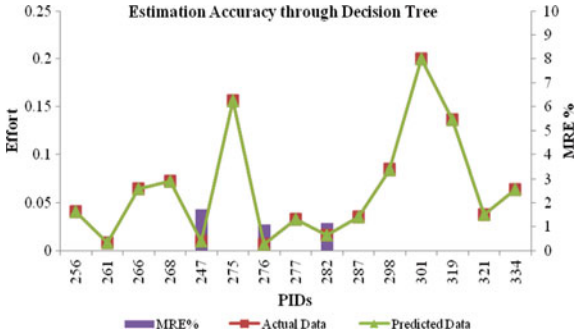


Fig. 7 Estimation plot for decision tree

train/test spilt, we divided our dataset into 400 and 100 PIDs, respectively (i.e., 80:20 ratio) and performed the ML techniques; the plots also show that the estimated values are very close to actual one.

6.2 Comparison Taxonomy with the Previous Models

The predicted results with our effort estimation model using decision tree, random forest and logistic regression methods were performed well in comparison with additional six approaches applied for the research. We have compared our results (using ML techniques on benchmark datasets) with the previous models. These previous models tabulated in comparison taxonomy have used the same dependent variable, but the independent variables vary from model-to model-implementation. From Table 3, it is clear that the logistic regression, decision tree and bagging techniques have outperformed the MMRE as well as PRED values of previous models tabulated in comparison taxonomy. The outcome of MdMRE is also satisfactory and lower as compared to the majority of the previous estimation models.

Table 3 Comparison taxonomy analysis for various ML techniques on benchmark dataset

Research publication (Reference No.)	Dataset used	Name of the ML technique used	PRED (25)%	MMRE %	MdMRE %
Proposed results	China (Chinese industrial dataset based on 499 completed project instances)	Linear regression	0.46	49.02	29.60
		Random forest	84	15.78	9.13
		Decision tree	100	0.39	0
		Bagging	30	57.97	42.47
		SVR	43	60.22	38.94
		Logistic regression	92	15.37	8.85
[10]	China (499)	Bagging	34.66	74.23	–
		Linear regression	36	17.97	–
		SVM	38.66	25.63	–
		ANN	11.33	143.79	–
		M5Rules	52	41,442	–
[26]	China (499)	ANN	22	90	–
		CART	26	77	–
		LSR (Least square regression)	33	72	–
		AABE (Euclidean distance)	57	38	–
		AABE (Minkowksi distance)	61	43	–
[20]	China (499)	Regressive	–	62.3	–
		ANN	–	35.2	–
		CBR	–	36.2	
[17]	China (499)	MART (Multiple additive regression tree)	88.89	8.97	–
		RBF	72.22	19.07	–
		SVR_Linear	88.89	17.04	–
		SVR_RBF	83.33	17.8	–
		Linear regression	72.22	23.3	–
[5]	China (499)	Additive regression (M5Rules)	80.59	16.83	–
		SVM for regression	86.47	41,502	–
[27]	China (499)	Linear regression	72.22	23.3	–
		RBF	72.22	19.07	–

7 Conclusion

In this research paper, we have applied six major ML techniques and their result in comparison with previous models or appaoches tabulated in comparison taxonomy. The dataset that we have utilised for getting the results were from Promise dataset repository. Further, we processed the datasets and the 19 attributes of Chinese dataset were reduced to 10 features in the experiment/test. The results show that the decision tree, logistic regression and random forest methods are the best in out of all six ML techniques or methods for predicting effort with MMRE values 15.37 and 15.78, respectively, along with the prediction level PRED(25) values significantly better and higher as compared to all the previously employed techniques presented in the literature. Hence, machine learning methods chosen are relatively suitable for finding the more accurate effort estimation results as compared to conventional or traditional approaches as well as with other ML techniques where initial test bed is not configured properly to estimate the accurate effort and cost estimation. Moreover, these techniques in combination with evolutionary and fuzzy techniques can reduce vagueness and impreciseness of datasets used for the implementation of model. Also, one can view this effort estimation problem as a classification problem by dividing whole effort into separate classes and later decide in which effort class the results are approaching or falling.

Acknowledgements I would like to express my appreciation toward the Director, Birla Institute of Applied Sciences, Bhimtal for providing the financial assistance and encouragement for the planning and development of this research work. Also, I would like to thank Dr. Abhinav Dinesh (Director, B.I.T., Ext. Center, Jaipur) for providing me the support and research facilities at B.I.T, Jaipur campus.

References

1. Albert A, Gaffney JE (1983) Software function source lines of code and development effort prediction: a software science validation. IEEE Trans Softw Eng 9:639–648
2. Yadav CS, Singh R (2014) Tuning of Cocomo II model parameters for estimating software development effort using GA for promise project data set. Int J Comput Appl 90(1):37–43
3. Subramanian GH, Pendharkar PC, Wallace M (2006) An empirical study of the effect of complexity, platform, and program type on software development effort of business applications. Emp Softw Eng 11:541–553
4. Chiu NH, Huang SJ (2007) The adjusted analogy-based software effort estimation based on similarity distances. J Softw 80:628–640
5. Goyal S, Parashar S (2018) Machine learning application to improve COCOMO model using neural networks. IJITCS 3:35–51
6. Srinivasan K, Fisher D (1995) Machine learning approaches to estimating software development effort. IEEE Trans Software Eng 21(2):126–137
7. Fedotova O, Teixeira L, Alvelos H (2013) Software effort estimation with multiple linear regression: review and practical application. J Inf Sci Eng 29:925–945
8. Haykin S (2000) Neural networks: a comprehensive foundation. Macmillan College Publishing Company, Boston

9. Jørgensen M, Boehm B (2008) Software development effort estimation: formal models or expert judgment. Int IEEE Softw
10. Finnie GR, Wittig GE (1997) A comparison of software effort estimation techniques: using function points with neural networks case-based reasoning and regression models. J Syst Softw 39:281–289
11. Moløkken K, Jørgensen M (2003) A review of surveys on software effort estimation. In: International symposium on empirical software engineering, pp 223–231 (2003)
12. Boehm BW (1994) Software engineering economics. Prentice Hall, Englewood Cliffs
13. Burgess CJ, Lefley M (2001) Can genetic programming improve software effort estimation? A comparative evaluation. Inf Softw Technol 43:863–873
14. Huang SJ, Chiu NH (2007) Applying fuzzy neural network to estimate software development effort. Springer Science Business Media, LLC, vol 30, pp 73–83
15. Attarzadeh I, Hockow S (2010) A novel Algorithmic cost estimation model based on soft computing technique. J Comput Sci 6(2):117–125
16. Malhotra R, Kaur A, Singh Y (2011) Application of machine learning methods for software effort prediction. In: Newsletter ACM SIGSOFT Software Engineering Notes, vol 35
17. Oliveira ALI (2006) Estimation of software effort with support vector regression. Neuro Comput 69: 1749–1753
18. Dave VS, Dutta K (2014) Neural network-based models for software effort estimation: a review. Artif Intell Rev 42(2):295–307
19. Reddy CS, Raju KVSVN (2009) A concise neural network model for estimating software effort. Int J Recent Trends Eng 1:188–193
20. Heiat A (2002) Comparison of artificial neural network and regression models for estimating software development effort. J Inf Softw Technol 44(15):911–922
21. Rajasekaran S, Vijayalakshmi Pai GA (2004) Neural networks, fuzzy logic, and genetic algorithms—synthesis and applications, PHI
22. Malhotra R, Jain A (2011) Software effort prediction using statistical & ML methods. IJACSA 2(1):145–152
23. Chiu NH, Huang SJ (2007) The adjusted analogy-based software effort estimation based on similarity distances. J Syst Softw 80:628–640
24. Promise Data repository. https://promisedata.org/repository/
25. Yurdakurban V, Erdo Gan N (2018) Comparison of machine learning methods for software project effort estimation. 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, pp 1–4
26. Smith RK, Hale JE, Parrish AS (2001) An empirical study using task assignment patterns to improve the accuracy of software effort estimation. IEEE Trans Software Eng 27:264–271
27. Gallego JJC, Rodriguez D, Sicilia MA, Rubio MG, Crespo AG (2007) Software project effort estimation based on multiple parametric models generated through data clustering. J Comput Sci Technol 22:371–378

Voice-Based Biometric Authentication

N. S. Wankhede

Abstract Voice or speaker recognition has become an important source of security nowadays. It has a wide scope in security, information services, industries and voice dialing. This implementation is to extract the speech sample, categorize it and find out where the person speaking is a registered user or not. For feature extraction, Mel-frequency cepstral coefficients (MFCC) has been used, and for feature matching, vector quantization has been used. MFCC technique gives accurate reading and comparatively faster than other techniques. The implementation is text-independent speaker identification which means that the matching process will not depend on the spoken words of a particular speaker and when something is spoken that particular sample will be matched to all the stored samples, hence making it 1:N matching. The GUI implementation is the uniqueness in this research work which can be used by banks and many other companies in private or public sector for authenticating the customers based on their voice as biometric signature.

Keywords Speaker recognition · Speaker identification · Verification · Mel-frequency cepstral coefficient · Vector quantization

1 Introduction

The voice recognition system is based on identifying the speaker using his or her speech sample coefficients. This implementation contains two major phases, the first one is feature extraction, and the second is feature matching. Feature extraction is the technique of extracting the coefficients of every speaker which can be done by various techniques such as linear predictive cepstral coefficients (LPCC), Mel-frequency cepstral coefficients (MFCC) and perceptual linear prediction coefficients (PLPC). The second phase is feature matching where the stored coefficients are matched to the input speech coefficients to check the similarity between the speakers and find out whether the new speaker is a stored speaker or

N. S. Wankhede (✉)
Department of Electronics and Telecommunication, Fr. C. Rodrigues Institute of Technology, Navi Mumbai 400703, Maharshtra, India
e-mail: nilashree.w@fcrit.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_22

not. Feature matching can be done using various techniques such as minimum mean square error (MMSE), vector quantization (VQ), hidden Markov model (HMM), Gaussian mixture model (GMM) and dynamic time wrapping (DTW). MFCC technique is used in the research for feature extraction, and vector quantization is used for feature matching.

2 Basics of Speaker Recognition

It has been understood that the anatomical structure of vocal tract is unique for every individual, and hence, the voice information can be used to differentiate between speakers and identify them. If this can be implemented, its application will find way for remote authentication of a person using speaker verification. Speaker recognition can be further categorized as speaker identification and speaker verification.

2.1 Speaker Identification and Speaker Verification

Speaker identification deals with identifying a speaker from a number of speakers. This is a 1:N techniques, that is, one to many technique, where one speaker coefficients are compared with all speakers stored in the database. Speaker verification technique deals with verifying the identity claim of a person. It either accepts or rejects the claim of a person. It is 1:1 technique which means only the database of a particular person will be matched to the current speaker. This technique is faster than identification. This technique also involves categorization into text-dependent and text-independent categories.

2.2 Text-Dependent and Text-Independent

Speaker identification and verification can be further divided into two methods, text-dependent and text-independent. Text-dependent approach includes the need of a particular text to be spoken while feature extraction and the same text to be spoken during matching phase.

In text-independent approach, there is no need of a particular text. Here, the speaker can speak anything randomly. Recognition rate is better in text-dependent than text-independent technique [1].

2.3 Closed System and Open System

Closed system consists of a limited number of speakers stored in the database, whereas an open system may have any number of trained speakers [1].

3 Feature Extraction

Using feature extraction, important features from speech sample can be extracted discarding the noise from it. This is actually a training phase in which a reference model for a speaker is obtained for each registered speaker who has to provide samples of their either pre-decided or random speech. There are various techniques used for feature extraction which include Mel-frequency cepstral coefficient (MFCC), linear predictive coding (LPC), pitch extraction, formant extraction and perceptual linear predictive coefficients (PLPs). In this implementation, the technique used for extraction is Mel-frequency cepstral coefficient (MFCC) which includes pre-emphasis of speech signal and various other processes like framing which follows windowing and also the fast Fourier transform and finally the Mel-frequency cepstrum [2].

4 MFCC

This section explains the procedure to extract a speakers MFCC parameters which has been used in this implementation work.

4.1 Pre-emphasis

The speech signal to be processed is recorded using microphone as input source, and then, it is subjected to pre-emphasis. Pre-emphasis is the process of passing the signal through high pass filter, this is required because the voiced section of the speech signal falls off at high frequencies. Also, the high frequency formants have small amplitudes compared to low frequency formants [2].

4.2 Framing

A short span of time needs to be selected from an audio signal which can be called a frame. To get reliable spectrum estimate, a shorter frame is selected [3].

4.3 Windowing

Various windowing techniques are used to minimize the signal discontinuities. Hamming window which has been proven to provide better frequency resolution

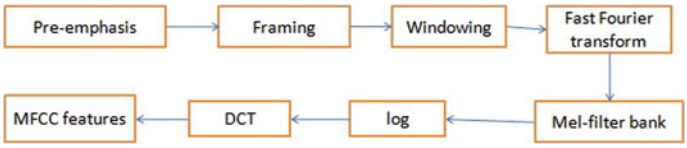


Fig. 1 MFCC block diagram

is applied on every frame [3]. The result of windowing is given as

$$Y(n) = X(n) * w(n)$$

where,

$Y(n)$—output signal

$X(n)$—input signal

$w(n)$—hamming window.

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right] \quad 0 \leq n \leq (N-1) \tag{1}$$

4.4 Fast Fourier Transform

Fast Fourier transform converts the time domain frame of N samples to frequency domain. FFT is to get the magnitude of frequency response of each frame [4].

Figure 1 shows block diagram to calculate MFCC parameters.

4.5 Mel-Frequency Cepstrum

Speech signal does not follow a linear scale. Hence, the Mel-frequency cepstrum is used where the representation of short-term power spectrum of sound which is based on linear cosine transform of a log power spectrum on a nonlinear Mel scale frequency [4].

4.6 Mel-Frequency Filter Banks

An input power spectrum is passed through a number of filter banks called Mel-filters. The output is an array of filtered values also called Mel spectrum, each corresponding to result of filtering the input spectrum through individual filter. Hence, length of output array is the number of filters created. The output of FFT is multiplied by a set of 20 triangular band-pass filters to get log energy of each triangular band-pass filter [5].

4.7 Discrete Cosine Transform

This is the final step in which the Mel spectrum coefficients are converted back to time domain using discrete cosine transform. The MFCCs are obtained [5].

5 Feature Matching

Feature matching is the technique of matching the obtained speech coefficients with the existing coefficients in the database. The features of an input sample are matched with all the samples stored as this is speaker identification mechanism. Feature matching includes various techniques such as vector quantization, mean square error, hidden Markov model, dynamic time wrapping, Gaussian mixture model, stochastic model and artificial neural networks. All the techniques have some advantage over the other. This implementation consists of vector quantization as the feature matching technique [6].

6 Vector Quantization

Vector quantization technique is mostly used for text-dependent systems, but in this implementation, it has been used for text-independent system. Vector quantization is used as a compression technique also, so here, it is used to compress the data and manipulate it so as to keep the useful information hence discarding information not required. Vector quantization technique can be used in voice recognition process. In the speaker recognition technology, the vector quantization method helps in creating a classification system for each speaker. The feature vectors are extracted using VQ from each speaker, and they are called code-words. Each code-word is used to construct a codebook, and thus, codebooks are created for each speaker who enrolls in the system.

In the speaker recognition phase, the data from the tested speaker is compared with the codebook of each speaker, and the differences are calculated. Linde–Buzo–Gray or LBG VQ has been used [6]

6.1 Linde–Buzo–Gray (LBG VQ)

Linde–Buzo–Gray algorithm was introduced by Yoseph Linde, Andrés Buzo and Robert M. Gray in 1980. This is a vector quantization algorithm to derive a good vector quantizer codebook. The set of output points is called as codebook of quantizer. The process of placing these output points is often referred to as codebook design. Linde–Buzo–Gray is based on the K-means algorithm which consists of large set of output vectors training set, initial set of K representative pattern. The representative pattern is updated by computing the centroid of training set vector [6]. The above process continues until all segments have been processed and uptil the new codebook is created. The aim of this procedure is to minimize any distortions in the data creating a codebook which is computationally optimized, while providing a suboptimal solution. Figure 2 gives LBGVQ algorithm.

The performance of VQ process is highly dependent on length successfully takes input from a user and stores it in the voice file which is operated [6] database in the .dat file.

To make the process a little simple for a new comer to use, GUI has been implemented. The program is tested in both normal command prompt mode and GUI mode.

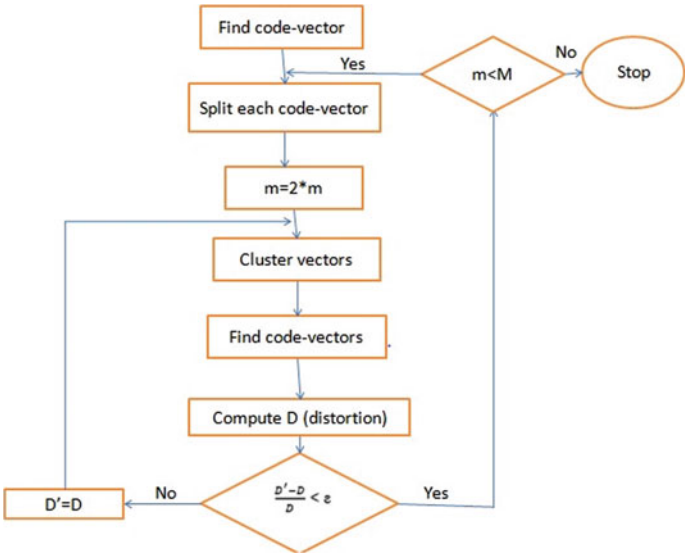


Fig. 2 LBGVQ algorithm

Results obtained are tallied and are satisfactory. There can be more modifications done in training the system and also in verification process achieve more accuracy in terms of speaker authentication for biometric authentication using voice.

7 Implementation Results

This section gives the implementation results for speaker authentication system demonstrated using MATLAB.

As a stage 1 testing process, initially four speakers two male and two female recorded their voice inputs, and using their speech input, 13 MFCC parameters were calculated for each speaker as shown in Table 1. Speaker 1 and 2 are male speakers, whereas speakers 3 and 4 are female speakers.

Figure 3 is a plot which helps in characterizing the speakers based on their extracted MFCC parameters. Various windows were developed as GUI so that any person can use the authentication software with ease. Figure 4 is the menu bar of the system. Four functions have been implemented where the first one allows the speaker to add a new sound from the microphone, the second one, speaker recognition from microphone used to recognize the speaker, database info is used to display the database, and then, delete database which is used to delete the database created. Exit function has also been added so that MATLAB can be successfully closed. Once a user chooses the option "Train," Fig. 5 is the new GUI page displayed. User can choose record option to create a database using a microphone. If database is already present, then the user can choose the load option.

Table 1 MFCC obtained for four speakers

S. No.	Speaker 1	Speaker 2	Speaker 3	Speaker 4
1	20.11082	19.26755	20.62373	13.27085
2	−8.35822	−8.08980	−7.94545	−2.84869
3	11.69873	11.76275	11.93348	12.52903
4	−5.97459	−6.26796	−6.4116	−2.39983
5	15.61956	15.33158	15.55458	10.1233
6	−7.67914	−7.97559	−7.58633	−1.12805
7	15.23478	15.5851	14.93706	8.600952
8	−6.15954	−5.83302	−5.94284	−0.91927
9	10.98147	11.17893	11.38241	8.962311
10	−2.84255	−3.03751	−2.61817	−0.51401
11	7.989414	8.03247	7.622978	7.56579
12	2.41966	−2.77832	−2.31688	0.506629
13	7.717682	7.98802	7.723391	8.037785

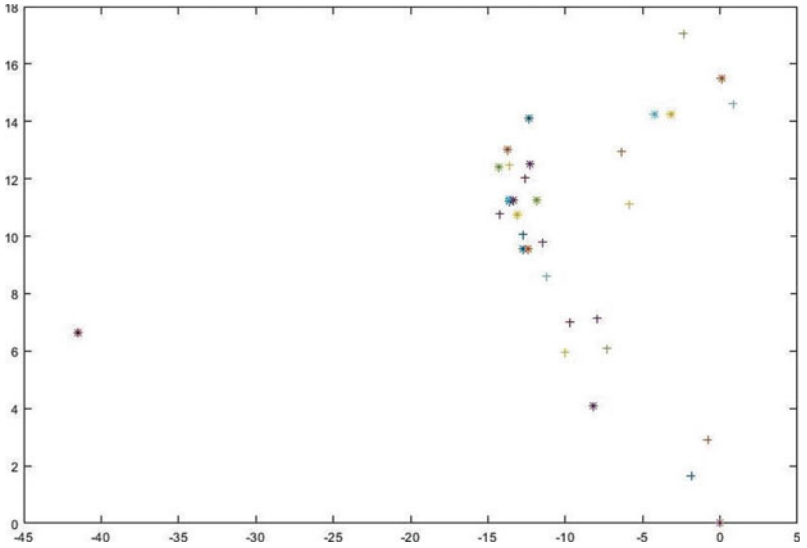


Fig. 3 Vector quantization plot

Fig. 4 Menu bar of the system

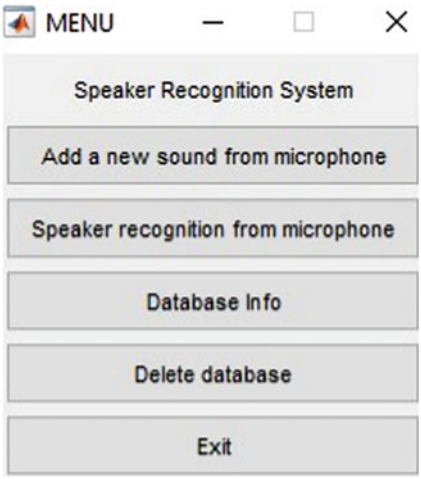


Figure 6 shows the first GUI on the screen. Here, train is for training the system for any user. After training the system, the user can click on test to test the system. Exit is used to exit from the program. The GUI displayed when user chooses "Test" in the first GUI page. This is used to verify the speaker.

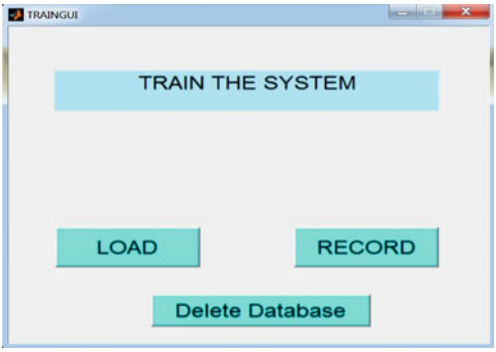


Fig. 5 Train the system

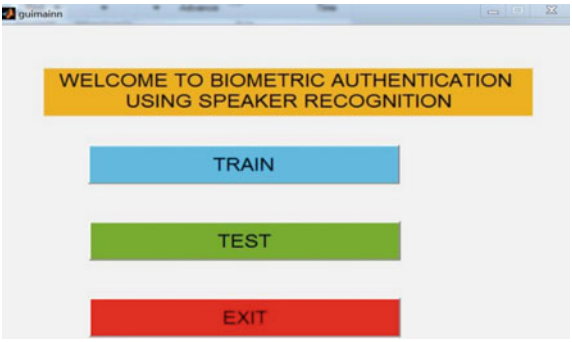


Fig. 6 GUI on the screen

8 Conclusion

The system was tested for 24 speakers. A database containing 24 speakers was made. Accuracy was found out to be 80%. This accuracy was predicted by trying to recognize ten speakers out of which eight were successfully recognized, for two speakers, the ID number used to change with another users ID number. It was concluded that we cannot rely completely on this type of voice authentication system for biometric verification, but it can definitely be used in most of the voice-based applications especially when there is a need of alternative verification and recognition of users is required in already existing biometric authentication systems. Voice authentication system can be an add-on feature which can co-exist with face recognition, fingerprint recognition or iris recognition system to provide multiple level security. Speaker recognition can be considered as one more level of any biometric authentication systems.

9 Future Work

The future work can be to implement vector quantization using other feature extraction technique like LPC and check the accuracy of that system and compare it with this technique. Also, if other speech parameters like vocal tract parameters and pitch values could be additional inputs to create a database for a particular speaker.

References

1. Sajjad A, Shirazi A, Tabassum N, Saquib M, Sheikhs N. Speaker identification & verification using MFCC & SVM. Int Res J Eng Technol 4(2):1950–1953
2. Adhav M, Kene JD (2015) Effective study of feature extraction methods for speaker identification. Int J Sci Eng Technol Res (IJSETR) 4(12):2015
3. Nijhawan G, Soni MK (2014) Speaker Recognition using mfcc and vector quantization. Int J Recent Trends Eng Technol 11(1):211–218
4. Dhonde SB, Jagade SM. Feature extraction techniques in speaker recognition. Int J Recent Technol Mechan Electr Eng 2(5):104–106
5. Tiwari V (2010) MFCC and its application in speaker recognition. Int J Emerg Technol 1(1):19–22. ISSN: 0975-8364
6. Samudre NA (2013) Text-independent speaker identification using Vector quantisation. Int J Eng Res Technol (IJERT) 2(8). ISSN: 2278-0181

Surface Potential Computation for Asymmetric Si–$Si_{1-x}Ge_x$ ID-DG MOSFET Following Ortiz-Conde Model

Anwita Nath, Farnaz Khanam, Swarnav Mukhopadhyay, and Arpan Deyasi

Abstract Surface potential for heterostructure double-gate MOSFET is analytically evaluated following Ortiz-Conde model where asymmetric effect is integrated in terms of considering different gate materials, and henceforth, their work functions for front and back gate. Variation of structural parameters is incorporated into the simulation for sub 10 nm channel length, and independent-gate approximation is taken into account by considering the separate effects of the gates. Fowler–Nordheim tunneling concept is introduced owing to the reduction of dielectric thickness, where the effect of external horizontal bias is analyzed with applied gate voltage. P-type Si-$Si_{1-x}Ge_x$ material composition is considered for simulation, where variation of mole fraction showed significant effect on surface potential. Lowering of surface potential is possible for gate-engineered asymmetry compared to the symmetric structure with identical structural and otherwise material composition. Results are important for investigating drain current in presence of quantum–mechanical effect.

Keywords Surface potential · Independent-gate MOSFET · Heterostructure · Ortiz-conde model · Asymmetric effect

A. Nath · F. Khanam · A. Deyasi (✉)
Department of Electronics and Communication Engineering, RCC Institute of Information Technology, Kolkata, India
e-mail: deyasi_arpan@yahoo.co.in

A. Nath
e-mail: nathanwita@gmail.com

F. Khanam
e-mail: farnazkhantabish@gmail.com

S. Mukhopadhyay
Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, India
e-mail: swarnavhome@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_23

1 Introduction

In the nano-age, the prime objective of chip design is to achieve lower power dissipation coupled with greater packing density [1]. The shortcoming associated with this objective is the severe short-channel effect [2], and several solutions in different angles are proposed by workers in the last decade. In the sub 50 nm region, one established solution is gate control mechanism, which speaks in favor of multigate transistors [3] compared to the existing single-gate technology. Several multigate structures are analyzed in last few years, and double-gate MOSFET is probably the most searched candidate among them [4–6] due to its simplicity in structural design as well as ease of mathematical analysis. This novel class of device already exhibited lower DIBL, moderate subthreshold swing, improved transconductance, and moreover better scalability [7, 8]; which is one of the prerequisite for VLSI circuit design [9].

Several architectures are proposed for better performance of double-gate MOSFET, among which tied-gate architecture is advantageous for higher drain current and better floorplanning [10]. But this architecture always leads to the oversimplified symmetric structure analysis. In order to make it asymmetric form material parameter point of view, independent-gate architecture is proposed [11] where individual role of both front and back gates is analyzed independently. This also ensures lower threshold voltage and power dissipation compared to other architectures.

Taur and Ortiz-Conde made path breaking contributions [4–6] in the progress of DG MOSFET, and their analytical findings are significant for both experimental and theoretical workers. However, their contributions are based on the symmetric lightly doped structures where both drift and diffusion mechanisms of carrier transport are considered. Also, the work of Ortiz-Conde is not so much analyzed for very small length devices, and sub 50 nm is not investigated in details. Also, the role of asymmetric structure in terms of gate material plays a major role for independent-gate devices, in terms of leakage current control. Henceforth, research can be carried out in this direction, and the present paper tries to throw some light into the arena. In order to proceed for the investigation, it is customary to calculate surface potential under the asymmetric condition, in presence of high-K dielectrics, and for different oxide thicknesses. The role of horizontal bias is compared with the unbiased condition, and the material composition of the heterostructure leads to tuning of surface potential. Results are summarized in next part after the brief description of analytical formulation, and work is concluded with significant findings.

2 Mathematical Formulation

Maximum potential will be at the center of the MOSFET, because the electric field is zero at the center between two gates. But due to heterojunction available, the maximum potential will not be at the center but shifted toward the front gate by an

amount of valance band discontinuity in the band structure. So, at the center, the potential will be slightly less than the maximum potential.

$$\phi_{\mathrm{oc}} = V + v_{\mathrm{t}} \ln\left(\frac{2\pi^2 \in_{\mathrm{s}} v_{\mathrm{t}}}{q n_i t_{\mathrm{s}}^2}\right) - \Delta E_{\mathrm{V}} \tag{1}$$

where ΔE_v is the valence band discontinuity, 'x' is the mole fraction of Ge, 'V' denotes applied horizontal bias, 't_{s}' represents thickness of total structure, φ_{oc} denotes potential at the center.

Surface potential in equilibrium condition according to Ortiz-Conde [5] is given by.

$$\varphi_{\mathrm{o}} = U_1 - \sqrt{U_1^2 + V_{\mathrm{gf}}\phi_{\mathrm{oc}}} \tag{2}$$

where $V_{\mathrm{gf}} = V_{\mathrm{gs}} - V_{\mathrm{fb}}$.

$$U_1 = \frac{1}{2}\left[-V_{\mathrm{gf}} + (1 + r_1)\phi_{\mathrm{oc}}\right]$$
$$r_1 = (At_{\mathrm{ox}} + B)\left(\frac{C}{t_{\mathrm{s}}} + D\right)\mathrm{e}^{-\mathrm{VE}} \tag{3}$$

3 Results and Discussion

Based on the Eq. (1), surface potential of the proposed asymmetric structure is computed and plotted as a function of gate voltage. Figure 1 shows the variation for different dielectric thickness at $V_{\mathrm{SD}} = 0.2$ V. From the plot, it is observed that increasing reverse gate voltage enhances surface potential. The change is at first takes linear shape, but after reaching the near-about saturation point, it becomes almost constant. With increase of oxide thickness, surface potential becomes almost constant early at higher negative gate voltage.

The variation of surface potential is due to the fact that when drain current moves from cut-off to active region, the rate of accumulation of charge decreases, and corresponding slope of surface potential changes. By introducing materials with higher dielectric constant, it is observed that the minimum value of surface potential decreases (negative scale). This is due to the fact that higher dielectric reduces the capacitance, which, in turn, reduces the accumulation of charge at surface. Therefore, surface potential reduces.

Figure 2 shows the variation of surface potential with various dielectric materials (high-K), and the result is compared with conventional SiO_2 material. Since high-K

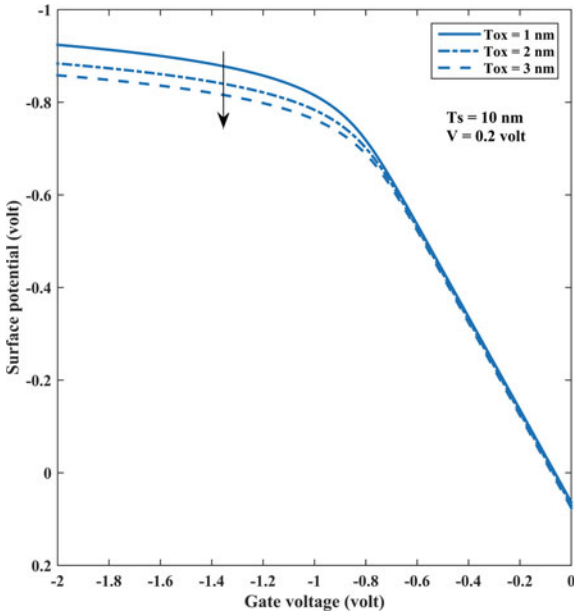


Fig. 1 Surface potential with gate voltage for different thickness of dielectric material

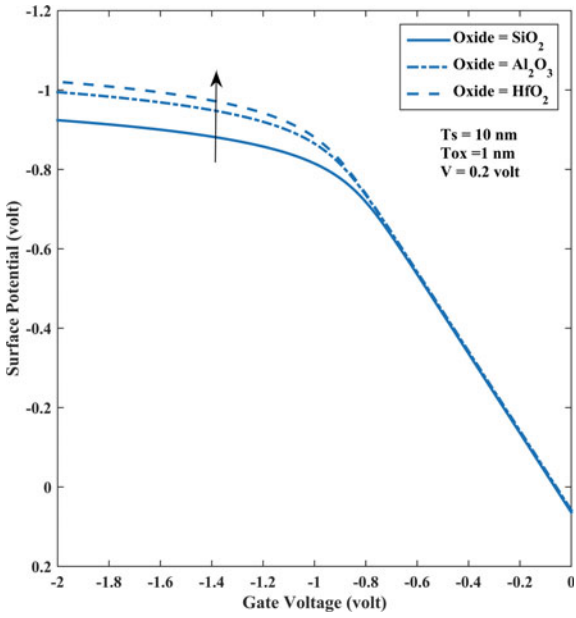


Fig. 2 Surface potential with gate voltage for different dielectric materials, and comparison is made with conventional dielectric

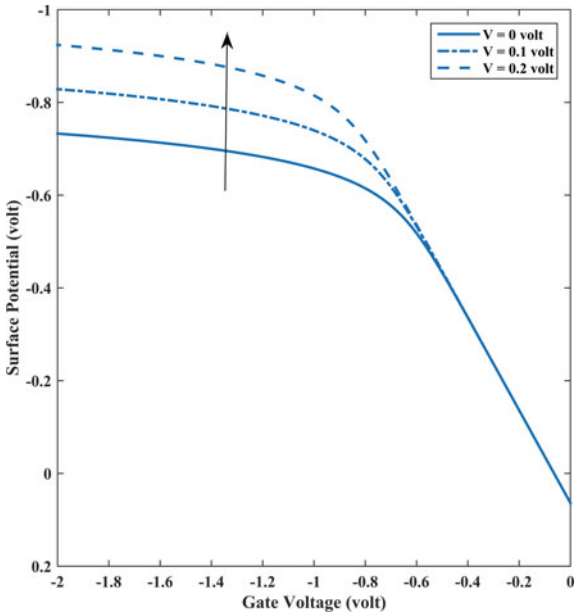


Fig. 3 Surface potential with gate voltage for different horizontal bias, comparison is made with unbiased condition

materials enhances the capacitive effect, thus, more charges are stored beneath the oxide–semiconductor interface, which causes higher potential.

By introducing horizontal bias, higher surface potential is achieved as it causes further band bending, which basically enhances minority carrier accumulation. This result is plotted in Fig. 3. This can be further explained with help of Fig. 4, where both center potential (following Ortiz-Conde model) [5] and surface potential are compared with gate voltage. After threshold, center potential becomes constant, as it is the surface potential computed in absence of horizontal bias. But if the effect of V_{SD} is added, then more charges are accumulated at the interface due to band bending, which causes higher surface potential (in negative scale). Thus, a difference is observed after the threshold point between center and surface potential, as depicted in Fig. 4.

With increase of Ge percentage in the SiGe channel, it is found that mobility of the carriers is increased, which causes the availability of more carriers at the surface form the bulk. These extra carriers, in turn, enhance the charge accumulation in the surface, which causes an increase of surface potential. The result is pictorially represented in Fig. 5. Since the mobility of carriers is increased once the Ge concentration exceeds 0.25, hence, the variations are taken after that threshold.

Keeping back gate as Ni, if the material of front gate is changed with reduction of work function, then surface potential increases. This asymmetric profile is represented in Fig. 6 and compared with the symmetric material composition [both front and back gate are constructed using Ni]. Because of asymmetry, more carriers are gathered near the surface, as work function of front gate is less than back gate.

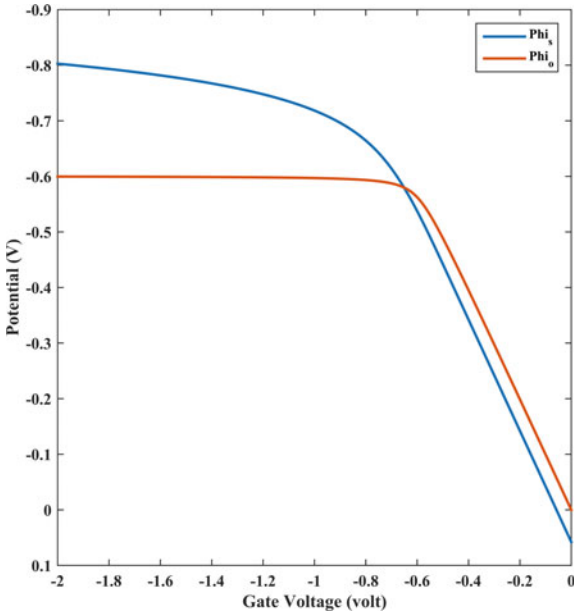


Fig. 4 Comparative study of surface potential and center potential with gate voltage

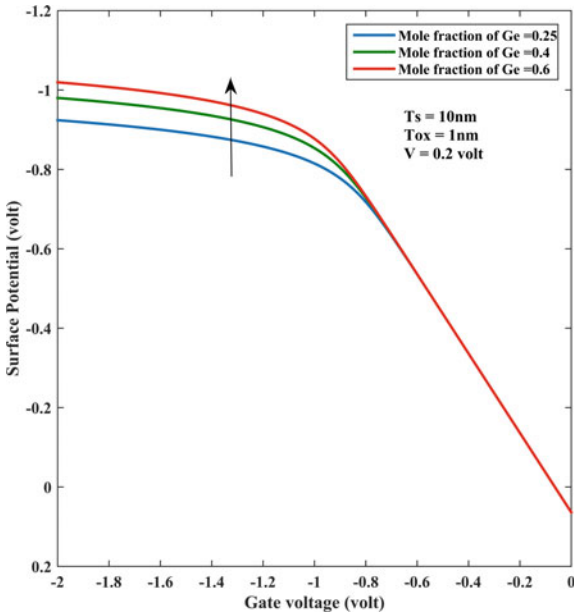


Fig. 5 Surface potential with gate voltage for different material compositions

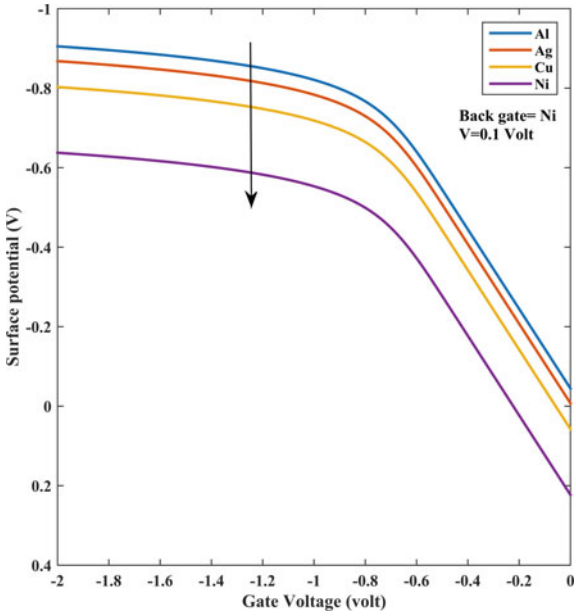


Fig. 6 Surface potential variation due to asymmetry in gate materials [comparison is performed with symmetric structure]

Therefore, more amount of band bending is observed, which helps in accumulation of carriers at surface. This causes higher surface potential.

A comparative study is summarized in following tables between asymmetric and symmetric structures under various biasing conditions and different materials.

Table 1 represents the effect of oxide thickness on surface potential with different gate voltages for both symmetric and asymmetric structures.

4 Conclusion

Comparative study between asymmetric [gate-engineered] and symmetric structures exhibit that surface potential can be lowered by intentionally choosing the gate materials, which effectively can be utilized for lowering of leakage current after threshold. It is found that higher Ge percentage [>0.25] can cause higher surface potential, which can be counterbalanced by higher dielectric thickness or high-K materials. Effect of horizontal bias plays a critical role in this aspect, for fine-tuning the surface potential in the range of interest.

Table 1 Effect of dielectric thickness on surface potential for asymmetric and symmetric structures [$T_{sub} = 10$ nm, $V_{SD} = 0$ V, HfO_2 as dielectric]

T_{ox} [nm]	Gate voltage [V]	Asymmetric surface potential [V]	Symmetric surface potential [V]
3	−0.4	−0.3397	−0.3998
	−0.8	−0.6453	−0.7055
	−1.2	−0.7213	0.7814
	−1.6	−0.7535	−0.8137
2	−0.4	−0.3409	−0.3999
	−0.8	−0.6593	−0.7183
	−1.2	−0.7408	−0.07998
	−1.6	−0.7742	−0.8332
1	−0.4	−0.3421	−0.3999
	−0.8	−0.6808	−0.7386
	−1.2	−0.7733	−0.8311
	−1.6	−0.8088	−0.8666

References

1. Chen Y-H, Chan WM, Wu W-C, Liao H-J, Pan K-H, Liaw J-J, Chung T-H, Li Q, Chang GH, Lin C-Y, Chiang M-C, Wu S-Y, Natarajan S, Chang J (2014) A 16 nm 128 Mb SRAM in High-K Metal-Gate FinFET Technology with Write-Assist Circuitry for Low-V Applications. In: Proceeding of IEEE international solid-state circuits conference, 2014
2. Kim YB (2010) Challenges for Nanoscale MOSFETs and Emerging Nanoelectronics. Trans Electr Electron Mater 11(3):93–105
3. Hiramoto T, Nagumo T (2006) Multi-Gate MOSFETs with back-gate control. IEEE international conference on IC design and technology, 2006
4. Taur Y (2001) Analytic solutions of charge and capacitance in symmetric and asymmetric double-gate MOSFETs. IEEE Trans Electron Devices 48(12):2861–2869
5. Ortiz-Conde A, Garcia-Sanchez FJ, Muci J (2005) Rigorous analytic solution for the drain current of undoped symmetric dual-gate MOSFETs. Solid State Electron 49(4):640–647
6. Ortiz-Conde A, García-Sánchez FJ, Muci J, Malobabic S, Liou JJ (2007) A review of core compact models for undoped double-gate SOI MOSFETs. IEEE Trans Electron Dev 54(1):131–140
7. Singh D, Panda S, Mohapatra SK, Pradhan KP, Sahu PK (2014) Static performance analysis on UTB-SG and DG MOSFETs with Si and III–V channel materials. International conference on high performance computing and applications, 2014
8. Deyasi A, Chowdhury AR, Roy K, Sarkar A (2018) Effect of high-K dielectric on drain current of ID-DG MOSFET using Ortiz-Conde Model. EDKCON, Nov 2018
9. Zhang W, Fossum JG, Mathew L, Du Y (2005) Physical Insights regarding design and performance of independent-gate FinFETs. IEEE Trans Electron Dev 52(10):2198–2206
10. Woo HJ, Jin KC, Kyu CY (2008) Universal potential model in tied and separated double-gate MOSFETs with consideration of symmetric and asymmetric structure. IEEE Trans Electr Dev 55(6):1472–1479
11. Liu YX, Masahara M, Ishii K, Tsutsumi T, Sekigawa T, Takashima H, Yamauchi H, Suzuki E (2003) Flexible threshold voltage FinFETs with independent double gates and an ideal rectangular cross-section Si-Fin channel. IEEE international electron devices meeting

An Optimized Hybrid Clustering Method Using Salp Swarm Optimization with K-Means

Shashank Gavel, Pallavi Joshi, Sudarshan Tiwari, and Ajay Singh Raghuvanshi

Abstract The field of data clustering is one of the commonly used techniques to analyze the data based on their structure. The use of clustering has found many applications in the field of data mining, image analysis, pattern recognition, and signal processing. This paper presents a hybrid scheme of data clustering by utilizing K-means clustering with salp swarm algorithm (SSA). The SSA is used to solve the problem of local optima for K-means and provide best suitable fitness value for the dataset. The experimental results show best suitable fitness value for different standard datasets by m the standard deviation. The experiment is carried out in MATLAB.

Keywords Clustering · K-means clustering · Salp swarm algorithm · SSA

1 Introduction

The problem for optimization has been frequently encountered in the field of science and engineering. The traditional problem was dependent upon being continuous, differentiable, unimodal and linear types of functions. At that time, to solve the mentioned traditional problem methods like Newton and quasi-Newton were used

S. Gavel (✉) · P. Joshi · A. S. Raghuvanshi
Department of Electronics and Communication, National Institute of Technology, Raipur, Raipur 492010, India
e-mail: sgavel.phd2016.etc@nitrr.ac.in

P. Joshi
e-mail: pjoshi.phd2017.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

S. Tiwari
Department of Electronics and Communication, Motilal Nehru National Institute of Technology, Allahabad 211002, India
e-mail: stiwari@mnnit.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_24

[1, 2]. However, nowadays, problems like non-differentiable, discontinuous, and nonlinear problems occur due to the availability of multiobjective applications. The need of an efficient mechanism to handle the datasets obtained from these problems becomes a necessary task. The inherent occurrence of complex real-world problems has given birth to various optimization algorithms such as ant colony optimization (ACO) [3], particle swarm optimization (PSO) [4], genetic algorithms (GA), artificial bee colony optimization (ABC) [5], and bat algorithms (BA).

Many evolutionary algorithms (EA) have tried to solve the complex problems by providing the optimum value for the utilized objective functions. In a similar way, this algorithm has shown significant performance in handling the large datasets by getting utilized in a hybrid manner. Clustering is one such technique that is used to handle the dataset by dividing them into different clusters having cluster centroid. Algorithms like K-means are in general used clustering algorithm that helps the dataset to form efficient cluster [6–8]. But this algorithm alone creates problem in handling large dataset or dataset having large number of features. K-means undergo problem of local conjunction and convergence that restricts the algorithm from obtaining optimal clusters with efficient cluster centroid. Several algorithms have tried to solve the problem of K-means by using them with efficient optimization algorithms. Author in [9] has utilized PSO with hybrid of K-means in order to obtain optimal cluster. Similarly, in [10], author has utilized gravitational search algorithm (GSA) with K-harmonic mean (KHarm) in order to obtain optimal clusters. Algorithms like ABC and GA has also been utilized in a hybrid manner in order to obtain optimal cluster with best suitable cluster centroid [11, 12].

The K-means algorithm is an important algorithm that was first proposed in [13, 14]. The application of K-means was to divide the data values into different sets of clusters. These sets are formed depending upon the data value having similar properties. The similar data value belongs to the same cluster and is different from those data values that belong different cluster as dissimilar as possible. However, beside this property, the K-means suffer problem of local conjunction and convergence that restricts it in the local optimum value. In this article, we have utilized Euclidean distance as an objective function for K-means clustering. We propose a new hybrid clustering technique by optimizing the fitness function of K-means clustering with salp swarm algorithm (SSA). The hybrid clustering is proposed to solve the problem of local optima for K-means clustering and to obtain optimal clusters. Standard UCI datasets have been utilized in order to judge the overall performance of the proposed method.

Rest of the paper is as follows: Section 2 includes the analytical techniques used in methodology, Sect. 3 covers the proposed methodology. Section 4 includes results and discussion section followed by conclusion in Sect. 5.

2 Analytical Techniques Used in Methodology

2.1 K-Means Clustering

K-means clustering is one of the frequently used clustering algorithms that has been utilized for many applications [5, 15]. The K-means algorithm is used to club the data points in the form of cluster by utilizing their properties and degree of similarity between them. The main concept of this algorithm lies in grouping the data in the form of clusters by providing them a best possible cluster centroid. In K-means algorithm, the Euclidean distance-based metrics are used in order to find the distance between the similar data value and choosing them as a part of clusters by analyzing the distance value. Let us consider a dataset, $R = \{r_1, r_2, r_3, \ldots\ldots r_n\}$ having 'k' clusters. The aim of the algorithm is to calculate the number of clusters in the dataset. For this purpose, firstly, the distance between the cluster its data point is to be calculated using the below fitness function, i.e.,

$$D(r_i, C) = \sqrt{\sum_{i=1}^{n} (r_i - C)^2} \tag{1}$$

here 'C' is the value of cluster that is given by $k = \{C_1, C_2, C_3, \ldots\ldots C_m\}$, where '$m$' belongs to number of clusters ($m > 2$).

2.2 Salp Swarm Algorithm (SSA)

Salp swarm algorithm is newly proposed meta-heuristic optimization algorithm that is inspired by the behavior of salps [16, 17]. Salps are transparent tissues which consist of barrel-shaped structure. These salps sometime forms a swarm, which is known as salp chain. The mathematical structure of SSA depends upon the group of leader as well as follower. The leader take position in front followed by the salps known as follower. Let us consider that the 'F' denotes the food source that act as a target for the swarm in a search space. The salp leader will update its position depending on the first salp's position, i.e.,

$$x_j^1 = \begin{cases} F_j + c1\big(\big(Ub_j - Lb_j\big)c2 + Lb_j\big), c3 \geq 0 \\ F_j - c1\big(\big(Ub_j - Lb_j\big)c2 + Lb_j\big), c3 < 0 \end{cases} \tag{2}$$

here x_j^1 represents the value of position of first salp in the jth dimension that act as a leader to the group. Ub_j and Lb_j are the values of boundaries decided, i.e., upper boundary and lower boundary. F_j is the position of food, $c1$, $c2$, and $c3$ are the randomly taken numbers in which $c1$ is used to balance the value of exploration and exploitation. The mathematical value of c1 in terms of number of iterations is

given by:

$$c1 = 2e^{-(\frac{4t}{T})^2} \tag{3}$$

where T is the total iterations in which 't' is the value of current iteration. The range of c2 and c3 is decided to be [0, 1] in which c3 manages to act as an indicator that provides the next position of search toward the positive or the negative end. Since to reach toward the final optimized value, the salps follow the Newton's second law of motion that provides the final value in terms of:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \tag{4}$$

3 Proposed Methodology

In this section, we proposed a hybrid clustering technique by utilizing K-means and SSA. Here, the Euclidean distance is used a fitness function and we have utilized SSA in order to optimized the fitness function. The optimized hybrid method will provide optimal value of cluster centroids. The originality of the paper lies in optimizing the fitness function and using it as a clustering algorithm by applying K-means. Figure 1 shows the overall architecture of the proposed algorithm, i.e., KMSSA. Here, firstly, the initial dataset is processed and the initial cluster centroid for each cluster is calculated using the fitness function as shown in Eq. (1). The initial centroid values act as an initial value of position to the optimization algorithm. The detail step-wise algorithm is shown below:

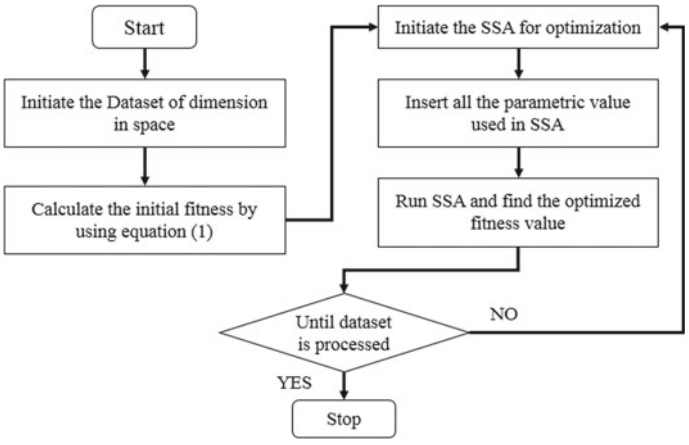


Fig. 1 Flow diagram for proposed hybrid KMSSA

Steps to apply hybrid KMSSA:

1. Initiate the dataset in the dimensional space.
2. Calculate the initial fitness value using the fitness function as shown in Eq. (1).
3. Initiate the SSA and insert all the parametric values, i.e., $c1$, $c2$, and $c3$ along with the upper bound and lower bound values.
4. Run the SSA algorithm and update the value using Eq. (4).
5. Update till minimum fitness value is obtained, i.e., efficient cluster with optimal value of cluster centroid.

4 Results and Its Discussion

To analyze the performance of the proposed KMSSA, we have compared the results with PSO, ACO, ABC and standard K-means algorithm. We have utilized standard datasets of UCI, i.e., detail is shown in Table 1 [18–20]. The table includes the detail description of the dataset used in terms of no. of attributes (features), no. of classes of data and no. of instances. These datasets are standard and are universally used to evaluate the performance of the clustering-based algorithms.

We have compared the results in terms of best fitness value (BFV), average fitness value (AFV) and worst fitness value (WFV) along with the standard deviation (SD) between the fitness value and total time taken by the algorithm to analyze the dataset.

For Iris dataset, the results obtained are shown in Table 2. The result obtained from the proposed method is promising as compared to other algorithms. The algorithm is run for 40 number of iterations and the best results are compared. The best fitness value obtained from the proposed algorithm is **94.20** with **0** standard deviation and the time taken is also less, i.e., KMSSA > ABC > PSO > ACO > K-means.

Table 1 Overall description of dataset

S. No	Dataset	No. of attributes	No. of classes	No. of instances
1	Iris	4	3	150
2	Wine	13	3	178
3	Glass	9	6	214
4	Cancer	9	2	683

Table 2 Results comparison for Iris dataset

Method	BFV	AFV	WFV	SD	Time (in sec)
KMSSA	**94.20**	**94.20**	**94.20**	**0**	**15.4**
PSO	96.894	97.233	97.898	0.348	30
ACO	97.10	97.17	97.80	0.367	24
ABC	94.106	94.61	94.9	0.009	n/a
K-means	97.33	106.50	120.45	14.63	0.4

Table 3 Results comparison for wine dataset

Method	BFV	AFV	WFV	SD	Time (in sec)
KMSSA	**16,285**	**16,285**	**16,285**	**0**	**16.8**
PSO	16,334	16,417	16,562	85.4	123
ACO	16,530	16,530	16,530	0	170
ABC	16,306	16,348	16,390	30.6	n/a
K-means	16,555	18,061	18,563	793.61	0.7

For wine dataset, the results obtained are shown in Table 3. The overall fitness value achieved using the fitness function is less as compared to the existing algorithms. The algorithm is run for 40 number of iterations for this dataset. The overall time taken by the algorithm to process the wine dataset is less than that of other algorithm.

For glass dataset, the results obtained are shown in Table 4. The overall result obtained is promising and is efficient as compared to other algorithms. Since the standard deviation is **0,** the fitness value achieved converges to a single value for number of iterations.

For cancer dataset, the overall performance of the proposed algorithm is superior to that of other existing algorithms. The overall result is shown in Table 5 for cancer dataset.

The overall time consumption for the proposed algorithm is much less than that of the other algorithms, i.e., for Iris dataset 15.4 sec, for wine dataset 16.8 sec, for glass dataset 28 sec, and for cancer dataset 15.2 sec. The result for the standard deviation

Table 4 Results comparison for glass dataset

Method	BFV	AFV	WFV	SD	Time (in sec)
KMSSA	**220.31**	**220.31**	**220.31**	**0**	**28**
PSO	270.57	275.71	283.52	4.55	400
ACO	269.72	273.46	280.08	3.58	395
ABC	230.55	230.98	231.23	0.013	n/a
K-means	215.4	235.5	255.38	12.47	1

Table 5 Results comparison for cancer dataset

Method	BFV	AFV	WFV	SD	Time (in sec)
KMSSA	**2971.8**	**2971.8**	**2971.8**	**0**	**15.2**
PSO	2973.5	3050.4	3318.8	110.8	123
ACO	2970.4	3046.0	3242.01	90.5	123
ABC	2964.3	2975.2	2989.6	12.5	n/a
K-means	2999.1	3251.21	3521.5	251.14	0.5

is found to be negligible for all the utilized dataset that states that the proposed algorithm nearly achieves best fitness value in limited number of iterations. This can decrease the overall time consumption by providing efficient performance.

5 Conclusion

In this article, a hybrid method for clustering using K-means with SSA, i.e., KMSSA is proposed. The SSA optimizes the fitness function for K-means clustering and enhances the performance. In the early stage of algorithm, the fitness function provides the temporary solution for clustering that is further optimized using SSA. To validate the performance of the proposed method, standard dataset has been utilized, i.e., UCI Machine Learning lab dataset. The results show that the proposed KMSSA has better performance that of the compared algorithms, i.e., PSO, ACO, ABC, K-means. The overall time consumption for the proposed algorithm is very less. So, by evaluating the results, we conclude that the proposed algorithm provides better clustering by taking less computational time. In the future, we will develop a new distance-based fitness function in order to achieve high performance by optimization.

References

1. Niknam T, Amiri B (2010) An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. Appl Soft Comput 10(1):183–197
2. Fischer A (1992) A special Newton-type optimization method. Optimization 24(3–4):269–284
3. Dorigo M, Di Caro G (1999) Ant colony optimization: a new meta-heuristic. In Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), vol 2, pp 1470–1477). IEEE
4. Kennedy J (2010) Particle swarm optimization. Encyclopedia of machine learning, pp 760–766
5. Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. Appl Soft Comput 8(1):687–697
6. Huang Z (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min Knowl Disc 2(3):283–304
7. Capó M, Pérez A, Lozano JA (2017) An efficient approximation to the K-means clustering for massive data. Knowl-Based Syst 117:56–69
8. Al-Yaseen WL, Othman ZA, Nazri MZA (2017) Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst Appl 67:296–303
9. Cui X, Potok TE (2005) Document clustering analysis based on hybrid PSO+ K-means algorithm. J Comput Sci (special issue) 27:33
10. Yin M, Hu Y, Yang F, Li X, Gu W (2011) A novel hybrid K-harmonic means and gravitational search algorithm approach for clustering. Expert Syst Appl 38(8):9319–9324
11. Krishnamoorthi M, Natarajan AM (2013) ABK-means: an algorithm for data clustering using ABC and K-means algorithm. Int J Comput Sci Eng 8(4):383–391
12. Kim KJ, Ahn H (2008) A recommender system using GA K-means clustering in an online shopping market. Expert Syst Appl 34(2):1200–1209
13. Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. J R Stat Soc Ser C (Appl Stat) 28(1):100–108

14. Gan G, Ng MKP (2017) K-means clustering with outlier removal. Pattern Recogn Lett 90:8–14
15. Kazemi MS, Kazemi K, Yaghoobi MA, Bazargan H (2016) A hybrid method for estimating the process change point using support vector machine and fuzzy statistical clustering. Appl Soft Comput 40:507–516
16. Mirjalili S, Gandomi AH, Mirjalili SZ, Saremi S, Faris H, Mirjalili SM (2017) Salp Swarm Algorithm: a bio-inspired optimizer for engineering design problems. Adv Eng Softw 114:163–191
17. Faris H, Mafarja MM, Heidari AA, Aljarah I, Ala'M AZ, Mirjalili S, Fujita H (2018) An efficient binary salp swarm algorithm with crossover scheme for feature selection problems. Knowl Based Syst 154:43-67
18. Fisher RA (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7(2):179–188
19. Aeberhard S, Coomans D, De Vel O (1994) Comparative analysis of statistical pattern recognition methods in high dimensional settings. Pattern Recogn 27(8):1065–1077
20. Street WN, Wolberg WH, Mangasarian OL (1993) Nuclear feature extraction for breast tumor diagnosis. In Biomedical image processing and biomedical visualization, vol 1905, pp 861–870. International Society for Optics and Photonics

An Investigation into Contemporary Developments in the Cloud Computing Related to Effectiveness in Data Storages, Service Providers and Technologies: A Review

Safwan A. S. Al-Shaibani, Parag Bhalchandra, G. D. Kurundkar, and R. S. Wasle

Abstract Cloud computing is a current trending in IT industry that interchanges computing and information away from the portable PCs and desktop into great data centers. The research literature on Clouds available across the globe lacks comprehensive information approximately the procedure of Cloud computing environment related to the data storages, what technologies are used and who are the service providers? This is primary intention behind this review paper as a microcosm of about Cloud computing. This paper reviews Cloud computing through its background to recent trends. Furthermore, these thoughts purpose to deliver a complete synopsis and a big picture to person who reads in this exciting area allowing them to possible research opportunities and avenues.

Keywords Cloud computing · Data storage · CSP · Amazon Web service (AWS)

1 Introduction

The Cloud has become a significant vehicle for conveying assets, for example, figuring and capacity to clients on their interest [1–4]. Late enthusiasm for distributed computing has been driven by new contributions of registering assets that are

S. A. S. Al-Shaibani (✉)
Information Centre, Government of Yemen, Sanaa, Yemen
e-mail: safwan.srtmu@gmail.com

S. A. S. Al-Shaibani · P. Bhalchandra · R. S. Wasle
School of Computational Sciecnes, S.R.T.M. University, Nanded 431606, MS, India
e-mail: srtmun.parag@gmail.com

R. S. Wasle
e-mail: rajeshwalse@gmail.com

G. D. Kurundkar
Department of Computer Science, SGBS College, Dist parbhani, Purna, MS, India
e-mail: gajanan.kurundkar@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_25

appealing because of flexible versatility, giving a noteworthy preferred position over the regular procurement and organization of hardware that was recently required. Being an ongoing slanting in IT, it has moved figuring and data from the work region and advantageous PCs into gigantic server ranches. The essential favored situation of conveyed processing is that customers do not have to pay for system or for its foundation neither for anticipated that work should manage such establishment and upkeep. The crucial target of appropriated registering is to use flowed resources, go along with them to achieve higher throughput and have the choice to deal with enormous extension computation issues. The Cloud figuring oversees virtualization, flexibility, interoperability, nature of organization and the movement models of the Cloud, specifically private, open and creamer [5–11]. The circulated registering includes gigantic data storing, virtual machine, security instrument, organizations, plan models and servers.

2 Background and Related Literature

It has dynamic provisioning that considers the strategy of affiliations subject to current intrigue necessities. This is done conventionally using programming robotization, engaging the unanticipated turn of events and narrowing of affiliation limit, fluctuating [12]. The relationship of relationship in the Cloud wires everything from using business applications to the latest application on the freshest moved cells, and controlled metering uses metering for overseeing and smoothing out the affiliation and to give declaring and charging information [13–15]. Thusly, purchasers are charged for relationship as demonstrated by the sum they have really used during the charging time parcel. Further, on-request self-association Cloud figuring licenses the clients to utilize Web associations and assets on request. One can login to a site at whatever point and use them. Therefore, wide system access since scattered figuring is completely electronic can be gotten to from any place and at whatever point. The favorable position pooling appropriated enrolling awards different inhabitants to share a pool of advantages [16, 17]. One can share a solitary physical case of apparatus, database and key structure. Moreover, expedient adaptability is certainly not hard corresponding the focal points vertically or on a level plane at whatever point. Scaling of advantages proposes the constraint of favorable circumstances for supervise broadening or lessening interest [18]. The points of interest being utilized by clients at some discretionary inspiration driving time are regularly viewed. Thusly, we utilize the Cloud considering the essential variables kept up by them, which wires, flexibility, that is Cloud-based associations are perfect for relationship with making or fluctuating data transmission requests. On the off chance that necessities are broadened, it is unquestionably not hard relative up your Cloud limit, drawing on the association's remote servers. Also, we can curtail it once more, subordinate upon the fundamentals [19]. In extra, mists support for catastrophe recuperation, where relationship of all sizes should put resources into excited cataclysm recuperation, yet for increasingly humble affiliations that come up short on the fundamental

money and wellness, this is as regularly as conceivable more a perfect than the real world [20]. Cloud is correct currently helping more affiliations stay away from that structure, private undertakings are twice as likely as more noteworthy relationship to have executed Cloud-based fortress and recuperation designs that additional time, keep up some fundamental decent ways from gigantic early hypothesis and climb distant bent as a critical piece of the game plan. It is fascinating to see the altered programming update utility of the mists. The brilliance of circled figuring is that, its servers are off-premise, far out of your closeness. Providers oversee them for clients and turn out conventional programming stimulates including security resuscitates, and therefore, client does not need to stress over relaxing around inertly keeping up the framework [21]. Thusly, capital-usage free-dispersed figuring removes the basic expense of equipment. Client can basically pay even more similarly as costs rise and worth an interest-based model. This develops the composed effort where different individuals in the social event of clients can get the chance to change and offer reports at whatever point, from any place, they are set up to accomplish significantly progressively together and improve [22]. Cloud-based work strategy and record sharing applications help them with making resuscitates progressively and give them full perceivable nature of their composed undertakings. Client can work from any place as client basically needs a Web relationship with get related with the working environment. The record control is in client's hold. The more operators and partners work together on reports, the more observable the essential for watertight record control. Prior to the Cloud, laborers expected to send files forward and in reverse as email relationship with be pursued by each client along these lines. In the end, ordinarily sooner, client ends up with a hazardous situation of clashing record substance, arrangements, and titles. Right when client gains the ground to scattered figuring, all files are dealt with halfway, and everybody sees one variant of this present reality. Logically perceptible noticeable quality recommends improved encouraged effort, which at last strategies better work and an inexorably supportive fundamental concern. Scattered enrolling in like way gives dynamically unquestionable security when physical loss of huge rigging contraptions occurs. This is considering the way that, the client's information is dealt within the Cloud, and client can get to it paying little psyche to what happens to client's machines [23].

3 Cloud Architectures and Models

The underlying foundations of mists lie in virtualization. With the presentation of virtualization, things have changed colossally. Virtualization improves asset usage and vitality productivity that point serving to significantly diminish server upkeep overhead and giving speedy catastrophe recovery provoking high openness. Virtualization has been huge for dispersed registering since it withdraws programming from gear and gives a framework to quickly reallocate applications across servers reliant on computational solicitations [24]. The virtualization was a critical development toward Cloud system. Virtualized circumstances supervised by internal system chiefs and

as usual virtualization stages do not give the consideration layer that engages Cloud organizations. For fogs, to enable a circumstance, a layer of pondering and on demand provisioning must be given on the top. This organization layer is a noteworthy nature of any Cloud condition as it covers the multifaceted design of the establishment and gives a Cloud the administrators interface to customers. Dependent upon the interface utilization, a Cloud the administrators interface can be gotten to through an organization dashboard, REST or SOAP Web organizations, programming APIs, or various organizations. For example, Amazon Web Services give access through an organization dashboard or REST/SOAP Web organizations [25]. The Cloud the officials interfaces like the Amazon manager support gives limits allowing customers to manage a Cloud lifecycle. For instance, customers can add new sections to the Cloud, for instance, servers, amassing, databases, stores and so on. Customers can use a comparative interface to screen the sufficiency of the Cloud and play out various exercises. Conveyed figuring structure can be isolated into two fragments: the front end and the back end. Both of them are related with each other through a framework, when in doubt the Web. Front end is what the client (customer) sees; however, the back end is the cloudiness of the system, as found in Fig. 1. The front end suggests the client part of a disseminated registering system; it includes interfaces and applications that are required to get to the Cloud stage [26–28]. The organizations like online email programs impact existing Web programs like Internet Explorer or Firefox. Various structures have novel applications that give organize access to clients.

The back end suggests the Cloud itself, where it contains titanic data amassing, virtual machine, security part, organizations, sending models and servers. Seeing of traffic, dealing with the system and client demands are directed by a central server. It sticks to explicit gauges, i.e., shows and uses unprecedented programming called the middleware. The middleware licenses orchestrated PCs to talk with each other [29, 30]. There are two sorts of circulated figuring models, first, the services model, with three central organizations, SaaS (Soft as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service). The resulting one is deployment model that included three basic send applications on public, private, community or hybrid Clouds [31, 32].

4 Cloud Data Storages

Capacity administrations permit organizations and clients to store data on the limit devices of outcast providers. Disseminated stockpiling is open on the Web and acquainted with customers as a ton of limit pools or bowls, accessible using rich interfaces, for instance, programming APIs, Web interfaces, or request line devices. The multifaceted nature of dispersed stockpiling configuration is gotten away from clients, yet toward the back, it is truly eccentric. It is normally made out of scattered amassing devices that are regulated by concentrated programming. There are limitations on programming and computation platforms related to passing on data

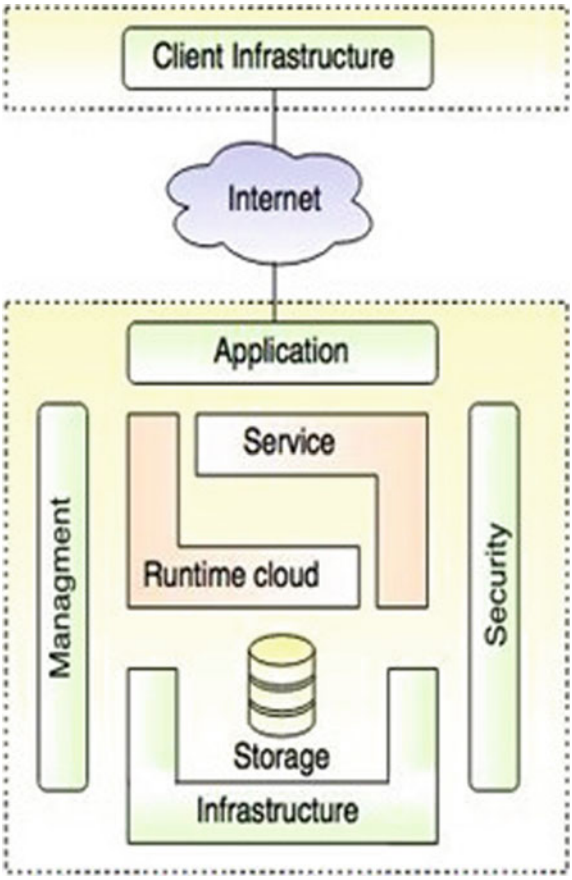


Fig. 1 Graphical view of Cloud computing architecture

from different connected devices. The IaaS is answerable for putting away information in mists [33]. Nonetheless, distributed storage may not meet each association's requirements. Potential detriments incorporate system dormancy, reliance on Web accessibility, security contemplations and constrained control. System idleness is higher contrasted with in house stockpiling on the grounds that the Cloud supplier's server farm is situated in an alternate geological area in any event a couple of systems jump from the customer's area. A customer who stores all information in an open Cloud and does not have a nearby duplicate is absolutely subject to Web availability. A Cloud supplier should offer elevated level security to keep away from data misfortune or bargain, and information move must be encoded. Attributes of average distributed storage are profoundly solid and repetitive, naturally adaptable, self-administration provisioning accessible for clients, available through rich interfaces (Web reassure, API, CLI) and pay-all the more just as expenses emerge portion model [34]. The Big data and disseminated processing are both are the snappiest moving advancements perceived on earth today. The Cloud enrolling is connected

with another perspective for the game plan of figuring establishment and gigantic data taking care of method for a wide scope of benefits [35]. Furthermore, some new Cloud-put together advances must be grasped with respect to the grounds that overseeing gigantic data for synchronous getting ready is inconvenient.

5 Cloud Service Provides

The Cloud service providers (CSP) are associations that offer framework organizations establishment or business applications in the Cloud. The colossal preferred position of using a Cloud expert center comes in capability and economies of scale. Rather than individuals and associations amassing their own system to support inside organizations and applications, the organizations can be purchased from the CSP, which offer the sorts of help to various customers from a typical establishment . The divisions, nevertheless, are not for each situation self-evident, a similar number of providers may offer various types of Cloud organizations; join standard Web or application encouraging providers. For example, for a firm which has started a Web encouraging association, getting Cloud provider, for instance, the Rackspace, and buy either PAAS or IAAS organizations are proposed. Many Cloud providers are focusing on unequivocal verticals, for instance, encouraging human administrations applications in an ensured IAAS condition. There are also trade-offs in the Cloud. As attempts move their applications and establishment to the Cloud, so they moreover give up control. Enduring quality and security are critical concerns. Various CSPs are focusing on giving raised degrees of organization and security, and PaaS and IaaS much of the time go with execution guarantees [36]. For model, Amazon a couple of colossal Cloud exercises helped cause to see the CSP in the mid 2000–2001. As CSPs have developed quickly and required new degrees of adaptability and the board they have largy affected figuring, stockpiling, and systems administration innovation. The CSP blast to a huge degree has driven interest for virtualization, in which equipment can be sectioned for access by various clients utilizing programming procedures. The development of the CSPs in the course of the most recent ten years has likewise determined probably the quickest development in innovation sections running from servers to switches and business applications.

6 Cloud Computing Technologies

There are various advancements that are employed after the distributed computing stage to make it solid, versatile and operational, and they are virtualization, grid registering, service-oriented architecture (SOA) and utility computing.

A. **Virtualization**: Virtualization is a technique that licenses sharing of a physical instance of advantage or an application among various customers or an affiliation. It does as such by giving out a reasonable name to a physical resource and giving a pointer to that physical resource when mentioned [37]. The primary use of this innovation is to give the applications a standard form to their Cloud customers. For instance, in the event that the refreshed rendition of the application is discharged, at that point, Cloud supplier should give the refreshed adaptation to their customers. For instance, VMware and Xen offer virtualized IT systems on demand. Virtual framework advances, for instance, virtual private network (VPN), bolster customers with a changed system condition to get Cloud assets. Virtualization strategies are the bases of Cloud registering since they render adaptable and adaptable equipment administrations. The multitenant design offers virtual separation among the different inhabitants, and along these lines, the associations can use and redo the application just as they individually have its personal specific illustration running, as seen in Fig. 2.

B. **Service-Oriented Architecture (SOA)**: Administration Oriented Computing presents and diffuses two significant ideas, which are additionally crucial for Cloud figuring, that is the Quality of Service (QoS) and Software as a Service (SaaS) [38] (Fig. 3).
Quality of Service distinguishes a lot of practical and non-useful traits that can be utilized to assess the conduct of an assistance from alternate points of view, and the Software as a Service presents another conveyance model for applications. It has been acquired from the universe of Application Service Providers (ASPs).

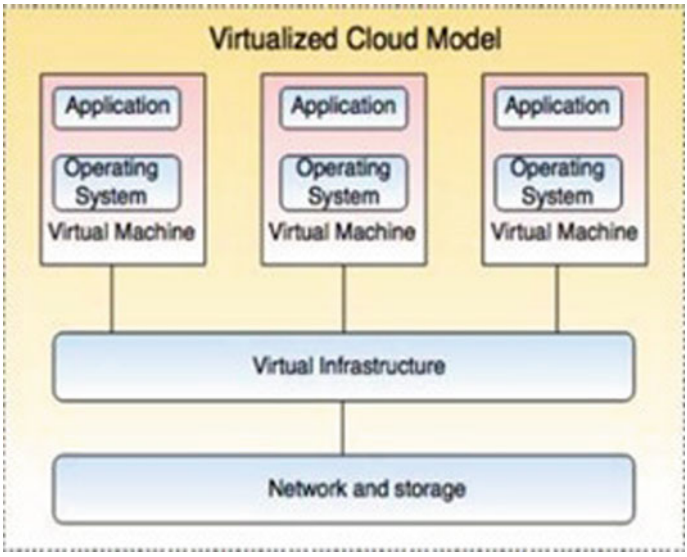


Fig. 2 Architecture of virtual Cloud model

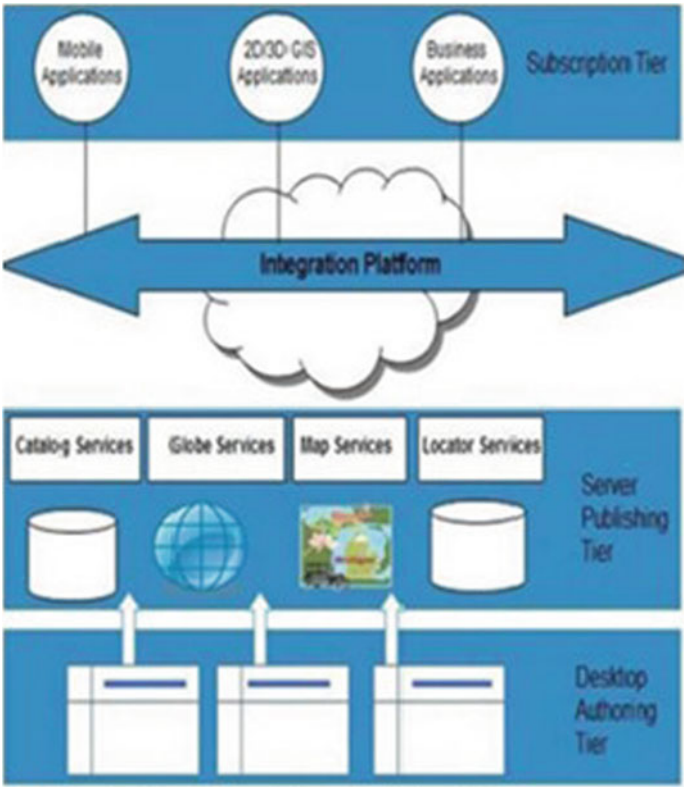


Fig. 3 Service-oriented architecture

C. **Grid computing**: The grid computing is characterized as circulated processing in which various PCs from numerous territories are associated together to accomplish a shared objective. The PC assets are extraordinary and topographically spread. In this manner, the matrix registering breaks a huge issue into smaller pieces. These smaller pieces are spread to frameworks that dwell inside the lattice. The lattice framework is proposed for sharing of assets through appropriated and enormous scope bunch figuring as found in observe Fig. 4. Lattice processing is mainstream in e-science, types of research that regularly require gigantic figuring force and joint effort between different information and processing administrations planning applications on grids can be a stunning task, especially while sorting out the flood of information over flowed enlisting assets. System work process structures have been made as a specific kind of a work strategy organization system arranged specially to make and execute a movement of a work procedure, or a computational or, data control steps, or in the grid arrangement.

A well-known grid computing venture is Folding@Home. The undertaking includes using unused registering forces of thousands of PCs to play out a

Fig. 4 Grid computing architecture

complex logical issue. The objective of the venture is "to comprehend protein collapsing, misfielding and related infections" [39].

D. **Utility Computing**: Utility figuring depends upon Pay-per-Utilize model. It gives computational assets to request as a metered advantage. All IT organizations, where grid processing, dispersed figuring, etc are used, they follow the idea of network registering. In these cases, evaluating on distributed computing is extremely unpredictable.

7 Comparative Analysis of Cloud Service Providers

There are many Cloud specialist co-ops accessible in the open market space to choose from, for example, Google, Microsoft, Amazon and so on. Huge numbers of them are monster names in IT. They offer various types of assistance at various expense. We will examine just three famous Cloud service providers, to be specific (I) Google Cloud Platform (ii) Microsoft Azure (iii) Amazon Web Services [40].

(i) **Google Cloud Platform**: Virtual machine movement is not upheld by Google; however, computer-generated machines can be relocated to Google Cloud by outsider suppliers like cohesive networks. Consequence of crude gadget representing pictures is bolstered by Google. Google bolsters is an auto-scalar that balances remaining burden with more or short requests.

(ii) **Microsoft Azure**: As Microsoft Azure is based upon Windows server, it is anything but difficult to move computer-generated mechanisms on open Cloud on account of numerous comparative programming and applications among neighborhood server farms and Microsoft Azure. Custom picture development is likewise a simple errand utilizing virtual hard disks. For scaling reason, Microsoft has a balance page in Azure interface that is taken care of physically or naturally. Purplish blue can have virtual systems in it, and it can likewise be associated with corporate system utilizing VPN.

(iii) **Amazon Web Services**: Amazon Web services (AWS) are as of now helpful in 190 nations in excess of a million clients with organizations, non-benefit associations and legislative associations. It is well known for providing a flexible, severely versatile and most cost-effective strategy to convey their sites and Web applications. AWS proposes the most consistent relocation way to association to transport computer-generated machines from the associations on premise hypervisor to open Cloud and the other way around, when required.

8 Conclusions

In this review paper, an endeavor is made to survey distributed computing foundation and its definitions alongside structures, models. In spite of the fact that each distributed computing stage has its own quality, one thing ought to be seen is that regardless of what sort of stage, there are heaps of unsolved issues. Some normal difficulties are connected with security and protection. The paper expects to give a method for keening the model and finding alternatives accessible for supplementing your innovation and association needs.

References

1. Josep AD, Katz R, Konwinski A, Gunho LEE, Patterson D, Rabkin A (2010) A view of cloud computing. Commun ACM 53(4)
2. yRon Gill CMA, C. F. M. (2011) Why cloud computing matters to finance. Strateg Finance 92(7):43
3. Priyadarshi R, Soni SK, Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-13-0776-8_27
4. Priyadarshi R, Gupta B (2019) Coverage area enhancement in wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-019-04674-y

5. Priyadarshi R, Singh L, Randheer, Singh A (2018) A novel HEED protocol for wireless sensor networks. In 2018 5th International Conference on Signal Processing and Integrated Networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474286
6. Moreno-Vozmediano R, Montero RS, Llorente IM (2012) Iaas cloud architecture: from virtualized datacenters to federated cloud infrastructures. Computer 45(12):65–72
7. Alamri A, Ansari WS, Hassan MM, Hossain MS, Alelaiwi A, Hossain MA (2013) A survey on sensor-cloud: architecture, applications, and approaches. Int J Distrib Sens Netw 9(2):917923
8. Priyadarshi R, Yadav S, Bilyan D (2019) Performance and comparison analysis of MIEEP routing protocol over adapted LEACH protocol. In Luhach AK, Hawari KBG, Mihai IC, Hsiung P-A, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer Singapore, pp 237–245. https://doi.org/10.1007/978-981-13-6295-8_20
9. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
10. Priyadarshi R, Gupta B, Anurag A (2020a) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. J Supercomput. https://doi.org/10.1007/s11227-020-03166-5
11. Priyadarshi R, Singh L, Singh A, Thakur A (2018) SEEN: stable energy efficient network for wireless sensor network. In 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474228
12. Lin Y, Shao L, Zhu Z, Wang Q, Sabhikhi RK (2010) Wireless network cloud: Architecture and system requirements. IBM J Res Dev 54(1):4–1
13. Weinhardt C, Anandasivam A, Blau B, Borissov N, Meinl T, Michalk W, Stößer J (2009) Cloud computing–a classification, business models, and research directions. Bus Inf Syst Eng 1(5):391–399
14. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
15. Priyadarshi R, Bhardwaj A (2017) Node non-uniformity for energy effectual coordination in WSN. Int J Inf Technol Secur
16. Priyadarshi R, Yadav S, Bilyan D (2019) Performance analysis of adapted selection based protocol over LEACH protocol. In Luhach AK, Hawari KBG, Mihai IC, Hsiung P-A, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 247–256. https://doi.org/10.1007/978-981-13-6295-8_21
17. Priyadarshi R, Gupta B, Anurag A (2020b) Wireless sensor networks deployment: a result oriented analysis. Wirel Pers Commun. https://doi.org/10.1007/s11277-020-07255-9
18. Rawat P, Chauhan S, Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circ Syst Comput. https://doi.org/10.1142/S0218126620502047
19. Subashini S, Kavitha V (2011) A survey on security issues in service delivery models of cloud computing. J Netw Comput Appl 34(1):1–11
20. Ren K, Wang C, Wang Q (2012) Security challenges for the public cloud. IEEE Internet Comput 16(1):69–73
21. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
22. Randheer, Soni SK, Kumar S, Priyadarshi R (2020) Energy-aware clustering in wireless sensor networks. In: Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-15-2854-5_39
23. Kumar S, Soni SK, Randheer, Priyadarshi R (2020) Performance analysis of novel energy aware routing in wireless sensor network. In: Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-15-2854-5_44
24. Priyadarshi R, Tripathi H, Bhardwaj A, Thakur A (2018) Performance metric analysis of modified LEACH routing protocol in wireless sensor network. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i1.5.9146

25. Spoorthy V, Mamatha M, Kumar BS (2014) A survey on data storage and security in cloud computing. Int J Comput Sci Mob Comput 3(6):306–313
26. Yang K, Jia X (2012) An efficient and secure dynamic auditing protocol for data storage in cloud computing. IEEE Trans Parallel Distrib Syst 24(9):1717–1726
27. Buyya R, Beloglazov A, Abawajy J (2010) Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. arXiv preprint arXiv:1006.0308
28. Amazon, Elastic compute cloud (EC2). https://aws.amazon.com/ec2/, January 2009
29. Sunitha K (2014) Virtualization components. In A survey on securing the virtual machines in cloud computing, vol 1 Issue 4, Int J Innov Sci Eng Technol. ISSN 2348-7968
30. Buyya R, Vecchiola C, Thamarai Selvi S (2013) Mastering cloud computing. McGrawHill, New Delhi
31. Priyadarshi R, Singh L, Kumar S, Sharma I (2018) A hexagonal network division approach for reducing energy hole issue in WSN. Eur J Pure Appl Math
32. Anurag A, Priyadarshi R, Goel A, Gupta B (2020) 2-D coverage optimization in WSN using a novel variant of particle swarm optimisation. In 2020 7th International Conference on Signal Processing and Integrated Networks, SPIN 2020. https://doi.org/10.1109/SPIN48934.2020.9070978
33. Priyadarshi R, Singh L, Sharma I, Kumar S (2018) Energy efficient leach routing in wireless sensor network. Int J Pure Appl Math 118(20):135–142
34. Ahmed M, Chowdhury ASMR, Ahmed M, Rafee MMH (2012) Grid computing in an advanced survey on cloud computing and state-of-the-art research issues. IJCSI-Int J Comput Sci 9(1)
35. Dillon T, Wu C, Chang E (2010) Cloud computing: issues and challenges. In 2010 24th IEEE international conference on advanced information networking and applications, pp 27–33. IEEE
36. Jadeja Y, Modi K (2012) Cloud computing-concepts, architecture and challenges. 2012 international conference on computing, electronics and electrical technologies (ICCEET). IEEE
37. Ferris JM (2018) U.S. Patent No. 9,930,138. U.S. Patent and Trademark Office, Washington, DC
38. Odun-Ayo I, Ananya M, Agono F, Goddy-Worlu R (2018) Cloud computing architecture: a critical analysis. In: 2018 18th international conference on computational science and applications (ICCSA), pp 1–7. IEEE
39. Markova O, Semerikov SO, Striuk AM, Shalatska HM, Nechypurenko PP, Tron VV (2019) Implementation of cloud service models in training of future information technology specialists. In: Proceedings of the 6th Workshop on Cloud Technologies in Education (CTE 2018), Kryvyi Rih, Ukraine, December 21, 2018, No 2433, pp 499–515. CEUR Workshop Proceedings
40. Joshi N, Shah S (2019) A comprehensive survey of services provided by prevalent cloud computing environments. In: Smart intelligent computing and applications (pp. 413–424). Springer, Singapore

High Throughput Novel Architecture of SIT Cipher for IoT Application

Zeesha Mishra, Shubham Mishra, and Bibhudendra Acharya

Abstract The advancement in network connectivity and data handling capabilities shows the tremendous growth of Internet of things (IoT) in physical life. The number of connected devices has been increasing in IoT applications, left severe security concerns. The problem of providing security solutions to resource-constrained devices leads to lightweight cryptographic algorithms. Secure IoT (SIT) is one of the lightweight cryptographic algorithm follows SPN and Feistel structure. SIT cipher works on 64-bit block and 64-bit key size. In this chapter, an efficient pipelined architecture of SIT lightweight block cipher has been proposed. Operating frequency of this proposed work has been improved by 17% and also throughput improved by more than 180% in comparison with LEA lightweight block cipher.

Keywords Cryptography · IoT · SIT · Lightweight · Block cipher

1 Introduction

The Internet of things (IoT) is a global network infrastructure of interconnected heterogeneous objects which envisions a future of the world with physical entities and virtual components that can be linked to enable a whole new class of applications and services [1]. The expansion in the computing capacities of the smart devices has fostered an ecosystem of connected physical objects that are accessible through the Internet. IoT innovations are at new born stage, growing at a tremendous pace and numerous technology improvements have happened within the integration of objects

Z. Mishra · S. Mishra · B. Acharya (✉)
Department of Electronics and Communication, National Institute of Technology, Raipur, Raipur 492010, India
e-mail: bacharya.etc@nitrr.ac.in

Z. Mishra
e-mail: zmishra.phd2016.etc@nitrr.ac.in

S. Mishra
e-mail: shubhammishra.0709@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_26

can bring convenience to the physical world [2]. Algorithms, which are useful to protect the resource-constrained IoT devices are falls in the field of lightweight cryptography [3].

A less complex lightweight algorithm called secure IoT (SIT) [4] was developed to encrypt and decrypt data. It is a 64-bit block cipher and uses 64-bit key to encrypt the data. SIT algorithm is based on both Feistel and substitution and permutation network (SPN) modes of approach. Hardware implementation tends to provide an efficient and reliable security as compared to software implementation [5]. There persists a trade-off between area, security and performance in designing any lightweight cryptographic algorithm.

1.1 Contribution

Novel pipelined architecture of SIT lightweight block cipher has been proposed in this chapter. These provide high throughput and good efficiency. Pipeline registers are given in the minimal critical path for increase the efficiency. Xilinx software has been used to implement this FPGA design.

1.2 Organization of Paper

This chapter organized in five sections. Following the introduction, SIT lightweight encryption algorithm presented in Sect. 2. Proposed hardware implementation of SIT lightweight block cipher explained in Sect. 3. Results and compression are presented in Sect. 4. Concluding remarks are outlined in Sect. 5.

2 SIT Lightweight Encryption Algorithm

SIT cipher is developed to encrypt and decrypt data. SIT is a symmetric lightweight block cipher with 64-bit key and 64-bit block sizes [4]. This lightweight block cipher designed for IoT application. To increase the complexity, SIT lightweight block cipher uses substitution and permutation network (SPN) and Feistel structure. SIT algorithm provides sufficient security and uses only five rounds. This is the biggest advantages of SIT, because some lightweight block cipher like AES [6], HIGHT [7], PICCOLO [8], LILLIPUT [9], LEA [10], QTL [11], XTEA [12], TEA [13] and Klein [14] uses more number of clock cycles as compared to SIT. SIT cipher consists of two modules and those are key schedule process and encryption process. Key scheduling process is round-based and consists of mathematical function. Operations performed in this section such as matrix concatenation, XOR, addition and shifting.

Table 1 Notation of the algorithm

Notation	Description
p	Plaintext
C_{key}	Cipher Key
F	Function
Kaif	16 bits of kbif to the function
$\oplus$	R
$\odot$	OR

Table 2 S and T values

Kkj	0	1	2	3	4	5	6	7
S1, S(Kkj)	3	F	E	0	5	4	B	C
S2, T(Kkj)	0	E	5	6	A	2	3	C
Kkj	8	9	A	B	C	D	E	F
S1, S1(Kkj)	D	A	9	6	7	8	2	1
S2, S2(Kkj)	F	0	4	D	7	B	1	8

Similarly, encryption process is diffusion of round-based key with plaintext. Some notations which have been used in key scheduling and encryption process are shown in Table 1. In key scheduling process of SIT uses S and T tables to increase the complexity of the algorithm. This is presented in Table 2. SIT lightweight block cipher's cipher [4] key scheduling and encryption process as proposed by M. Usman, et al. described further in this section.

1. Key Scheduling Key scheduling process involves the division of 64-bit plaintext into 4 parts and after that in this way by concatenating each part produces 4 blocks. This 16-bit data will be given to the F-function, and F-function works on the basis of S and T values, which are shown in Table 2. After the process of F-function, it provides 4 × 4 matrix and gives 16-bit values. This 16-bit value performs unique concentration operation and generates 4 keys. The key names are Kk1, Kk2, Kk3, Kk4 and XOR of all of them give Kk5.

Algorithm 1

Input: 64bit Key

Output: 64 bit K1 to K5

1. $C_k \leftarrow 64\ bits$
2. $F \leftarrow 16\ bit\ Data$
3. $Ka_i f \leftarrow f(Ka_i f)$
4. $for\ (i = 0; i \leq 4; i++)$
5. $K_i = Ka_i f;$
6. $K5 = k5 \oplus Ka_i f$

2. Encryption Process Encryption process starts after key generation process. This process uses some logic operations like XOR and XNOR for more confusion diffusion. To complete five rounds, it employs shifting, swapping substitution and permutation. This encryption process uses 64-bit plaintext directly. 64-bit plaintext divided into 4 parts and each part of plaintext (16-bit) forwarded to first round function. Swapping, shifting and XNOR operation also performed in first round function.

Similar operations performed in 3 more rounds. Finally, in last round, similar operations performed expect swapping operation of 16-bit blocks.

Algorithm 2

Input: 64bit Key

Output: 64 bit K1 to K5

1. $for\ j = 0\ to\ 5$ {
2. $for\ i = 0\ to\ 3$ {
3. $P_{X_{i,j}} \leftarrow P_{X_i} \odot K_j;$
4. $P_{X_2} \leftarrow P_{X_2} \oplus f(P_{X_1} \odot K_j)$
5. $P_{X_3} \leftarrow P_{X_3} \oplus f(P_{X_4} \odot K_2)$
6. $P_{X_4} \leftarrow P_{X_4} \odot K_j;$

}

}

3 Hardware Implementation of Sit Lightweight Block Cipher

SIT algorithm architecture divided into two parts first one is key generation process and second is encryption process and this are explained further in this section.

Key Generation Architecture Key generation process is the generation of key for different rounds of encryption process. In each round, key scheduling process encodes key and provides it to the encryption process for further processing. Key scheduling process starts with a 64-bit input key value which goes through key expansion block to produce four different 16-bit blocks for concatenation operations. These values of 16-bits have been shown in Fig. 1 as Kb1f to Kb4f. F-function consists of S and T whose values are shown in Table 2 and these provides confusion in algorithm. Output of F-function is a 4 × 4 matrix (16 elements) that gives Ka1f to Ka4f (16-bit) values for key generation input as shown in Fig. 1. Key generation comes up with four middle keys as Kk1, Kk2, Kk3 and Kk4. There is a fifth key, which is XOR of all middle keys.

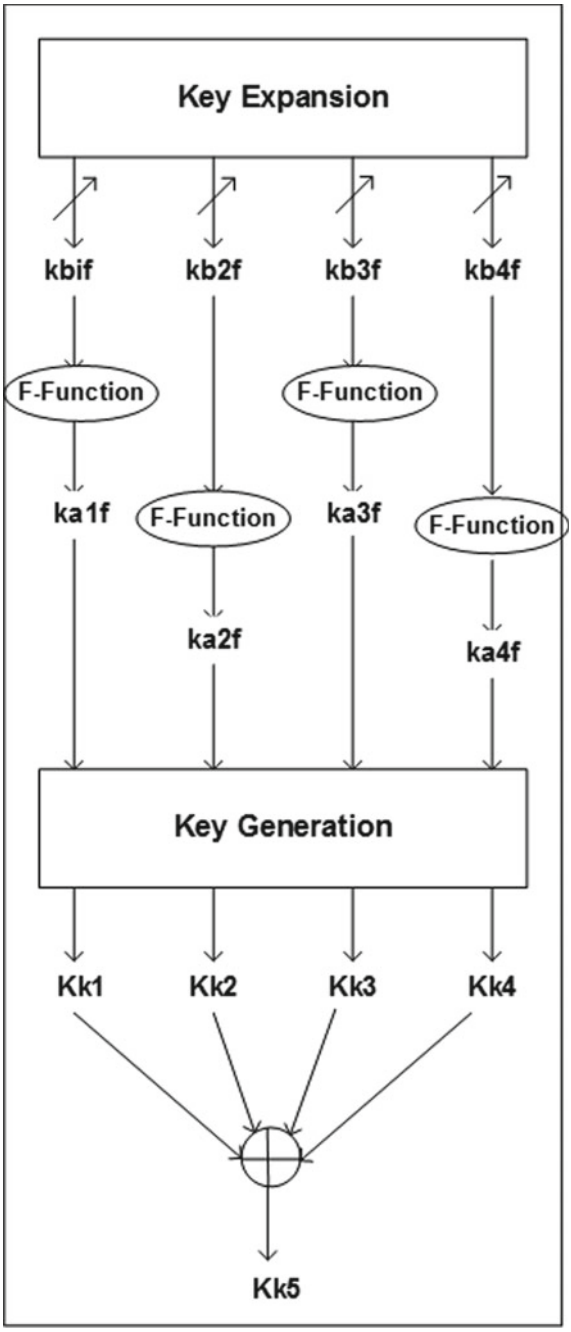


Fig. 1 Key generation architecture for SIT encryption algorithm

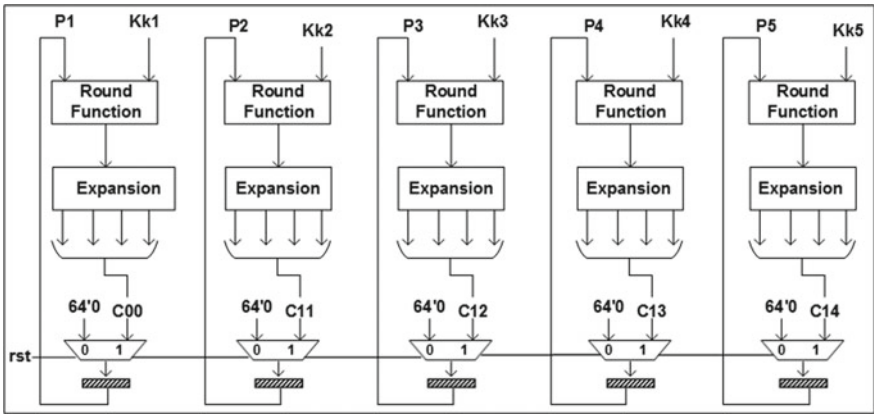


Fig. 2 Architecture for SIT encryption algorithm

Encryption Architecture Encryption process of SIT is consists of round function, expansion and multiplex operations. Input provided to encryption process is 64-bit plaintext data. This input goes through round function operation and passes through bit expansion process. Each round goes by these provided operations and at last selection process has been done by multiplexer. This multiplexer is synchronized by reset signal and issues output at the same time for each round. Whatever results are stored in the P5 at the end of the fifth round, is ciphertext.

In the encryption process of the architecture uses five pipeline registers. It will increase the efficiency of architecture. Each round generates 64-bit middle encrypted plaintext, as we know that there are five rounds in SIT algorithm. Hence, after fifth round whatever result will be stored in pipeline registers that are ciphertext (Fig. 2).

4 Results and Comparison

The proposed architectures implemented in Verilog hardware description language (HDL) Xilinx ISE for FPGA implementation. FPGA is a low cost, high speed and flexible functional programmable logic device that can be configured for a wide variety of applications. The hardware utilization of proposed architecture results using virtex-7 family xc7vlx330t device tabulated in Table 3.

Proposed design is focused on high speed IoT applications and the synthesis tools were also fixed based on the design goals optimization. The results of various lightweight block ciphers are compared with proposed architecture by using Table 4, which presents FPGA implementation of block ciphers. This comparison is made based on various hardware performances in terms of area performance and speed performance. In this chapter, we compared proposed results with some lightweight cryptographic algorithms like LEA [15], CLEFIA [16], KATAN [17], AES [18],

Table 3 FPGA implementation of SIT algorithm in XC7VLX330TT

S. No.	Parameter	Results
1.	No. of LUTs	760
2.	No. of flip flop	128
3.	No. of slices	345
4.	Latency	5
5.	Max. frequency (Mhz)	479
6.	Throughput (Mbps)	6131.2
7.	Throughput per area	17.77
8.	Static power	0.177
9.	Dynamic power	0.008
10	Total power	0.185

Table 4 FPGA implementation results of block cipher

Algorithm	Device name	Block size	LUT	FFs	Slices	Max. frequency	Throughput	Throughput per area
LEA [15]	XC5VLX50	128	360	382	122	225.038	1200.20	9.83
CLEFIA [16]	XC2VP7	128	1944	409	1040	50.42	140.29	0.134
KATAN [17]	XC35200	64	–	–	1054	–	–	0.23
AES [18]	XC5VLX85				28598	644.33	–	2.88
Hummingbird [19]	XC3S200-5	16	473	120	273		160.4	0.59
[20]	XC6VLX130T	–	10160	2052	3144	127	–	11.82
This Work	**XC3S200-5**	**64**	**1072**	**464**	**550**	**123.34**	**1578.75**	**2.87**
This Work	**XC5VLX50**	**64**	**1066**	**383**	**422**	**265.97**	**3404.41**	**8.06**
This Work	**XC6VLX130T**	**64**	**928**	**287**	**305**	**443.95**	**5682.56**	**18.63**
This Work	**XC5VLX85**	**64**	**1066**	**382**	**416**	**286.01**	**3672.96**	**9.1**

Hummingbird [19], and [20]. LEA having more block size and simple architecture but this proposed work gives 17% higher frequency in compression LEA [15]. From the compression table, KATAN [17] throughput is less as compared to proposed work with the same platform. Hummingbird [19] is also a small size block cipher but its throughput is less as compared to proposed work of this chapter. AES is complicated block cipher uses complex S-boxs. AES [18] slices are more as compared to proposed. CLEFIA [16] is a 128-bit block cipher and it more complicated block cipher having more slices are more as compared to the proposed work. Graphical representation of the maximum frequency on different FPGA devices is presented in Fig. 3.

Equation (1) presents the calculation of throughput.

$$\text{Throughput} = \frac{\text{Max. Frequency} \times \text{Block Size}}{\text{Latency}} \tag{1}$$

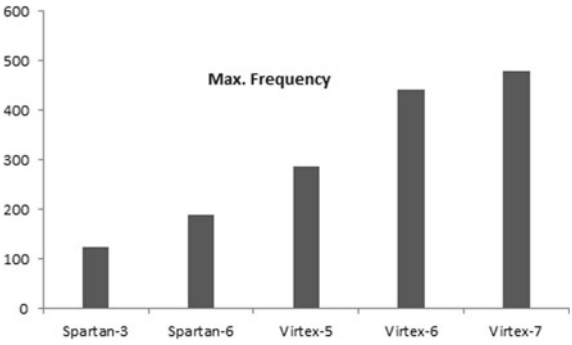


Fig. 3 Graph presentation of maximum frequency in different FPGA devices

Equation (2) represents throughput per area (throughput/area) and for FPGA implementation area represents number of slices.

$$\text{Throughput/Area} = \frac{\text{Throughput}}{\text{Area in slice}} \tag{2}$$

Throughput and throughput per area are two important parameters for pipelined architecture, and this represents the efficiency of the architecture. It depends on maximum frequency, block size, latency and slices.

5 Conclusion

This chapter presented novel pipelined architecture of SIT lightweight block cipher. Use of pipeline registers in the proposed design provides batter throughput. FPGAs implementation results of lightweight block cipher compared with the proposed work. Operating frequency of this proposed work has been improved by 17% and also throughput improved by more than 180% in comparison with LEA lightweight block cipher.

Acknowledgements This work has been carried out under Information Security Education Awareness (ISEA) project phase – II & SMDP-C2SD project funded by Ministry of Electronics and Information Technology (MeitY), Govt. of India in the Department of Electronics and Communication Engineering at National Institute of Technology Raipur, India. Authors are thankful to the Ministry for the facilities provided under this project.

References

1. Xu T, Wendt JB, Potkonjak M (2015) Security of IoT systems: design challenges and opportunities. IEEE/ACM International Conference on Computer-Aided Design. Dig. Tech. Pap. ICCAD, vol 2015, pp 417–423

2. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of Things (IoT): a vision, architectural elements, and future directions. Futur Gener Comput Syst 29(7):1645–1660
3. Singh P, Acharya B, Chaurasiya RK (2019) A comparative survey on lightweight block ciphers for resource constrained applications. Int J High Perform Syst Archit 8(4):250–270
4. Usman M, Ahmed I, Imran M, Khan S, Ali U (2017) SIT: a lightweight encryption algorithm for secure internet of things. Int J Adv Comput Sci Appl 8(1):1–10
5. Mohd BJ, Hayajneh T, Vasilakos AV (2015) A survey on lightweight block ciphers for low-resource devices: Comparative study and open issues. J Netw Comput Appl 58:73–93
6. Banu JS, Vanitha M, Vaideeswaran J, Subha S (2013) Loop parallelization and pipelining implementation of AES algorithm using OpenMP and FPGA. 2013 IEEE international conference emerging trends in computing, communication and nanotechnology, ICE-CCN 2013, pp 481–485
7. Hong D et al (2006) HIGHT: a new block cipher suitable for low-resource device. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 4249 LNCS, pp 46–59
8. Shibutani K, Isobe T, Hiwatari H, Mitsuda A, Akishita T, Shirai T (2011) Piccolo: an ultra-lightweight blockcipher. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 6917 LNCS, pp 342–357
9. Berger TP, Francq J, Minier M, Thomas G (2016) Extended generalized Feistel networks using matrix representation to propose a new lightweight block cipher: lilliput. IEEE Trans Comput 65(7):2074–2089
10. Lee D, Kim DC, Kwon D, Kim H (2014) Efficient hardware implementation of the lightweight block encryption algorithm LEA. Sensors (Switzerland) 14(1):975–994
11. Li L, Liu B, Wang H (2016) QTL: a new ultra-lightweight block cipher. Microprocess Microsyst 45:45–55
12. Kaps JP (2008) Chai-tea, cryptographic hardware implementations of xtea. In: International conference on cryptology in India Springer, Berlin, pp 363–375
13. Wheeler DJ, Needham RM (1995) Tea, a tiny encryption algorithm. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 1008, pp 363–366
14. Gong Z, Nikova S, Law YW (2011) KLEIN : a new family of lightweight block ciphers. In: International workshop on radio frequency identification: security and privacy issues. Springer, Berlin, pp 1–18
15. Mishra Z, Ramu G, Acharya B (2019) Hight speed low area VLSI architecture for LEA encryption algorithm. In: Proceedings of the third international conference on microelectronics, computing and communication systems, pp 155–160. Springer, Singapore
16. Sekhar A (2017) Design of a new lightweight encryption for embedded security. Int Res J Eng Technol 4(6):398–402
17. De Canni C (2009) "KATAN & KTANTAN—a family of small and efficient hardware-oriented block ciphers. In: International workshop on cryptographic hardware and embedded systems, pp 272–288. Springer, Berlin
18. Soltani A, Sharifian S (2015) An ultra-high throughput and fully pipelined implementation of AES algorithm on FPGA. Microprocess Microsyst 39(7):480–493
19. Engels D, Fan X, Gong G, Hu H, Smith EM (2010) Hummingbird : ultra-lightweight cryptography for resource-constrained devices. In: International conference on financial cryptography and data security, pp 3–18. Springer, Berlin
20. Imran M, Rashid M, Shafi I (2018) Lopez Dahab based elliptic crypto processor (ECP) over GF(2163) for low-area applications on FPGA. 2018 International Conference on Engineering Emerging Technologies. ICEET 2018, vol 2018, pp 1–6

Design and FPGA Synthesis of an Efficient Synchronous Counter with Clock-Gating Techniques

Saurabh Kumar Singh, Mangal Deep Gupta, and R. K. Chauhan

Abstract In the proposed work, we focused on clock-gating-based synchronous counter. This paper depicts the designing of high-speed synchronous counter with low dynamic power dissipation using clock-gating method. We study the various design technique to overcome the dynamic power dissipation in the synchronous circuit. After analysing the several techniques, we focused on clock-gating technique to minimize the dynamic power dissipation and compare the proposed counter to the previous counter in terms of latency, on-chip power dissipation, utilized area, and maximum operating frequency. The clock-gating technique enables for an improvement of 24, 36, and 31%, respectively, for latency, area, and maximum operating frequency as well as maintaining the on-chip power dissipation as the prior synchronous counter. The design proposal of 4-, 8-, and 16-bit synchronous counter is built by Verilog HDL code and synthesis is carried out with Spartan 3 FPGA on ISE design suit 14.2 Tool.

Keywords Synchronous counter · Clock gating · Latency · On-chip power dissipation · Verilog HDL · Spartan-3 FPGA

1 Introduction

A counter is a special type of register that goes through a pre-determinant sequence of state. As the name offer, counter enforces counting such as time and electronics pulses. Counters are extensively used in processor, calculator, real-time clock, etc.;

S. K. Singh (✉) · M. D. Gupta · R. K. Chauhan
Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, India
e-mail: srbhsingh448@gmail.com

M. D. Gupta
e-mail: mangaldeepgcet@gmail.com

R. K. Chauhan
e-mail: rkchauhan27@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_27

in modern technology, high-speed counters are simulated which is used in machine moving control, digital clock, etc. Initially, counters are two types, synchronous and asynchronous; clock is the fundamental distinction between both of them. In asynchronous counter, the flip-flop is triggered either by the clock pulse or by the output of neighboring flip-flop, but in case of synchronous counter, all the flip-flop are triggered by an ordinary clock pulse.

In modern automation technology, some event is very fast which cannot be detected in the program cycle. To detect such high-speed event, introducing a new technique term as high-speed counter (HSC). HSC is advantageous when determining the speed of rotary motion in case of only one or few pulses per rotation, and a part of this HSC is applicable at automation and process control, programmable logic controller, motor and stepper drives, and motion control application. In designing of minimum dynamic power and highly reliable synchronous counter, there are two crucial factors occurring: first one, in synchronous counter, all the flip-flops are triggered by a common clock pulse and clock consumes maximum power and second circuitry of synchronous counter is very complex and it requires connection oriented protocol.

Design of minimum power and high performance synchronous counter is crucial factor; Minimum power is serious issue in recent time, because nowaday's designer uses high frequencies and small chip size. Dynamic power dissipation factor has become crucial not only from the point of perspective of reliability but they have assumed greater significance by the advent of portable battery-driven device [1, 2]. In designing of low power consumption and high-speed synchronous circuit, particularly as battery-driven circuitry, clock-gating technique has been broadly studied and designed. It efficiently reduces dynamic power dissipation and enhances the speed in various synchronous circuits because clock gating eliminates unnecessary switching activity when the circuit is in active operation mode. Clock gating is working on the principle of adding more logic circuit to prune the clock tree. Designing of lower dynamic power and highly reliable synchronous circuit, clock is a crucial factor because maximum dynamic power dissipation in any synchronous circuit is due to clock. It is observed that clock contains almost 40% of total power [3]. We must take effort to reduce the clock power reduction like voltage swing reduction, buffer insertion, and clock routing. In synchronous circuit clock activity, consume unnecessary power to overcome power consumption reducing unwanted clock switching [4]. In clock-gating terminology, cut the clock of disable portion of the circuitry at result flip flop in disable portion do not have in conducting state only active state of flip flop consume power by the result dynamic power consumption going to zero [1]. Dynamic power dissipation is caused by switching activities of the circuits and short circuit current and the clock gating reduces the switching activity of the synchronous circuit. Two kinds of clock gating: latch-based clock gating and latch free clock gating. Latch-based clock gating is edge sensitive; on the other hand, latch free clock gating is level sensitive. This is the fundamental differences between both the techniques.

Clock-gating is a advanced technique to design synchronous circuit with minimum dynamic power dissipation.

In literature, there are several power efficient methods introduced to overcome the dynamic power dissipation of the synchronous circuit at different level. Hossain et al. [3] proposed double edge trigger technique to overcome the dynamic power dissipation. Conventional flip-flop is sensitive only at rising edge or falling edge of the clock. Therefore, only 50% impact is used and rest part of clock is contributing in power dissipation. A double-edge-triggered flip-flop can therefore be used to attain energy saving. However, double-edge trigger is slightly slower and more complex than single edge trigger flip-flop. Wu et al. [4] proposed a method known as quasi-synchronous derived clock. In this method, the master clock originates the synchronous clock that block flip-flop in the circuit from the unwanted switching activity, but the disadvantage of this method, given the circuitry a enormous delay. Keote et al. [5] proposed 3-bit binary up counter using modified quasi-state energy recovery logic. In this work, use two-phase clock supplies and 180-degree phase shift signal to overcome the circuit switching activity. This paper improves the drawback of [4] quasi-synchronous derived clock sequential circuit. This technique is more power efficient and faster than [4]. Pedram et al. [6] presented minimum power counter using priority encoding, in case of state assignment, k state variables are used to convey 2^k distinct states. However, if the number of functional states l is not equal to 2^k, i.e., $l \leq 2^k$, then there will exist $(2^k - l)$ unnecessary states. To save the power, remove the unused state code, using priority encoding. This technique is restricted with its complex circuitry because it has needed external priority encoder circuit. Gautam et al. [7] proposed a mode-10 counter using an improved clocked adiabatic logic circuit (ICAL). Adiabatic logic improves the synchronous circuit by means of recycling the energy that is going too dissipated in the form of heat. Adiabatic logic improves the circuit in term of saving the energy but it has drawback of higher supply voltage and lower speed of operation. Katreepalli et al. [1] proposed a new power efficient and highly reliable counter circuit with clock-gating technique. According to the author optimization of the circuit depending on its architecture. Optimization of clock-driven circuit such as synchronous circuit, in terms of low power utilization and highly reliability is very crucial factor because 50% of power dissipation within the clock. In this paper, the author proposed a powerful synchronous counter design based on the clock-gating technique. This design reduces the drawback of [7], and improves the synchronous counter in term of low power dissipation and high reliability. Several literatures suggested the application based on clock-gating technique. Using clock-gating technique, design ALU for memory application, 4-bit Johnson counter with power gating technique based on two-phase CPAL circuit [2, 8]. Power saving methods in hardware models have become critical in latest years, particularly for mobile device. A double-edge-triggered transmission-gate flip-flop with a clock-gating feature reduces the 33.14% power on average when it compares to previous work on edge trigger flip-flop [9]. Papers [10–12] define the different clock gating method to decrease the power dissipation. These papers discuss the distinct clock-gating technique that can be used to optimize the VLSI circuit at RTL level and the various problem connected with the implementation of these RTL-level power optimization techniques. Paper [13–15] suggested the merits of FPGA and described various application of FPGA. According to literature, FPGA provides the

opportunity to develop versatile, customizable, and low cost alternatives, therefore, likely application range from space application.

After analysing the several techniques to decrease the dynamic power of synchronous counter circuit [16–20], we introduce clock-gating technique in our proposed design. In our proposed work, we are interested on FPGA design technique. The FPGA design in between PLD and ASICs because their functionality can be built in the field like PLD, and FPGA implemented externally large and complex logic circuit and contains millions of logic gates like ASICs. Designing process of FPGA is very fast because it has not included several designing steps like layout, masking, etc. FPGA is ready for burning our Verilog code. FPGA designing eliminates the time-consuming floor-planning, place, route, and timing analysis. FPGA as the name suggests, that field programmable gate array, it has ability to reprogram the current action. This designing technique is good for prototype device and synthesis is much easier than ASIC. FPGA is applied in signal processing system, image processing, and enhancement and used in ASIC emulation and verification.

In suggested work, we design an advance synchronous counter based on clock-gating technique to achieve higher speed and less power dissipation as compared to counter [1]. In proposed job, we design 4-, 8-, and 16-bit synchronous counters and synthesis with Spartan 3 FPGA on ISE design suit 14.2 tool. Further, compare the proposed synchronous counter in terms of on-chip power dissipation, latency, maximum operating frequency, and area with counter [1] and examine the on-chip power dissipation at various frequencies.

2 Previous Synchronous Counter

In order to reduce the dynamic power dissipation of the synchronous counter, a new design proposed in [1]. The suggested 4-bit low power synchronous counter design is shown in Fig. 1. This counter is quite similar to traditional counter but it is working on the principle of clock-gating technique to decrease the dynamic power dissipation. In this design, [1] used T flip-flop in toggle mode, flip-flop toggle at positive edge. In this design, clock buffer network is introduced, that is built using sequence of repeater and this is helpful for removing the clock skew with minimal effort and distributing the clock to the network. In [1], AND gate followed by N-MOS transistor works as clock gating to the network. To design n-bit counter n, T flip-flop are used in toggle mode, and LSB flip-flop directly synchronized with master clock through clock buffer network. LSB flip-flop always toggle at each period of clock, the adjacent flip-flop toggle if the desire state of the LSB flip-flop is logically high then control N-MOS transistor passes the clock signal, otherwise block the clock signal to prevent the unwanted switching activity. Further, next flip flop toggle when both the previous flip-flop output states are logically high, on the other hand if any previous flip-flop output state are logically low then the clock of the next flip-flop going to block to prevent the unwanted switching activity. In [1], maximum power consumption is due to clock signal; it consumes on average 25–40% of total power. In

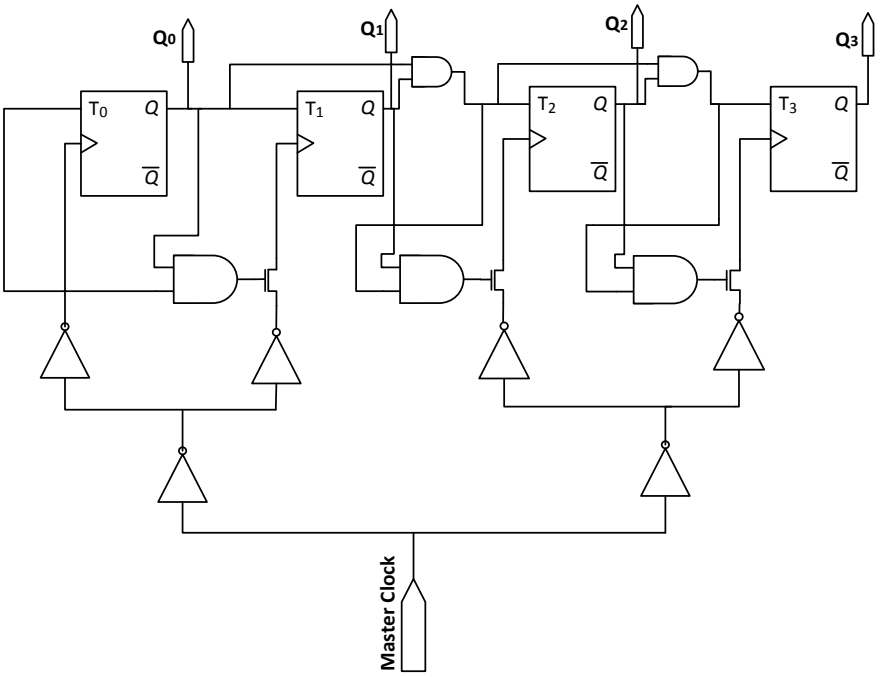


Fig. 1 4-bit low power synchronous counters [1]

synchronous counter clock forces, the entire flip-flop changes their state. If the next state of the flip-flop is the same as the previous state during the switching operation, this flip-flop in holding mode, so this type of activity of the flip flop is unnecessary and produce unwanted power dissipation. To prevent such type of unwanted transition, clock-gating technique is very useful. Figure 2 shown the output waveform of counter [1].

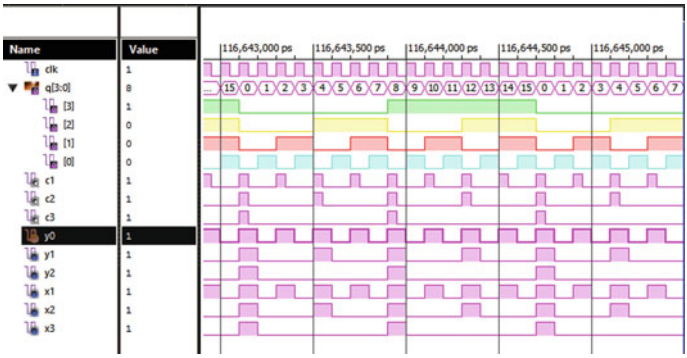


Fig. 2 Waveform of the 4-bit counter [1]

3 Proposed Synchronous Counter

Low power synchronous counter as further improved in terms of on-chip power dissipation, area (LUT), and propagation delay. The proposed 4- and 8-bit low power synchronous counter design is shown in Figs. 3 and 4, respectively. Our suggested work of the synchronous counter is quite different when compared to the one in Fig. 1,

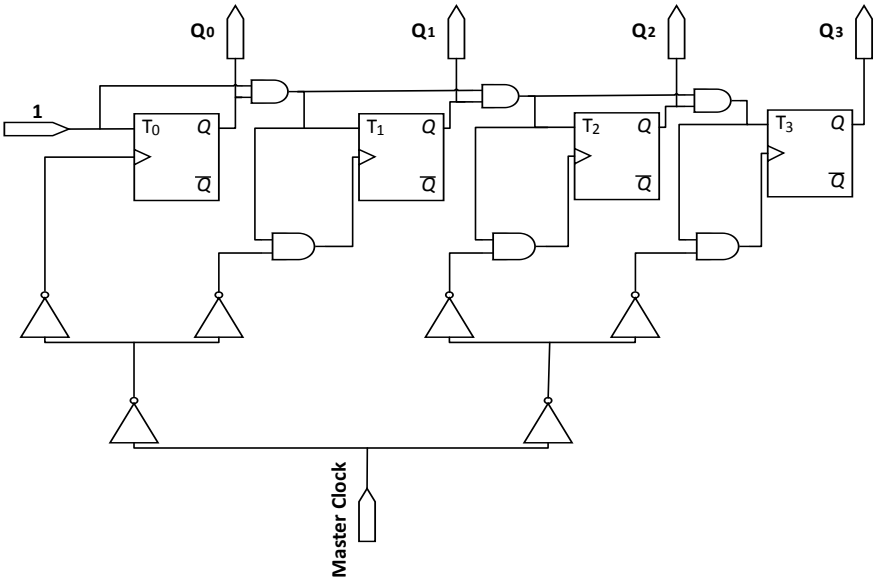


Fig. 3 Proposed 4-bit low power synchronous counter

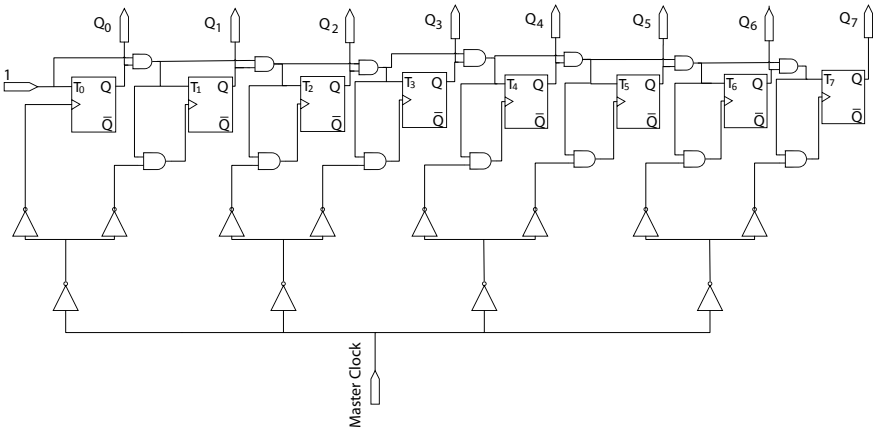


Fig. 4 Proposed 8-bit low power synchronous counter

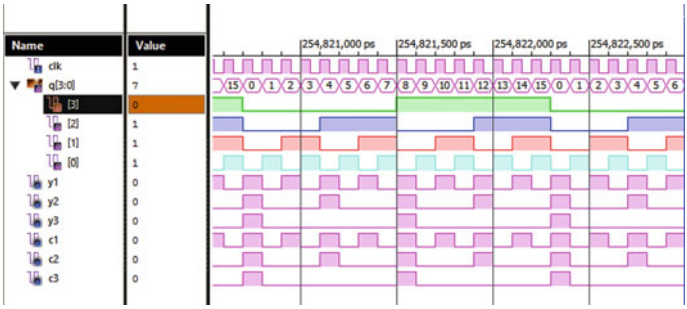


Fig. 5 Waveform of the proposed, 4-bit counter

and working on the different clock-gating techniques. For n-bit counter design n, T flip-flops are used in toggle mode and a logic AND gate using for clock gating instead of AND logic followed by NMOS transistor. The procedure is as follows: the LSB flip-flop toggles at each clock period and adjacent flip-flop toggles when the output state of the previous flip-flop is logically high and further flip-flop will toggle when the desired state of all previous flip-flop is logically high; otherwise, the flip-flop will not switch. AND logic gate worked as clock gating, when the output state of the flip-flop is logic high, then AND gate passes the clock signal to synchronize the next flip-flop; otherwise, it blocks the clock signal to prevent the unwanted switching activity within the counter. A cascaded T flip-flop design structure can be observed in this work, which has voided the key counter feature. This design strategy using T flip-flop is more concerned with tracking next-state change activity. The energy contribution due to clock also occurs during both phases T_{on} and T_{OFF} but power dissipation owing to shift T_{OFF} can be prevented by properly designing the excitation function for the clock controlled T input making T flip-flop the preferred option for low energy design [1]. Using the clock-gating method, energy consumption can be regulated. By regulating the clock signal, the gating impact minimizes the switching activity. It is evident from Figs. 3 and 4 that the suggested counter design reduces the circuit complexity that is beneficial during the fabrication. Figure 5 shows the output waveform of the proposed 4-bit synchronous counter.

4 Results and Discussion

To assess the performance and effectiveness of the recommended work 4, 8, & 16-bit counter based on advance clock-gating technique, we simulate and synthesize these outcomes on ISE design suit 14.2, Spartan 3 FPGA at 100 MHz using Verilog HDL code.

From Fig. 6, it is clearly evaluated that proposed counter is quite faster than counter [1]. From Fig. 6, it can be seen that the latency decreases for suggested

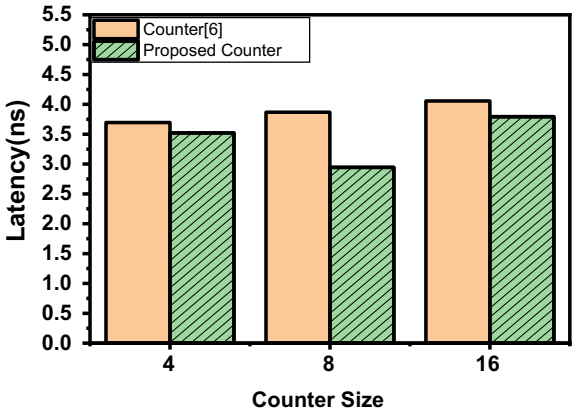


Fig. 6 Shown the relation between counter [1] and proposed counter in terms of latency

counter using the suggested technique is 4.6%, 24%, & 6.4%, respectively for 4, 8, &16-bit counter. Latency is inversely proportional to the speed.

The latency reduces using the suggested methodology will be additional considerable if the structure extended to broad bit width as well as enhanced operating bandwidth.

From Fig. 7, it is clearly seen that the suggested counter has higher speed than the counter [1]. When we compared proposed counter to the counter [1], the maximum operating frequency increases as 5%, 31%, & 7%, respectively, for 4, 8, & 16-bit counter. Proposed counter is quite advantageous to detect the fast event in the program cycle

Using the proposed methodology, maximum operating frequency rises, which is more important if the structure is extended to a wide bit width. 8-bit counter shown the more significant outcome in terms of maximum operating frequency.

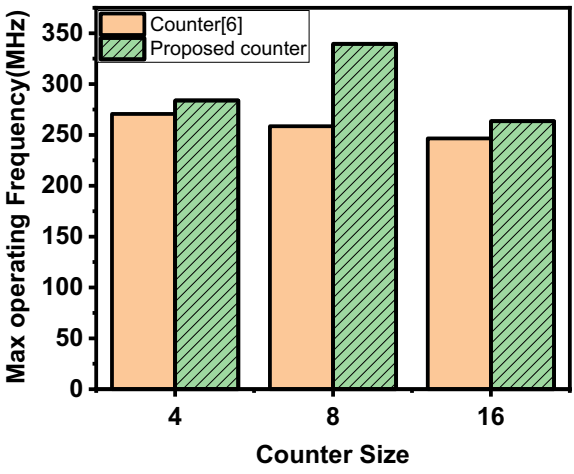


Fig. 7 Shown the relation between counter [1] and proposed counter in terms of maximum operating frequency

From Fig. 8, it is obvious that the proposed counter has less area than counter [1]. In FPGA design technology, area of the circuit, define in number of occupied slices. In proposed counter, the number of occupied slices is much lesser than counter [1]. Using the suggested methodology, no of occupied slices decreases will be more essential if the structure expanded to a large bit width and it is effective during the IC fabrication process. Comparing the suggested counter-to-counter [1], Number of occupied slices reduces by 25%, 36%, & 36% for 4, 8, & 16 bits respectively.

It is evident from Fig. 9 that the proposed counter has same on-chip power dissipation. This result shown that we preserved the on-chip power dissipation and improved the proposed counter as 36%, 24%, & 31%, respectively, for area, latency, and speed. Comparative analysis of the counter [1] and proposed counter is shown in Tables 1 and 2.

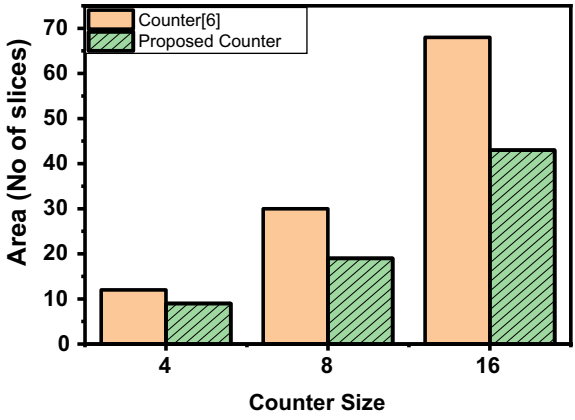


Fig. 8 Shown the relation between counter [1] and proposed counter in terms of area

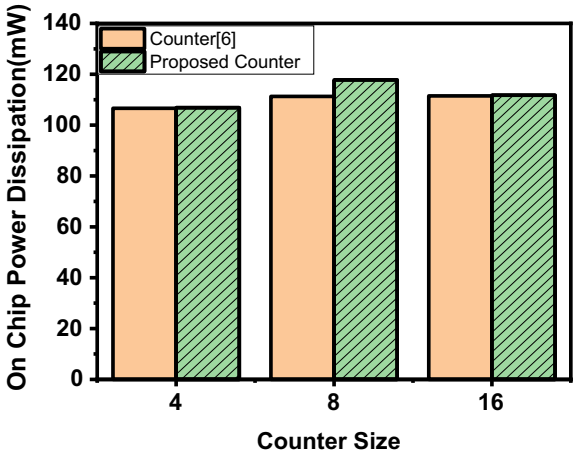


Fig. 9 Shown the relation between counter [1] and proposed counter in terms of on-chip power dissipation

Table 1 Performance comparison

Parameter	counter [1]			Proposed counter		
	4-bit	8-bit	16-Bit	4-bit	8-bit	16-bit
Max frequency (MHz)	270.614	258.542	246.567	283.938	339.628	263.581
Latency (ns)	3.695	3.868	4.056	3.522	2.944	3.794
On-chip power dissipation (mW)	106.63	111.30	111.51	106.87	117.78	111.76
Number of occupied slices	12	30	68	9	19	43

Table 2 On-chip power dissipation at different frequency

Frequency (MHz)	On-chip power dissipation (mW)					
	Counter [1]			Proposed counter		
	4-bit	8-bit	16-bit	4-bit	8-bit	16-bit
50	106.52	111.20	111.40	106.77	117.72	111.64
100	106.63	111.30	111.51	106.87	117.78	111.76
150	106.75	111.41	111.62	106.96	117.85	111.88
200	106.86	111.51	111.73	107.05	117.91	112.00
240	106.95	111.59	111.82	107.12	117.96	112.11

5 Conclusion

In this brief, we are presenting an advanced clock-gating network for power efficient and high-speed synchronous counter. The suggested design strategy takes advantage of a clock gating that offers the capacity to design power efficient and high-speed synchronous counter with minimum circuit complexity. In comparison to counter [1], we see an improvement in area and speed with minimum power dissipation. Area thereby decreases and speed involved increases in proposed counter design. Proposed synchronous counter based on AND-gate clock-gating technique minimizes the hardware complexity, maximizes the speed of the counter as well as minimizes the area. Once we compare proposed synchronous counter on different bit-width with counter [1], we find the proposed counter shown the improvement in term of latency, areas and maximum operating frequency while preserving their on-chip power dissipation. The latency is improved by 24% and increases the maximum operating frequency by 31%, which means proposed counter is faster than counter [1]. Area also decreases by 36%, which shows that massive improvement is in proposed counter because it reduces the circuit complexity, and less area is required during fabrication process. We synthesize the on-chip power dissipation at different frequencies and found that when the frequency increases then on-chip power dissipation also increases.

References

1. Katreepalli R, Haniotakis T (2019) Power efficient synchronous counter design. Comput Electr Eng 75:288–300
2. Kandasamy N, Telagam N, Devisupraja C. Design of a low-power ALU and synchronous counter using clock gating technique. https://doi.org/10.1007/978-981-10-6875-1_50
3. Hossain R, Wronski LD, Albicki A (1994) Low power design using double edge triggered flip-flops. In: IEEE transactions on very large scale integration (VLSI) systems, vol 2, no 2, pp 261–265
4. Wu X, Wei J, Pedram M (2000) Low-power design of sequential circuits using a quasi-synchronous derived clock. In: Proceedings 2000 design automation conference (IEEE Cat. No. 00CH37106). Yokohama, Japan, pp 345–50
5. Keote ML, Karule PT (2016) Modified quasi static energy recovery adiabatic logic implementation for low power application. In: 2016 2nd international conference on advances in computing, communication, and automation (ICACCA) (Fall), Bareilly, 2016, pp 1–6
6. Wu X, Pedram M (2000) Low power sequential circuit design using priority encoding and clock gating, power electronics and design, 2000. In: Proceedings of the 2000 international symposium on ISLPED'00. pp 143–148
7. Gautam M, Nirmal U, Jain R (2016) Low power sequential circuits using improved clocked adiabatic logic in 180 nm CMOS processes. 2016 international conference on research advances in integrated navigation systems (RAINS), Bangalore, 2016, pp 1–4
8. Bhargave G, Uniyal S, Sheokand P (2016) Low power adiabatic 4-bit Johnson counter based on power-gating CPAL logic. In: 2016 Second international innovative applications of computational intelligence on power, energy and controls with their impact on humanity (CIPECH), Ghaziabad, pp 297–301
9. Wang X, Robinson WH (2010) A low-power double edge-triggered flip-flop with transmission gates and clock gating. In: 2010 53rd IEEE International Midwest Symposium on Circuits and Systems, Seattle, WA, pp 205–208
10. Wu Q, Pedram M, Wu X (2000) Clock-gating and its application to low power design of sequential circuits. In: IEEE transactions on circuits and systems I: fundamental theory and applications, vol 47, no 3, pp 415–420
11. Shinde J, Salankar SS (2011) Clock gating—a power optimizing technique for VLSI circuits, 2011 Annual IEEE India Conference, Hyderabad, pp 1–4
12. Pedram M, Abdollahi A, Low power RT-level synthesis techniques: a tutorial. Department of Electrical Engineering. University of Southern California
13. Aguirre Morales JD, Marc F, Bensoussan A, Durier A (2018) Simulation and modelling of long term reliability of digital circuits implemented in FPGA. Microelectr Reliab 88–90
14. Radha S, Krishna RH, Pandi NP, Varghese S, Nagabushanam P (2018) Floor planning of 16 bit counter design for health care applications using 180 nm technology in cadence tool. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp 484–487
15. Gupta T, Verma G, Kaur A, Pandey B, Singh A, Kaur T (2015) Energy efficient counter design using voltage scaling on FPGA. 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, pp 816–819
16. Rajashekar P, Malathi M (2011) Power efficient multiplexer using DLDFF synchronous counter. IEEE J Solid State Circ 42:284–289
17. Nehru K, Shanmugam A, Darmila Thenmozhi G (2012) Design of low power ALU using 8T FA and PTL based MUX circuits. In: IEEE conference on advances in engineering, science and management, pp 45–49
18. Rivoallon F (2001) Reducing "switching power with intelligent
19. Thakur A, Mehra R (2016) Power and speed efficient ripple counter design using 45 nm technology. In: 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES), Delhi, pp 1–4

20. Agrawal T, Kumar A, Sharma J (2018) Environment friendly frequency scaling based counter design. In: 2018 9th international conference on computing, communication and networking technologies (ICCCNT), Bangalore, pp 1–4

Estimating Energy Consumption for Various Sensor Node Distributions in Wireless Sensor Networks

Pallavi Joshi, Shashank Gavel, and Ajay Singh Raghuvanshi

Abstract A wireless sensor network uses sensor nodes with sensing, manipulating and communication abilities. The energy efficiency is one of the major challenges for WSN as it survives on batteries. Most of the energy is consumed by communication and data processing. Data aggregation is the best way to address such challenges. The in-network data aggregation mainly focuses on these problems which are energy constraint in the sensor networks. The main task in the data aggregation algorithms is to gather data and aggregate it in an energy-efficient manner so as to increase the network lifetime. In this paper, we have studied the random deployment of sensor nodes using eight different random distributions with and without clustering and their impact on the K-means and K-medoids clustering algorithms. Simulation results show that, for a dense WSN scenario, the K-medoids clustering algorithm gives better results for two sensor nodes distributions namely: Beta and Uniform distributions. Also, we carry out a brief survey on different data aggregation algorithms and their comparison on the basis of network lifetime, communication delay, data accuracy and energy efficiency. In the end, we conclude our work with possible future scope.

Keywords Wireless sensor network · In-network data aggregation · Sensor node distributions · K-means clustering · K-medoids clustering

P. Joshi (✉) · S. Gavel · A. S. Raghuvanshi
Department of Electronics and Telecommunications Engineering, National Institute of Technology, Raipur, Raipur 492010, India
e-mail: pjoshi.phd2017.etc@nitrr.ac.in

S. Gavel
e-mail: sgavel.phd2016.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_28

1 Introduction

A sensor node is composed of typical sensors which senses environmental parameters like stress, pressure, temperature, light, gas and many more. These sensed qualities are in the form of electrical signals which have to be further calibrated to calculate the values of corresponding physical properties. A wireless sensor network is formed by collection of such sensor nodes [1]. A typical wireless sensor network is depicted in Fig. 1 [2].

The nodes are deployed basically in two structures—random and deterministic. A wireless sensor network with both node deployments is depicted in Fig. 2.

The nodes in the wireless sensor network exhibit very limited energy and power sources. Mostly, the nodes are battery operated and these batteries are non-replenishable. Due to their operation in harsh environmental conditions they have limited computational as well as communication capabilities. Thus, sensor nodes are more prone to attacks and failures. The applications such as habitat monitoring, military surveillance, forest fire detection do not require human intervention. So, the main challenges in the path of WSN is to conserve the energy of nodes and improving the network lifetime [3].

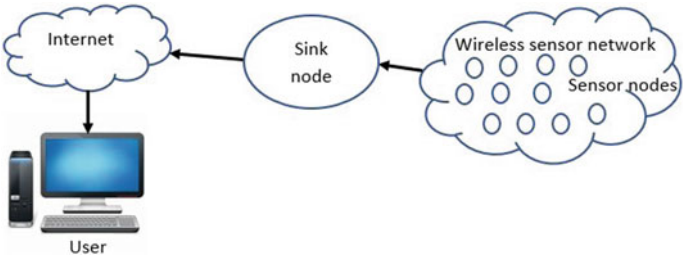


Fig. 1 Typical wireless sensor network

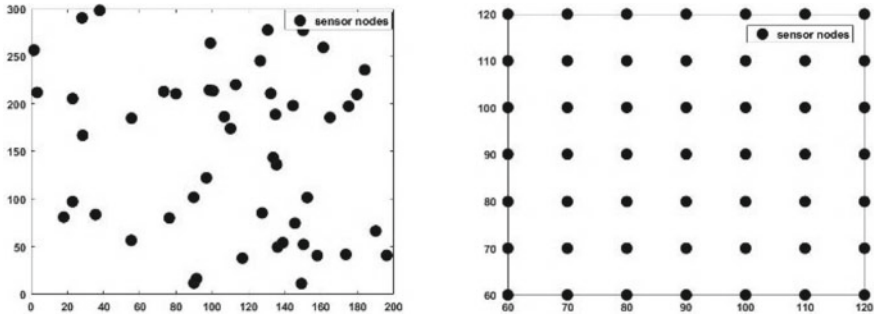


Fig. 2 **a** Random node deployment, **b** deterministic node deployment

To overcome the challenges, there is a need to reduce the data transmission delay between the nodes. Aggregation of data would be the intelligent technique for such problems in WSN. The notion behind data aggregation is to accumulate data at intermediate nodes from various source nodes in the network, thus, reducing packets per transmission [4]. Data aggregation (DA) is used frequently to address these issues which makes it the most acquainted method in many wireless network scenarios. It is the means of combining and compressing the data, thereby reducing the redundancies and anomalies in the data. There have been many works which are already carried out in this field. The efficient utilization of node energy is to improve sensing accuracy and data gathering capability and is presently one of the important research aspects wireless sensor network. For dense WSNs, the data transfer needs more hops to reach to sink node, in this case, the nodes in the proximity of gateway nodes are more likely to drain their energy and become dead. Thus, it is quite a difficult task to ensure minimum energy consumption with maximum node lifetime [5].

2 Background

Data aggregation is basically the collection of most critical data from sensor nodes to some intermediate nodes and transmission of the accumulated data to the sink node. This task is done energy efficiently and with least latency. Both energy and latency are important constraints in the applications involving data aggregation such as habitat monitoring. It is a difficult task to build such an efficient data aggregation model which can enhance the lifetime of sensor nodes in a wireless network. Some of the factors which can decide the efficiency in energy in the network are the employed routing protocol, architecture of the network, data aggregation approach, etc. We focus on some data aggregation approaches to determine the energy efficiency of a network. Some challenges in data aggregation are:

- Data accuracy
- Energy efficiency
- Network lifetime
- Latency.

Table 1 gives the overview of different of data aggregation approaches with examples of some implemented protocols [6].

In the first approach, the protocol used for the transmission of data from start node to sink depends on the network architecture and topology. The second approach deals with the network flow-based aggregation and the third one considers quality of service parameters of the network for data aggregation. Since we are dealing with the structure-based data aggregation, we focus on in-network data aggregation approaches.

Table 1 Types of data aggregation schemes based on different parameters in a WSN

S. No.	Data aggregation on different parameters in sensor networks	Types of data aggregation protocols
1	Data aggregation protocol based on network architecture	1. DA for flat network-pull diffusion, SPIN 2. DA for hierarchical network (a) Tree-based EADAT (b) Cluster-based LEACH (c) Chain-based PEGASIS
2	Network flow-based data aggregation protocol	1. DA for lifetime maximization-MLDA 2. Network correlated data gathering-SPT
3	QOS aware data aggregation protocols	1. DA for optimal information extraction 2. DA for congestion control and reliability

2.1 In-network Data Aggregation

When data aggregation is done along the routes, the aggregated data packets exhibit data from various source nodes and if one of the packets is ruined that information is also lost. In the context of sensor networks, in-network data aggregation protocols should have desirable characteristics like: minimum number of data for setting a route tree, reliable data transmission, maximum aggregation rate and maximum overlapping routes.

The in-network data aggregation deals with the different ways in which the nodes between source and destination forward the data to the sink node and gather data from various source nodes. A data aggregation aware routing algorithm is the key component of in-network DA [7]. The in-network DA is different from classic approach of DA aware routing which follows shortest path routing to forward data to the sink node. The notion behind in-network DA is the synchronized data transmission among nodes. In in-network data aggregation algorithms, the data from a node is kept in waiting until neighboring nodes are also available with the data, this leads better aggregation opportunity. It improves its performance and saves the energy. Some in-network aggregation approaches are:

2.2 Tree-Based Approaches

This approach works on hierarchical organization of sensor nodes. Some special aggregation points are elected which define a direction from source to sink to forward data. A tree structure is organized in the network, then the sink node transmits the queries and all the data is gathered at the same node of that tree when routing occurs [8]. Thus, the aggregated packets are forwarded to neighbor nodes and so on. Some of the tree-based data aggregation approaches are the shortest path tree (SPT) algorithm [9], greedy incremental tree approach (GIT) [10], center at nearest (CNS) algorithm [9], tiny aggregation service (TAG) [11, 12].

This approach has some drawbacks:

1. It does not respond well to fault tolerance. When there is a loss of a packet at a level in the tree, the whole subtree becomes waste and the data cannot be forwarded through that part.
2. It is more prone to link failure and node failure.
3. It is not a reliable approach in terms of cost.

2.3 Cluster-Based Approaches

These approaches are also very similar to the tree-based approaches consisting of network organized hierarchically. In these approaches, clusters are formed from nodes and each cluster has a collector node called cluster head. The cluster leaders aggregate the data from each cluster, thereby forwarding the aggregated data to the sink node. The best example for such approach is low-energy adaptive clustering hierarchy (LEACH) [13, 14]. Another cluster-based approach is the information fusion-based role assignment (InFRA) algorithm [15], only event detecting nodes form clusters. Then, the cluster heads integrate the data and forward it to the destination node. In InFRA algorithm as clusters are build, the cluster heads select shortest path to maximize information fusion. One demerit of InFRA algorithm is whenever a new event occurs that information must be disseminated in whole network so that other nodes can be informed of its occurrence and also the aggregated coordinators distance is updated. It makes the communication cost high and limits the scalability [16].

2.4 Chain-Based Approaches

In this approach, each node sends data to the closest neighbor. All nodes are grouped in a linear chain for data aggregation. Power Efficient Data Gathering Protocol for Sensor Information system (PEGASIS) is the best example which employs this approach [17]. It uses greedy algorithm to form a chain of nodes since it assumes that the sensor nodes have global information about the network. The node which is very far from sink node initiates the formation of chain and with each step a close neighbor of it is chosen as the successor node, in this way, the chain formation proceeds.

2.5 Structureless Approaches

Very few routing aware algorithms use structureless approach, one of them is data-aware anycast algorithm (DAA) [8]. It forwards data to the sink via one-hop neighbors. It uses both spatial and temporal aggregation which means the aggregated

packets have the tendency to meet at same node and at same time. But this approach has also some limitations that is this approach is not cost effective for larger networks because it cannot aggregate all packets as the size of network increases.

3 Proposed Work

The paper proposes the impact of applying different sensor node distributions on the clustering and energy consumption of nodes. A wireless sensor network scenario with field dimensions 500 × 500 is considered with 100 nodes. The sink node is at fixed position whose coordinates are (400,100). These nodes are deployed randomly using eight different distributions namely: Normal, Weibull, Gamma, Exponential, Beta, Poisons, Cauchy and Uniform [18].

The two clustering algorithms are implemented namely k-means and k-medoids on all eight sensor distributions for a WSN scenario. In this experiment, the nodes are grouped into four clusters. The number of clusters can be changed. We are observing the impact of these clustering algorithms on different node distribution scenarios. The average energy consumed by a node per transmission and the number of operational nodes per transmission is calculated for all node distributions. A comparison is done between k-means and k-medoids on the basis of the performance metrices is done.

4 Simulation Results

The following figures show the node deployments and simulations performed in the experiment. Simulations results are obtained in MATLAB software.

Figure 3 shows the random node deployment with 100 number of nodes and a sink node at the location (400, 100). The nodes are deployed using eight different random distributions. Figure 4 shows the clustering of the same nodes by k-means algorithm. Figure 5 shows clustering by k-medoids algorithm. Table 2 shows values of average energy consumed by a node per transmission calculated for each distribution which are used to deploy 100 nodes with K-means clustering. Table 3 shows the values of average energy consumed by a node per transmission calculated for each distribution which are used to deploy 100 nodes with K-medoids clustering. Figure 6 shows the decreasing number of operational nodes per transmission for 8 node distributions employing K-means clustering. Figure 7 shows the number of operational nodes calculated per packet transmission for 8 node distributions employing K-medoids clustering.

It is observed that for k-means clustering, the distributions namely: Normal, Exponential and Cauchy consume minimum average energy by a node per transmission. For k-medoids clustered network the distributions namely: Beta, Poisson, Cauchy consume minimum average energy by a node per transmission. Also, the number of operational nodes decreases for more transmissions in the K-medoids clustering,

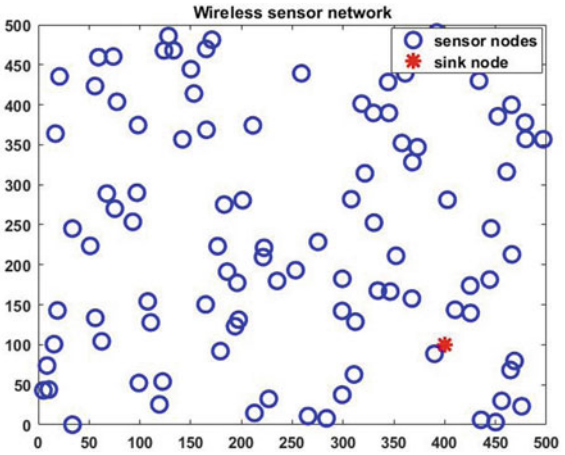


Fig. 3 Wireless sensor network scenario with 100 nodes and a sink node

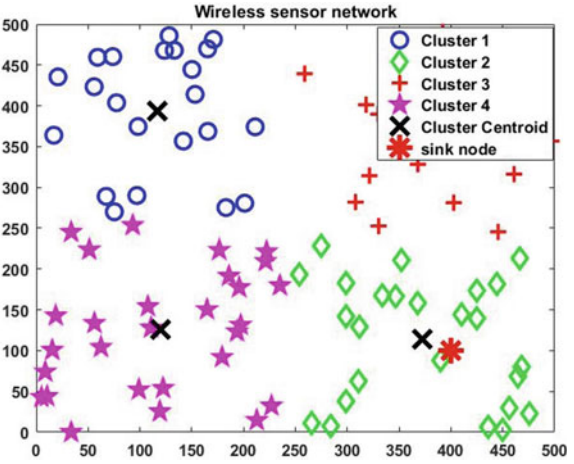


Fig. 4 Nodes are grouped into 4 clusters by applying the K-means clustering algorithm

whereas for K-means clustering the number of operational nodes decreases rapidly for less transmissions that is the energy of nodes deceases rapidly. For K-means clustered network, the performance of the operational nodes is not appreciable for Normal, Weibull, Exponential, Beta, Poisson and Cauchy distributions.

5 Discussions

Verma et al. [18] proposed the event-based routing protocol which uses 8 sensor node distributions and their impact on sense count receive count and receive redundant count. They took into account the flooding theory which has disadvantages like large

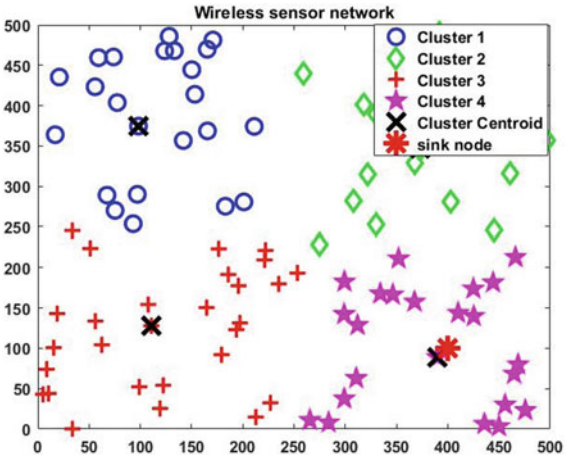


Fig. 5 Nodes are grouped into 4 clusters by applying the K-medoids clustering algorithm

Table 2 Average energy consumed by a node per transmission in the network with k-means clustering for 8 sensor node distributions

S. No.	Name of the distribution	Average energy consumed by a node per transmission
1	Normal	0.0004605
2	Weibull	0.01879
3	Gamma	0.03746
4	Exponential	0.000575
5	Beta	0.03834
6	Poisson	0.03949
7	Cauchy	0.001268
8	Uniform	0.03477

Table 3 Average energy consumed by a node per transmission in the network with k-medoids clustering for 8 sensor node distributions

S. No.	Name of the distribution	Average energy consumed by a node per transmission
1	Normal	0.03092
2	Weibull	0.03149
3	Gamma	0.03354
4	Exponential	0.02762
5	Beta	0.001956
6	Poisson	0.000175
7	Cauchy	0.000125
8	Uniform	0.01521

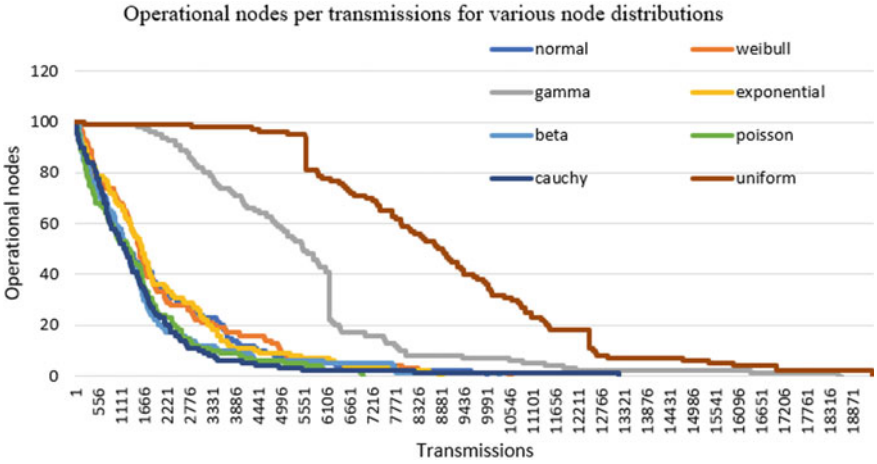


Fig. 6 Number of operational nodes per transmission for 8 node distributions in the K-means clustered network

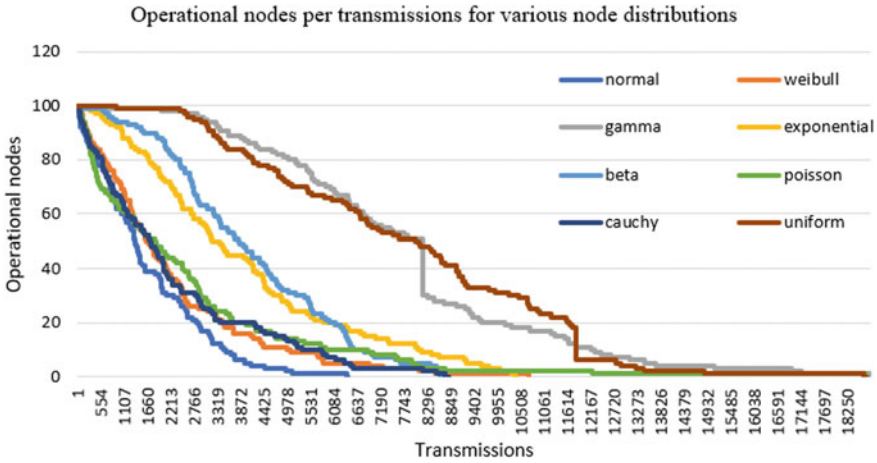


Fig. 7 Number of operational nodes per transmission for 8 node distributions in the K-medoids clustered network

bandwidth and energy requirement. In this paper, clustering is introduced which minimizes the energy consumption in a sensor network. Yu et al. [19] considered cluster-based routing protocol to compare network lifetime for non-uniform and random sensor node distributions. The network lifetime for non-uniform distribution is better but the problem of energy imbalance throughout the network arises, which has to be addressed. Table 4 shows that the average energy consumed in the network is less when clustering is employed than the values of energy obtained when the clustering is not used [20].

Table 4 Comparison of average energy consumption in the network with clustering and without clustering [21]

Node distribution	With clustering	Without clustering
Normal (k-medoids)	0.03092	0.068382
Uniform (k-medoids)	0.01521	0.081355
Normal (k-means)	0.0004605	0.070870
Uniform (k-means)	0.03477	0.195974

6 Conclusion

The main contribution of this paper is the implementation of eight types of random sensor distributions and their analysis for different performance metrices as well as their impact on clustering. By observing the results, it can be concluded that for a large sensor network k-medoid clustering performs well over k-means for different sensor node distributions. The results show that the network with Uniform and Beta sensor distributions employing K-medoids clustering gives better results as compared to other distributions and ensures high network lifetime. Also, a smaller number of transmissions are required in case of k-medoid clustering. This work may be extended by introducing the mobility in intermediate nodes and also scheduling in nodes to save the energy dissipation in the entire network [21, 22]. Further, the clustering can be made optimal to make the network more energy efficient.

References

1. Li C, Zhang H, Hao B, Li J (2011) A Survey on routing protocols for large-scale wireless sensor networks. Sensors 11(4):3498–3526
2. Rajaram ML, Kougianos E, Mohanty SP, Choppali U (2016) Wireless sensor network simulation frameworks: a tutorial review: MATLAB/Simulink bests the rest. IEEE Consum Electron Mag 5(2):63–69
3. Wu X, Chen G, Society IC, Das SK (2008) Avoiding energy holes in wireless sensor networks with nonuniform node distribution. 19(5): 710–720
4. Rajagopalan R, Varshney PK (2006) Data aggregation techniques in sensor networks: a survey recommended citation data aggregation techniques in sensor networks: a survey. Sensors 11: 3498–3526
5. Hua C, Yum TSP (2008) Optimal routing and data aggregation for maximizing lifetime of wireless sensor networks. IEEE/ACM Trans Netw 16(4):892–903
6. Fasolo E, Rossi M, Widmer J, Zorzi M (2007) In-network aggregation techniques for wireless sensor networks: a survey. IEEE Wirel Commun 14(2):70–87
7. Dhasian HR, Balasubramanian P (2013) Survey of data aggregation techniques using soft computing in wireless sensor networks. IET Inf Secur 7(4):336–342
8. Fan KW, Liu S, Sinha P (2006) On the potential of structure-free data aggregation in sensor networks. In: Proceedings of IEEE INFOCOM, 2006
9. L. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," *Proc. - Int. Conf. Distrib. Comput. Syst.*, vol. 2002-Janua, pp. 575–578, 2002.
10. Intanagonwiwat C, Estrin D, Govindan R, Heidemann J (2002) Impact of network density on data aggregation in wireless sensor networks. Proc Int Conf Distrib Comput Syst 457–458

11. Madden S, Franklin MJ, Hellerstein JM, Hong W (2002) TAG: a tiny aggregation service for ad-hoc sensor networks∗. Oper Syst Rev 36(Special Issue):131–146
12. Madden S, Szewczyk R, Franklin MJ, Culler D (2002) Supporting aggregate queries over ad-hoc wireless sensor networks. In: Proceedings of 4th IEEE Work Mobile Computing System and Applications WMCSA, pp 49–58
13. Goyal R (2014) A review on energy efficient clustering routing protocol in wireless sensor network. 3(5):2319–2322
14. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wirel Commun 1(4):660–670
15. Nakamura EF, Loureiro AAF, Frery AC (2007) Information fusion for wireless sensor networks. ACM Comput Surv 39(3):9-es
16. Younis O, Krunz M, Ramasubramanian S (2006) Node clustering in wireless sensor networks: recent developments and deployment challenges. IEEE Netw. 20(June):20–25
17. Dasgupta K, Kalpakis K, Namjoshi P (2003) An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. IEEE Wirel Commun Netw Conf WCNC 3(C):1948–1953
18. Verma VK, Singh S, Pathak NP (2014) Comprehensive event based estimation of sensor node distribution strategies using classical flooding routing protocol in wireless sensor networks. Wirel Netw 20(8):2349–2357
19. Yu J, Qi Y, Wang G, Gu X (2012) A cluster-based routing protocol for wireless sensor networks with nonuniform node distribution. AEUE Int J Electron Commun 66(1):54–61
20. Sangwan R, Duhan M, Dahiya S (2013) Energy consumption analysis of ad hoc routing protocols for different energy models in MANET. 6(4): 48–55
21. Huang H, Savkin AV (2017) An energy efficient approach for data collection in wireless sensor networks using public transportation vehicles. AEUE—Int J Electron Commun 75:108–118
22. Bagaa M, Younis M, Ksentini A, Badache N (2014) Reliable multi-channel scheduling for timely dissemination of aggregated data in wireless sensor networks. J Netw Comput Appl 46:293–304

A Statistical Method to Predict the Protein Secondary Structure

Smita Chopde and Jasperine James

Abstract The statistical method has simplified the processes for prediction of protein structure in economical way and reduced the complex methods reacquired for finding the structure. In statistical method, various machine learning methods are used. For finding the protein structure, this paper has used neural network for this purpose. This paper attempts to illustrate this method by using the feed forward neural network and deep learning toolbox for the stated purpose. The protein structure has made it possible to define the 3D structure of the protein. This has simplified the efforts of the biochemist to develop suitable drugs for various diseases with suitable power of the drug as well as the medical practitioner to work on the chronic diseases like cancer and other genetic diseases.

Keywords Neural network · Secondary structure · Classifier · Confusion matrix

1 Introduction

Protein structure forecast is one of the significant territories of research for the specialists in the region of bioinformatics. It includes expectation and examination of large-scale particles. Proteins are the essential atoms of all the living life form. There are special chains of amino acids. They speak to the one of a kind three-dimensional structures which permit the analyst to do perplexing natural capacity. The determination of protein succession is given by the amino acids. There are 20 diverse amino acids which are framed via carbon, hydrogen, nitrogen, oxygen and sulfur. Proteins have four unique structures referenced beneath [1–3]:

S. Chopde (✉) · J. James
Department of Electronics and Telecommunication Engineering, Fr. Conceicao Rodrigues Institute of Technology, Vashi, Navi Mumbai, India
e-mail: smita.chopde@fcrit.ac.in

J. James
e-mail: jasperinesimon6295@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_29

1. Primary: In form of amino acids in prime structure.
2. Secondary: Nearby well-defined extremely regular sub-structure, alpha helix, beta pieces and coil are the secondary protein structures.
3. Tertiary: Three-dimensional structure.
4. Quaternary: Complex of several polypeptide chains.

The general way to deal with anticipate the optional structure of protein is finished by contrasting the amino corrosive succession of a specific protein to arrangement of the well-known database. In protein, optional structure expectation amino acids successions are inputs and coming about yields is adaptation or the anticipated structure which is the blend of alpha helix, beta sheets and circles. Natural concept: Homology is the nuts and bolts of bioinformatics. In gnomic, bioinformatics is the capacity of quality anticipated by homology method.

1.1 Homology Technique

It keeps the rule that if the grouping of gene A, whose capacity is known, is homologous to the arrangement of quality B may share An's information. In the auxiliary, bioinformatics homology is utilized to figure out which some portion of protein is significant in structure development and communication with other protein. In homology demonstrating method, this data is utilized to anticipate the structure of protein once the homologous structure protein is known. At present, the best way to foresee the structure is to anticipate it dependably In protein structure, it is critical to discover the adaptation class. A ordinary protein grouping.

Protein sequence.

AAABABABCGPCQPPPAAAQQAQQA

1.2 Conformation Class

HHHH EEEE HHHHHHHH

where H means helical, E means extended and banks are remaining coiled conformation. The accuracy is given by

$$Q = [(P_\alpha + P_{\beta +} P_{\text{coil}})/T] \times 100\%$$

where P_α, P_β P_{coil} are the number of residues predicted correctly in state alpha helix and beta strand and loop and T is the total number of residues.

1.3 Need of Protein Secondary Structure Prediction

The micro-molecules can be predicted and analyzed. Structure function relation can be provided which helps to find particular function of particular protein structure. On protein structure synthesis, the function can be added or removed. The structure prediction will give the information of unknown protein which can be viral which help the biochemist to design the new drug. This reduces the sequence structure gap. The protein structure of membrane protein can as well be determined by the computational tool which would be slightly incapable by the experimental techniques.

2 Method

2.1 Dataset

The Rost-Sander data set [4] is used in this paper which contains a protein structure which is relatively used large domain and has greater density and length. The file RostSanderDataset.mat is the set available in the MATLAB 16 version [5, 6] can conveniently and usually used for the analysis of the protein structure as the structural assignment is known for every protein sequence.

2.2 Neural Network Architecture

Neural system is a solid methodology for learning the Structural country—helix, sheet or circle of every buildup in a given protein, created absolutely at the protein basic styles found all through a tutoring stage. To guarantee profitability of the outcomes, the worldwide irregular generator is reset to a spared state available in the document stacked.

The neural system contains of one information layer, one masked layer and one yield layer. The neural system has just one covered up layer, therefore, it is referred as simple artificial neural network (ANN) which nearly works like the human brain. Similar to human neural system which has the veins for transition of sign, denrites as the message maker neurons as the places for transmission of the message or sign produced at the denrites of the hub.

The information layer is practically like the nodes present in the human neural framework utilized for social event the data of the amino corrosive grouping. The data if further utilized for encoding a sliding window of each information amino corrosive assortment and also probabilities are made on the state of the buildup situated at the focal socket of the window. The length of the Window 17 is estimated related on the realities of the factual connection to be had between the optional structure of a given buildup work and the 8 deposits on the similarly aspect of the forecast socket [7, 8].

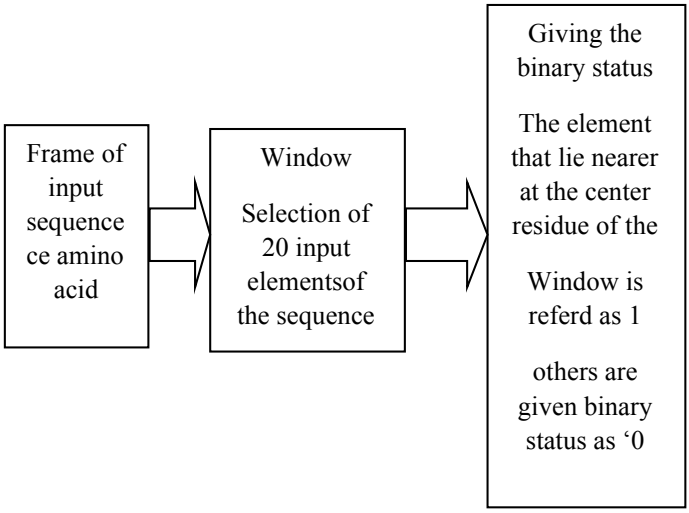


Fig. 1 Summarizing the event of grouping the amnio acid sequence and giving the binary status

Every window region is encrypted the use of a double exhibit of dimension 20, here, to make reference to that the cluster size 20 is considered as it was watched this gave the precise prerequisite, with every window having one component for each amino corrosive kind. Each gathering of 20 data sources, the component speaking to the amino corrosive kind inside the prominent job is going to one, at the indistinguishable time, as unwinding others responses are set to zero. Consequently, the info layer comprises of $R = 17 \times 20$ input gadgets, I.E. 17 gatherings of 20 data sources. This event of grouping the input can be summarized in the following flow graph as shown in Fig. 1

First, create possible protein sequence similar to a sliding window of size W with the aid of create a Henkel matrix. In this matrix formation, the residue within the *i*th column represents the sub-series from the *i*th role inside the unique collection. Then, for every location of the residue in the window, an array of dimension 20 is created and fixed *j*th residue in the sequence is fixed to 1 if the residue lies within the nearer vicinity of the residue considered inside the window has a numeric representation equal to *j*.

The output layer of neural network contains of three units, one unit for every of the structural kingdom or classes taken in to account which might be encoded the usage of a binary collection. To generate a goal matrix for the neural community, we initial attain from the information all of the structural possibilities of all feasible subsequence parallel to the sliding window, then take into account the important role in individually window and remodel the parallel structural opportunities the use of the binary encoding: 1 zero for coil, 0 1 zero for sheet, 0 zero 1 for helix. Before forming the neural network, first carry out normalization of the input sequences and

goal matrix factors representing the region of the residues from the center residues of the window taken into consideration.

2.3 Creating Neural Network

The secondary structure expectation can be perception of as a pattern appreciation problem which broadly used for identification of the textual content or digit using the neural network. In gift situation, the neural network is trained to recognize the structural state in '1' or '0' of the primary residue maximum probably to occur while the unique residues in the given sliding window are discovered. The pattern recognition for neural community is created the use of the center and target matrices well-defined above and identifying a hidden layer of dimension 3.

2.4 Training the Neural Network

The pattern recognition network utilizes conjugate inclination calculation for training. In deep picking up information on gadget field documentation, various calculations are to be had for a rundown of characteristic. In patter notoriety network after the each tutoring cycle, the tutoring grouping is offered to the system over the sliding window, one buildups at a time. The concealed element change the sign got from the enter layer by means of the switch highlight logsig to flexibly a yield sign that is between and near both 0 or 1 reproducing the terminating of neuron [1, 7]. Loads are balanced all together that the mistake among the discovered yield from each unit and the favored yield sure through the objective network is limited. During instruction, the preparation apparatus window opens and shows the turn of events. The preparation data comprising of the arrangement of rules, the presentation norms, the type of blunder considered, etc., are proven. The feed forward network is utilized for the example notoriety of the incoming buildup with the buildup put at the center of the sliding window [9].

One of the regular problem associated with feed forward network is at some point of neural network education is information outfitting, wherein the community tends to memorize the statistics/samples giving 100% accuracy which is not the desired end result on account that no gadget is one hundred% accurate and is blind in gaining knowledge of a way to oversimplify to the new scenario. The default method for enhancing the generalization is referred to as early preventing while assuming that accuracy reaches near to 95–98% and is composed in divided the available education statistics set into three substates: (1) The training set which is used for calculating the gradient and informing the network masses and biases. (2) The validation set, whose faults are examined at some step in the training techniques as it inclines to increase while data is over equipped three. The test facts whose faults may be used to assess the first-rate of the department of facts set.

When the usage of the function educate, by way of default the information is casually separated so that 60% [1, 10, 11] of the trials are allotted to the training set, 20% [10, 12, 2] of the validation set and 20% to the check set, the opposite sorts of segregating can be implemented via requiring the property Internet. Divide fnc.

The function plot execute shows the movements of the training, validation and test mistakes as training group pass as shown in parent Fig. 2a–c respectively.

The training breaks when one of numerous conditions is met. Consider an illustration in training, the training technique breaks when the validation faults growths for a precise wide variety of iteration say 6 or the most wide variety of acceptable generation is touched a thousand (Fig. 3).

2.5 Analyzing the network response

To training the network response, we see the confusion matrix by seeing the productions of the trained network and connecting them to the predictable results (targets) (Fig. 4).

The diagonal factors in the network show the quantity of buildup place that are linear the inside found buildup inside the window for separately basic elegance. The off-inclining cells sign the arrangement of deposits that are misclassified, e.g., Helical spots expected as snaked areas. The slanting perspectives consent to the statement that are proficiently estimated which include the quantity of observation and the rate of total no. of observation in each cell. The segment on the far appropriate of the grid shows the portion of all models foreseen to have a place with every polish which are proficiently and inaccurately marked. These measurements are much of the time known as the accuracy or awesome plausible qualities and the bogus revelation charge, respectively. The column at the most minimal of the network shows the level of the entirety of the models having a place with every style which may be accurately and erroneously characterized. These measurements are frequently alluded to as the remember and bogus poor cost separately. The line inside the bottom of the lattice shows the general precision.

We remember the receiver operating characteristics (ROC) bend, a plot of the real worthwhile charge (affectability) versus the bogus positive rate (1-particularity) (Fig. 5).

Refining the neural network for accurate results:

The neural network this is characterized is entirely straightforward. To acquire improvements in the forecast exactness, we need to endeavor one of the accompanying:

Increment the wide assortment of preparing vectors. Expanding the scope of successions committed to preparing reacquires a huge handled database of protein structure with an estimated dispersion of wound, helical and sheet components.

Increment the amount of data sources esteems. Expanding the window size or including extra pertinent realities which incorporate biochemical homes of the amino acids are the substantial choices.

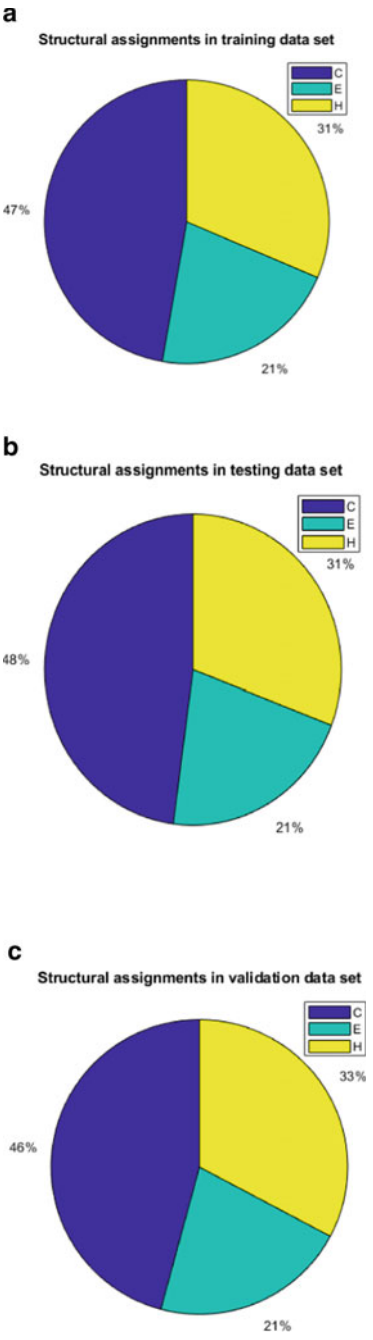


Fig. 2 **a** Structural assignment in training data set, **b** structural data set in testing data set, **c** structural data set in testing data set

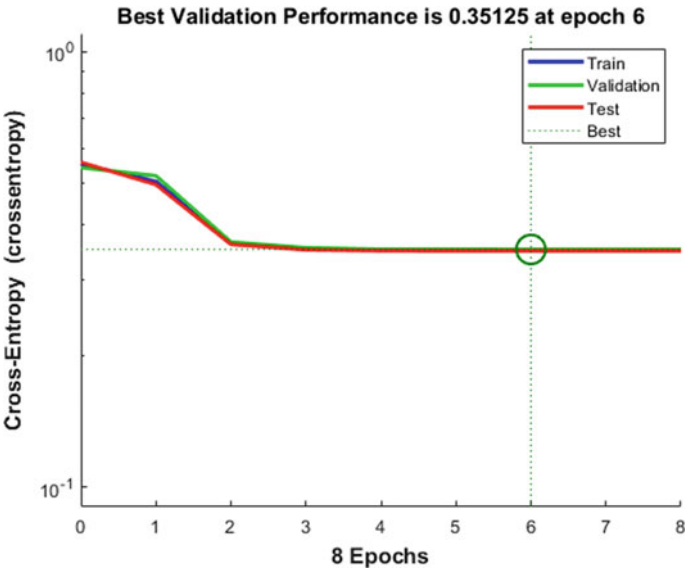


Fig. 3 Validation performance for several iterations

Utilize an unmistakable tutoring set of rules. Different calculations shift in memory and speed necessity.

Expanding the quantity of covered up neurons. By including progressively concealed devices we by and large accomplish an additional cutting edge connect with the capacity for better anyway we should careful not to overfit the data.

2.6 Accessing Network Performance

The assessment of structure expectation in component might be determined by utilizing figuring forecast lovely lists which suggest how pleasantly a chose country is envisioned and whether the overprediction or underprediction has took place. We characterize the file pcObs(S) for nation S ($S = $ C, E, H) on the grounds that the quantity of buildups viably anticipated in the realm S, isolated by utilizing the amount of deposits situated inside the country S. Similarly, we layout the record pcPred(S) for the realm as the scope of deposits accurately anticipated in the realm S, partitioned by the amount of deposits expected in country S (Fig. 6).

These quality records are helpful for the understanding of the forecast precision. Truth be told in situations where the expectation systems tend to overpredict/underpredict a given express, a high/low forecast exactness may be a curio and does not give a proportion of value to the method itself.

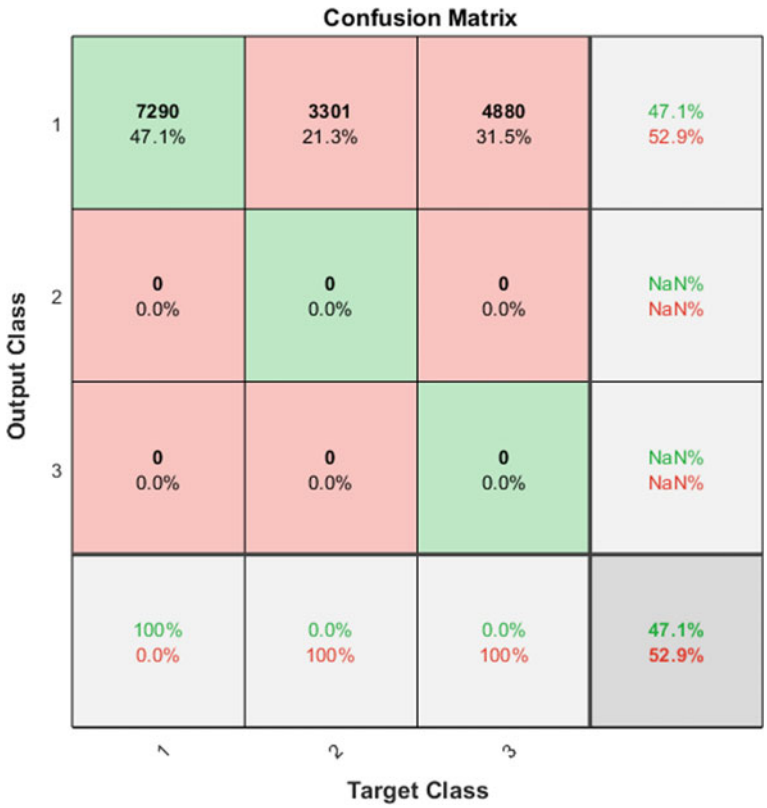


Fig. 4 Confusion matrix showing number of residues correctly classified

3 Conclusion

The technique existing here predicts the position of the structural state of a assumed protein residue based on the structural state of the residue placed at the center of the sliding window selected for finding the nearby residue that lie almost closely to the center located residue in the sliding window. However, there were restrictions observed when predicting the contents of the amino acid molecules present in the protein, such as the smallest length of individually amino acid residue. This situation was observed when a helix and sheet is allotted to any group of four or more continuous residues. To include this type of information, multi-neural network can be designed to predict the amino acid sequence followed by finding the structure element, i.e., primary, secondary or tertiary structure.

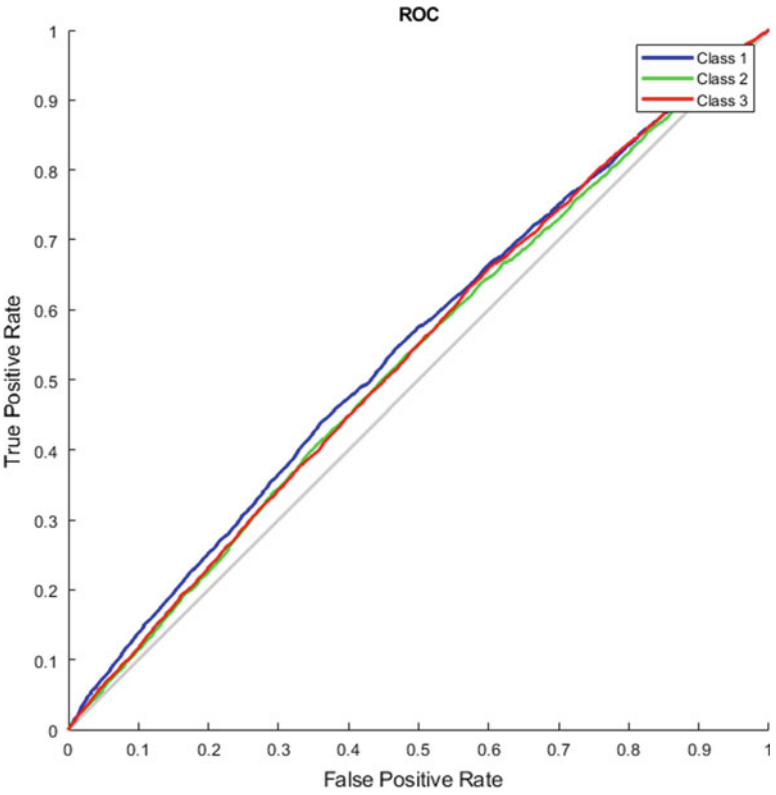


Fig. 5 Receiver operating characteristics

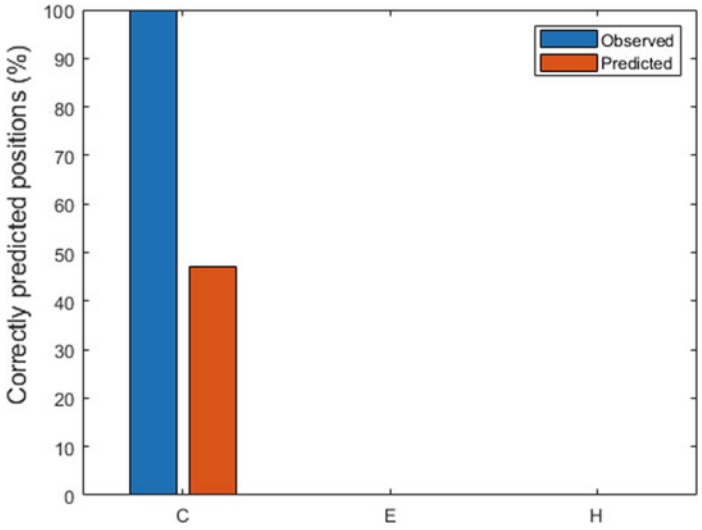


Fig. 6 Comparison of quality indices of prediction

References

1. Micheal G, Protein bio informatics: from sequence to function, Kindle edn.
2. Leaning D (2018) From basics to practice volume 1 and 2, Andrew Glassner. The Imaginary Institute, Seattle
3. Agostino M, Practical bio informatics
4. Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 232(2):584–599
5. Goodfellow I, Bengio Y, Courville A (2017) Deep learning. MIT Press, Cambridge
6. www.mathwork.com
7. Holley LH, Karplus M (1989) Protein secondary structure prediction with a neural network. PNAS 86(1):152–156
8. Murphy KP, Machine learning a probabilistic perceptive. The MIT Press, Cambridge
9. Kheller JD, Namee BM, Dárcy A (2018) Fundamentals of machine learning foe predictive data analysis. MIT Press, Cambridge
10. Kabsch Wand Sander C (1983) How good are prediction of protein secondary structure? FEBS Lett 155(2):179–182
11. Barnes MR, Bio Informatics for genetics, 2nd edn.
12. Haykin S. Neural network and machine learning, 3rd edn. Prentice and Hall Publication

Image Edge Detection by Mean Difference Thresholding

Smita S. Hande

Abstract Edge detection is very essential step for numerous image processing tasks. In this paper, the different techniques of edge detection, *i.e.*, Roberts, Prewitt, Sobel, Isotropic, Laplacian of Gaussian (LoG), Difference of Gaussian (DoG) and Canny are discussed and also the diagrammatic difference between the some of these techniques are reviewed. This paper describes a simple edge detection technique based on thresholding, which is used to make fast process of edge detection and represents results in more efficient way. The validation of proposed algorithm is done by MSE and PSNR parameters. There are many edge detection techniques available, but any single technique is not suitable for all the applications due to the variation in quality of an image and the variability in shape of an image. So, the selection of the technique is dependent on the requirement of the applications.

Index Terms Edge detection · Gradient · Thresholding · Laplacian of Gaussian (LoG) · Difference of Gaussian (DoG) · Mean square error · PSNR

1 Introduction

An image $f(x, y)$ is mostly a two-dimensional function, where x and y are contiguous coordinates and at a given coordinate, the amplitude of this function gives the value of intensity of an image. When the f value, i.e., intensity of pixel and the value of x, y are finite, then the image is a digital image [1]. A large no. of edge detection techniques forming a group is one of the very old developed and widely used in image processing called content-based [11]. The main aim of edge detection is to find fast intensity variations in image brightness. These intensity variations make boundaries between different parts of image. Boundaries of regions are created for objects, overlapping objects, background, etc. The process of edge detection reduces

S. S. Hande (✉)
Department of EXTC, Fr. Conceicao Rodrigues Institute of Technology, Sector 9A Vashi, Navi Mumbai 400703, Maharashtra, India
e-mail: smita.hande@fcrit.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_30

the required data to be processed and analyzed. Edge detection is the first step in many image processing applications such as image compression, feature extraction, identification of objects in a scene, segmentation, registration, shape recognition and face recognition.

Many gradient operators have been introduced and used for image processing in the literature, such as Robert's edge detection, Sobel edge detection, Prewitt, LoG, DoG and Canny [2–5]. Edges are detected by use of first derivatives, which are called gradient, or second derivatives, which are called Laplacian. Laplacian is very much sensitive to noise as it uses more data because of the nature of the second derivatives.

The 5 × 5 LoG filtering mask introduced by Marr–Hildreth [9] is a default mask, where the mask can be extended dimensionally from the need of the image processing application. For noisy images, the edge detection is difficult to implement, as edges and noise both are high frequency contents. Edge detection operator required to be selected as per gradual changes of intensities results of poor focus or refraction of the objects having boundaries. The problems of missing true edges, false edge detection, edge localization and high computational time can be prevented. The objective of this research work is comparison and analysis of the performance of the various edge detection techniques under different conditions.

This research work covers the following: Sect. 2 introduces the previous works concerned with different techniques of edge detection introduced in processing of images. Section 3 presents the conventional Roberts, Prewitt and Sobel edge detection operator for gradient images and a brief description of Isotropic, LoG and DoG edge detection. Section 4 gives the proposed scheme of thresholding-based edge detection. The results are discussed in Sect. 5 and in Sect. 6 some concluding remarks are discussed.

2 Edge Detection Related Works

The pixels on the boundary are connected with each other in the edge detection technique, to form the edge between two regions which differ in the intensity value [1, 6]. These can be used to find discontinuities in intensities of gray level images. The basic four types of edges to find the different image shapes are shown in Fig. 1.

The operators are used to locate diagonal, horizontal or vertical edges and corner of edges as well. Operators are matrix of 3 × 3 pixels, coefficient values which are used to analyze neighborhood pixel values. The operator. The resultant operator value is the sum of product of pixel value and operator coefficient values. The different types of edge detection operators which are used for image edge detection are grouped as: [6]

(1) First-order derivative [12]

- Roberts
- Prewitt operator
- Sobel operator

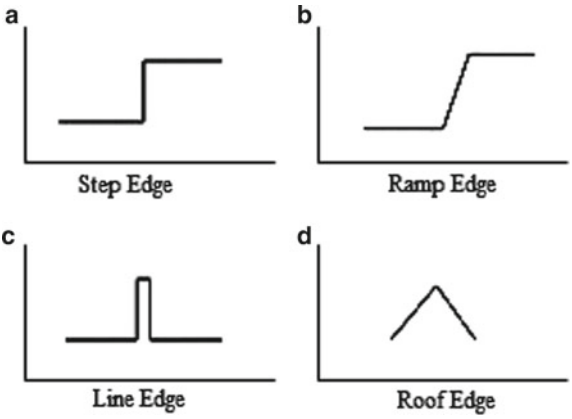


Fig. 1 Basic four types of edges [1, 6]. **a** Step edge, **b** ramp edge, **c** line edge, **d** roof edge

(2) Second-order derivative

- Laplacian operator

(3) Other

- LoG [4, 8]
- DoG
- Canny [7].

Gradient operators of first-order derivative generate image gradient value. Equation (1) shows first-order derivatives of image for value determination. *Gx* and *Gy* are horizontal and vertical components of local derivative. The mask of operator is used to find values of *Gx* and *Gy*.

$$\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right] \tag{1}$$

$$\begin{aligned} \text{mag}(\nabla f) &= \left[Gx^2 + Gy^2\right]^{1/2} \\ &= |Gx| + |Gy| \end{aligned} \tag{2}$$

The direction of gradient operator is as:

$$\alpha(x, y) = \tan^{-1}(Gy/Gx) \tag{3}$$

To calculate the gradient, two operators in opposite orientation horizontal and vertical are used. Edge points are found by thresholding and localization of local extrema values from a set of operator results. The edge direction at (x, y) point is perpendicular to the direction of the gradient vector.

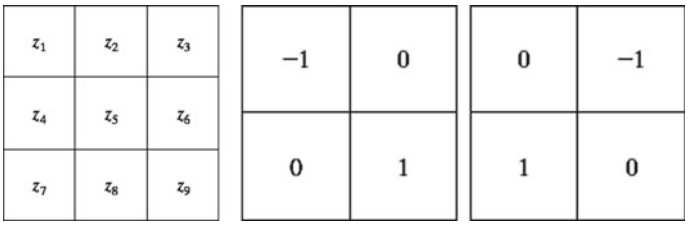


Fig. 2 Convolution masks for Robert operator [1]

3 Conventional Edge Detection Operators

This section describes most widely used conventional operators and its mask.

3.1 Robert Cross-Gradient Operators

This operator has a simple contiguity mask 2 * 2. Perpendicular and parallel directions are determined, respectively, by applying these two mask separately to the image. This provides results of *Gx* and *Gy*.:

$$Gx = (z_9 - z_5) \tag{4a}$$

$$Gy = (z_8 - z_6) \tag{4b}$$

The even size masks do not have a center of uniformity so are difficult to implement (Fig. 2).

3.2 Prewitt Operator

By filtering an image with the kernel or left mask, this operator can be implemented. The image is again filtered with other mask to perform the square of pixels values. These two results are added and root is computed.

$$\begin{aligned} \text{mag}(\nabla \text{f}) \approx &|(\mathbf{z}_7 + \mathbf{z}_8 + \mathbf{z}_9) - (\mathbf{z}_1 + \mathbf{z}_2 + \mathbf{z}_3)| \\ &+|(z_3 + z_6 + z_9) - (z_1 + z_4 + z_7)| \end{aligned} \tag{5}$$

The 3 * 3 masks for Prewitt are shown in Fig. 3.

Fig. 3 Convolution masks for Prewitt operator [1]

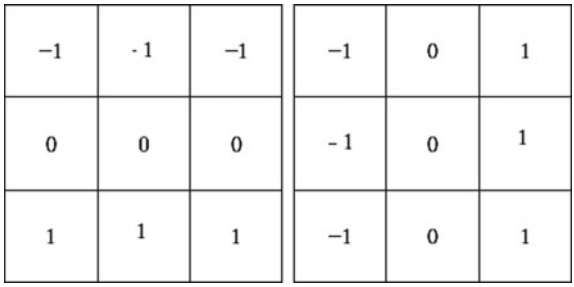
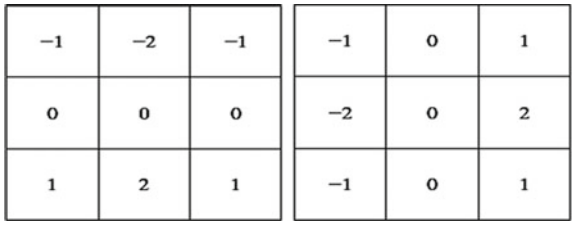


Fig. 4 Convolution masks for Sobel operator [1]



3.3 Sobel Operator

Sobel operator is widely used in edge detection methods. By taking the difference between the third and first rows of 3 * 3 image region the approximation of the partial derivative in x-direction is implemented by left mask or kernel of Fig. 4. The y-direction mask is implemented by taking the difference between the third and first columns to approximate the derivative.

$$\begin{aligned} \text{mag}(\nabla f) \approx\ & |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| \\ & + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)| \end{aligned} \tag{6}$$

3.4 Laplacian Operator

Second-order derivative operators are based on detecting points of zero-crossing values. These points are nothing but local extreme pixels values of image. Equation (7) shows second-order derivatives value determination method of an image, where ∇^2 f is calculated by a single mask of the operator.

$$\nabla^2 f = \partial^2 f / \partial^2 x + \partial^2 f / \partial^2 y \tag{7}$$

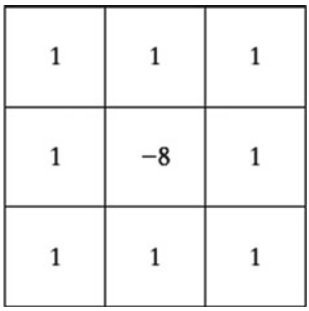


Fig. 5 Widely used masks for Laplacian operator [1]

These results of edge detection are almost similar to first-order derivative. In these operators, only one operator is used but lose information about orientation of the edge results and these operators are sensitive to sharp noise.

The gradiant for Laplacian is found by Eq. (8) as shown below.

$$\left|\nabla^2 f\right| = |\{(Z_1 + Z_2 + Z_3 + Z_4 + Z_6 + Z_7 + Z_8 + Z_7) - 8 \times Z5\}| \quad (8)$$

The widely used second derivative mask of Laplacian is shown in Fig. 5.

The limitations of Laplacian are Laplacian gives very large values for noisy images and thus ruins the entire image. An undesirable effect of double edges is produced by the Laplacian operator which complicates segmentation. Laplacian is unable to detect the direction of edge.

3.5 LoG the Laplacian of Gaussian Operator

Laplacian of a Gaussian (LoG) is the most popular operator, this is the combination of Gaussian operator for smoothing and Laplacian second-order operator [4, 8, 10]. Also known as Marr–Hildreth operator. This operator is used in many image comparing algorithms, which results in a map of marked points of detected edges with identical size as that of input image. Usually, images are converted into easily readable maps for human. All standard edge detection methods are vulnerable to blur noise. Only small area of the image is analyzed due to the size of operator matrix. Larger matrix is preferred.

$$\nabla^2 h = \frac{(r^2 - \sigma^2)}{\sigma^4} \mathrm{e}^{\frac{r^2}{2\sigma^2}} \quad (9)$$

This function is also known Mexican hat function, as it looks Mexican hat, when it is plotted. Only one mask to compute the second-order derivative is used by Laplacian edge detector unlike the Sobel operator (Fig. 6).

Fig. 6 Widely used masks for LoG operator. [1]

0	0	−1	0	0
0	−1	−2	−1	0
−1	−2	16	−2	−1
0	−1	−2	−1	0
0	0	−1	0	0

3.6 DoG Difference of Gaussian Operator

The Laplacian of Gaussian can be approximated by the difference between two Gaussian functions.

$$\nabla^2 G \approx G(x, y; \sigma_1) - G(x, y; \sigma_2) \tag{10}$$

The DoG is not separable. By simply taking the difference between two Gaussian convolutions with different sigma values an efficient implementation is possible.

4 Proposed Scheme

In an image, the edge information is found by searching the relationship a pixel has with its neighborhoods.

If a pixel's value is similar to pixel values those around it, it can be said there is no edge at that point. If a pixel has neighbor pixels with widely varying intensity levels, it may present an edge point.

4.1 Algorithm

1. Set the threshold value '*T*' by mean value of all standard deviation values of columns.
2. Find all eight difference values of each pixel with its eight neighborhood pixels. Say I1 to I8.
3. If any of this I1 to I8 difference value is greater than 'T' edge is detected so save '1' in image else save '0'.
4. '1' indicates large intensity variations due to edge and '0' indicates no edge detected.

4.2 Validation

Validation of the results is done by MSE and PSNR on images. PSNR is used to measure the quality of lossy and lossless compression, e.g., images. Validation between original data and the noise is the error after compression. PSNR is an approximation to human perception of reconstruction quality. A high PSNR indicates the reconstruction is of high quality, in some cases, it may not. Mean squared error is used to define the PSNR,

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

The PSNR is defined as:

$$\begin{aligned} \text{PSNR} &= 10 \cdot \log_{10} \left(\frac{\text{MAX}_1^2}{\text{MSE}} \right) \\ &= 20 \cdot \log_{10} \left(\frac{\text{MAX}_1}{\sqrt{\text{MSE}}} \right) \end{aligned} \tag{10}$$

Here, MAXI is 255 for images as it is the maximum possible pixel value of the image.

5 Results

This section gives results of all conventional edge detectors and the proposed scheme.

Figure 7 shows the results of implemented algorithms for some conventional gradient operators and Laplacian operator. Result shows better edge detections for Sobel and Laplacian algorithm than the Prewitt operator. Figure 8 shows the results of Sobel operator and the proposed scheme of thresholding-based edge detection algorithm. The results of proposed scheme depend on the threshold value selection. So by keeping less value and a large value results are observed on many test images. Figure 8a shows false edge detections due to small value of T and Fig. 8b shows loss of data due to large value of selection of T.

Figure 9 shows results for threshold T selected as mean of standard deviation values of all columns. For better results, one can select the appropriate value of threshold. Threshold value also depends on the image quality and the application.

The validation of the proposed algorithm by parameters MSE and PSNR is shown in Table 1. The comparison with conventional algorithm shows that the proposed algorithm is simple but gives the good result, almost similar to conventional methods.

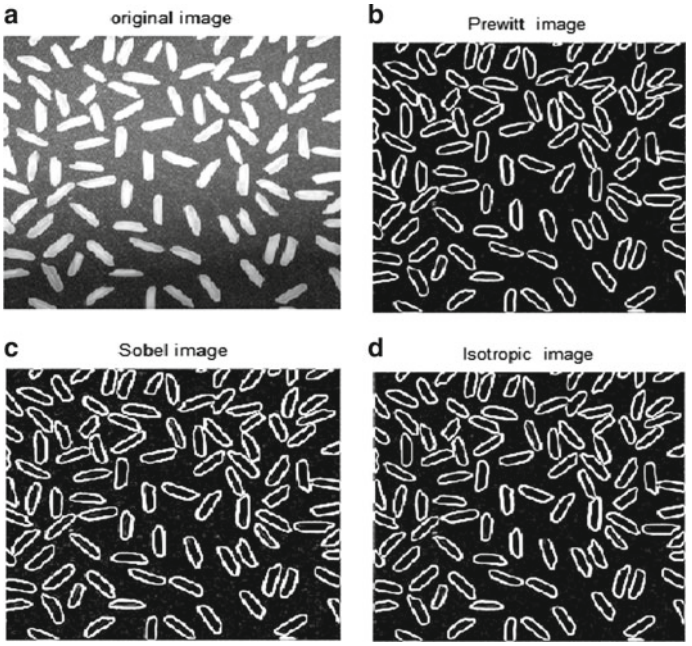


Fig. 7 Results of conventional edge detection operators. **a** Original image, **b** Prewitt, **c** Sobel, **d** Isotropic

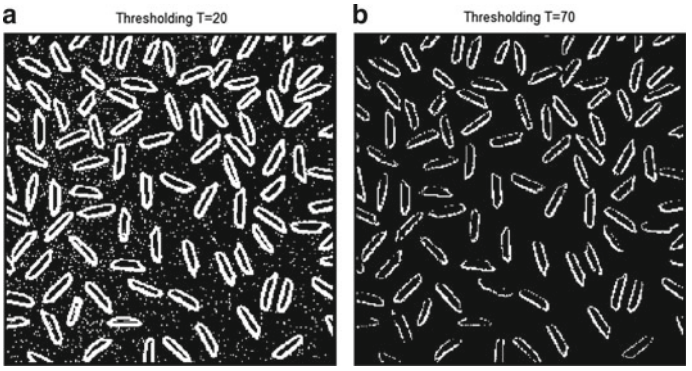


Fig. 8 Results of proposed scheme for different values of T. **a** Result for $T = 20$, **b** result for $T = 70$

6 Conclusion

In this paper, a study of gradient-based and Laplacian-based edge detection operators, which are most commonly used techniques, is carried out as a start of research work. The algorithm is developed using MATLAB. Prewitt edge detector which is

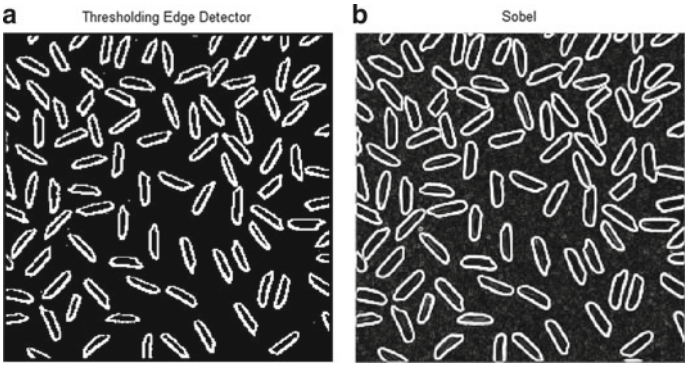


Fig. 9 Results of proposed scheme. **a** Proposed scheme, **b** Sobel operator

Table 1 Validation by MSE and PSNR

Algorithm		Prewitt	Sobel	Isotropic	Proposed Algorithm
Coin	MSE	232.56	227.73	230.64	231.11
	PSNR (dB)	24.46	24.55	24.50	24.49
Rice	MSE	205.86	195.59	201.53	197.73
	PSNR (dB)	24.99	25.21	25.08	25.17
Cell	MSE	236.34	229.29	233.41	203.96
	PSNR (dB)	24.39	24.52	24.44	25.03
Cameraman	MSE	181.08	169.65	176.26	194.16
	PSNR (dB)	25.55	25.83	25.66	25.24

a gradient-based algorithm and is an appropriate way to estimate the magnitude and orientation of an edge but it has a drawback of being sensitive to noise. Sobel and Laplacian operator give better results than Prewitt. The proposed scheme is a simple edge detection technique based on thresholding, which is used to make process of edge detection fast and represents results in more efficient way. The proposed scheme result depends on the threshold value selection. If the threshold value is too less, it results in false edge detection and if the threshold value is large it results in loss of data. The proposed scheme cannot be used to find the direction of edge as in case of other gradient operators.

References

1. Gonzalez and Woods (2008) Digital Image Processing, 3rd edn. Pearson Edition, India
2. Liao PS, Chen T-S, Chung P-C (2001) A fast algorithm for multilevel thresholding. J Inf Sci Eng 17(5):713–727

3. Aurich V, Weule J (1995) Nonlinear Gaussian filters performing edge preserving diffusion. In: Proceedings of 17th Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) Symposium, Sept. 13–15, Springer, Bielefeld, Germany, pp 538–545
4. Basu M (1994) A Gaussian derivative model for edge enhancement. J Pattern Recogn 27(11):1451–1461
5. Canny J (1986) A computational approach to edge detection. IEEE Trans Patt Anal Mach Intell 8:679–698
6. Jain AK (2002) Fundamentals of digital image processing, 2nd edn. Prentice Hall, Upper Saddle River
7. Senthilkumaran N, Rajesh R (2009) A study on edge detection methods for image segmentation. In: Proceedings of international conference on mathematics and computer science (ICMCS-2009), vol I, pp 255–259
8. Mitra B (2002) Gaussian based edge detection methods—a survey. IEEE Trans Syst Manand Cybern 32:252–260
9. Marr D, Hildreth E (1980) Theory of edge detection. Proc R Soc Lond Ser B, Biol Sci 207:187–217
10. Jia D, Xiao C-L, Sun J-G (2015) Edge detection method of Gaussian block distance. In: 2015 IEEE International Conference on Image Processing (ICIP), pp 3049–3053, 27–30 Sept 2015
11. DingHui GH (2009) A survey of edge detection in image processing. J Capital Normal Univ 30
12. Zhu Z, Liu G, Liu Q (2009) Study of first-order edge detection algorithm. Modern Electronics Technique 24
13. Chen W-d, Dong Z-L, Li Y (2008) Rice edge detecton based on canny operator. J HenanUniv Technol (Nat Sci Edn) 29

Mathematical Modeling and Simulation of a Nanorobot Using Nano-hive Tool for Medical Applications

B. H. Raghunath, H. S. Aravind, N. Praveen, P. Dinesha, and T. C. Manjunath

Abstract A brief review of the nanorobots that are currently used in the biomedical engineering to cure various types of diseases is being presented. In the context, we are carrying out a literature survey for the treatment of cancer using the nanotechnology concept. The way we have picked is the union of nanotechnology and medicine. The mix of nanotechnology into medication is probably going to get some new difficulties restorative treatment. Nanorobot is a superb vision of medicine in the future. The most exceptional nanomedicine includes the utilization of nanorobots as small-scale specialists. Advancement in nanotechnology may allow us to build artificial red blood cells called respirocytes capable of carrying oxygen and carbon dioxide molecules (i.e., functions of natural blood cells). Respirocytes are nanorobots, small mechanical gadgets intended to work on the atomic level. Respirocytes can give a brief substitution to characteristic platelets in the instance of a crisis. Subsequently, respirocytes will truly change the treatment of coronary illness. We can imagine a day

B. H. Raghunath
Dept of ECE, VTU RS, Acharya Institute of Technology, Soldevanahalli, Bangalore, Karnataka, India
e-mail: raghunath@acharya.ac.in

H. S. Aravind
Dept of ECE, JSS Academy of Technical Education, Dr Vishnuverdhana Road, Bangalore 560060, Karnataka, India
e-mail: aravindhs1@gmail.com; aravindhs@jssateb.ac.in

N. Praveen
Department of Electrical and Electronics Engineering, Faculty in Engineering, Ibra College of Technology, Ibra, Sultanate of Oman, Oman
e-mail: praveen0374@gmail.com

P. Dinesha
Department of ECE, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India
e-mail: dineshprg@gmail.com

T. C. Manjunath (✉)
Department of ECE, DSCE, Bangalore, Karnataka, India
e-mail: dr.manjunath.phd@ieee.org

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_31

when you could infuse billions of these nanorobots that would skim around in your body. A standout among the most sensible and almost doable accomplishments is the cure for growth which is one of the primary centers of this work. Nanorobots could carry and deliver large amounts of anti-cancer drugs into cancerous cells without harming healthy cells, reducing the side effects related to the current therapies. These nanorobots will have the capacity to repair tissues, clean veins and aviation routes, change our physiological capacities. The work presented in this paper is this research work of the research scholar student that was undertaken under the guidance of the doctorates which just provides a brief review of the applications of the nanorobots that could be used in the medicine for the curing of the cancer treatment and is just a review paper, which serves as a basis for all the students, faculties as a base for carrying out the research in this exciting field of nanorobotics. At the end, once the review is over, we are presenting a simulation which we had developed in nanohive simulation tool for biomedical engineering applications.

General Terms Nanorobots · Medicine

Keywords Cancer · Treatment · Coronary · Artery

1 Introduction

A nanorobot can be defined as an artificially fabricated object able to freely diffuse in the human body and interact with specific cell at the molecular level. The figure shows a schematic representation of a nanorobot that can be activated by the cell itself when it is needed. Stress induced by disease or infectious attack generally leads to changes in the chemical content of the cell which in turn trigger the nanorobots. Nanorobots can be coated with different agent depending on their application or tissue destination. The external shell is a crucial point because it has to be recognized as a part of the body (inert coating) and be able to release different ideal matrix that it is not toxic at the nanometer level. The pore size can be tuned permitting release of different size sizemolecules (tunable porosity).

A rigid shell like silica is an molecules. The surface is easily functionalizable with simple chemical methods, but most important is that the silica is not biodegradable permitting a long-term activity in the body. Nanorobots will be able to analyze each cell type surface antigens to recognize if the cell is healthy, what the parent organ is, as well as almost all information about the cell, and using chemotactic sensors, certain molecules and cells could be identified and easily targeted for action. A 1 cm^3 injection of nanorobots would be able to deliver selectively into cells at least 0.5 cm^3 of chemical agent, and the sensors could test levels of the chemical to guard against an accidental overdose. Self-assembly and nano-manipulation are two main ways for the production of nanorobots. Figure 1 shows the robots which are used to attack the cancerous cells.

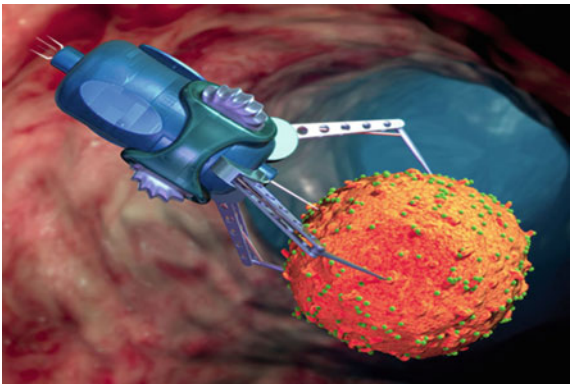


Fig. 1 Nanorobots which are used to attack the cancerous cells

2 Overview

We will describe a mobile robot (nanorobot) that can be created with the existing technology that can be used to seek out and destroy inimical tissue within the human body that cannot be accessed by other means. The construction and use of such devices would result in a number of benefits. It would provide either cures or at least a means of controlling or reducing the effects of a number of ailments. Nanotechnology has the potential to radically increase our options for prevention, diagnosis, and treatment of cancer. Nanotechnology may also be useful for developing ways to eradicate cancer cells without harming healthy neighboring cells. Nanotechnology uses therapeutic agents that target specific cells and deliver their toxin in a controlled, time-released manner. As a syringe is today used to inject medication into the patient's bloodstream,tomorrow, nanorobots could transport and deliver chemical agents directly to a target cell.

Nanokiller (i.e., nanorobot) could find and repair damaged organs, detect, and destroy a tumor mass. They would be able to communicate their positions, operational statuses, and the success or failure of the treatment as it progresses. They would tell you how many cancer cells they have encountered and inactivated. Respirocytes identify tumors and then allow the nanokiller to kill cancerous cells with a tiny but precise amount of a chemotherapy drug. It would not only find cancers in their earliest stages before they can do damage or spread, but also deliver a small amount of a drug targeted directly at tumors, which would cause little or no side effects. Nanomedicine could result in noninvasive devices that can enter the body to determine glucose levels, distinguish between normal and cancerous tissue, and destroy the tumor in the initial stage itself. This nanorobot has vibrating cilia-like structures with inbuilt nanosensors to detect the cancerous tissue. This nanorobot shells are specially coated with gold that allow them to attach to malignant cells, and they can deliver the drug internally.

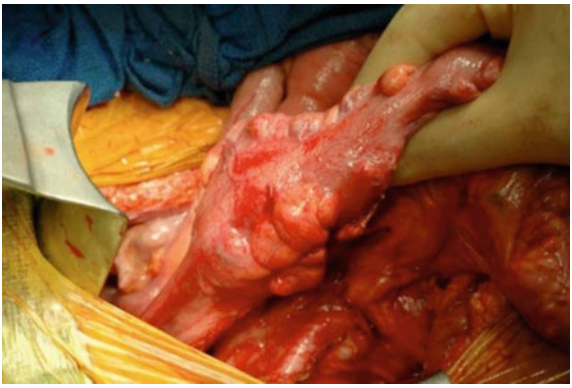


Fig. 2 Removal of cancer by surgery using the conventional method

3 Conventional Methods of Cancer Therapy Treatments

Most people with cancer receive surgery, chemotherapy, radiation therapy, or other conventional therapies at some point during treatment, and many will have a combination of these treatments. Injection of drugs affects both cancerous and non-cancerous cells in conventional method. So, the main types of cancer treatment include: surgery, radiation therapy, chemotherapy, immunotherapy, targeted therapy, hormone therapy. Figure 2 shows the removal of cancer by surgery using the conventional method.

4 The Major Drawbacks of the Conventional Methods Used in the Cancer Therapy

Injection of drugs affects both cancerous and non-cancerous cells in conventional method. There are various side effects with the conventional methods as it effects both the cancerous and non-cancerous cells. Surgery and radiation therapy remove, kill, or damage cancer cells in a certain area which also affects healthy cells. The time to heal is longer with the conventional methods that include methods like surgery and radiation therapy. Some of the side effects that occur when conventional method is followed are.

Fatigue—This is a persistent feeling of physical, emotional, or mental exhaustion. Cancer-related fatigue differs from feeling tired due to lack of rest. Receiving multiple treatment types may increase your fatigue, for example, having chemotherapy and radiation therapy.

5 The Modern Technology Used to Cure Cancer Therapy

Nanotechnology uses therapeutic agents that target specific cells and deliver their toxin in a controlled, time-released manner. As a syringe is today used to inject medication into the patient's bloodstream, tomorrow, nanorobots could transport and deliver chemical agents directly to a target cell. Nanokiller (i.e., nanorobot) could find and repair damaged organs, detect and destroy a tumor mass. They would be able to communicate their positions, operational statuses, and the success or failure of the treatment as it progresses. They would tell you how many cancer cells they have encountered and inactivated. Respirocytes identify tumors and then allow the nanokiller to kill cancerous cells with a tiny but precise amount of a chemotherapy drug.

It would not only find cancers in their earliest stages before they can do damage or spread, but also deliver a small amount of a drug targeted directly at tumors, which would cause little or no side effects. Nanomedicine could result in noninvasive devices that can enter the body to determine glucose levels, distinguish between normal and cancerous tissue, and destroy the tumor in the initial stage itself. This nanorobot has vibrating cilia-like structures with inbuilt nanosensors to detect the cancerous tissue. This nanorobot shells are specially coated with gold that allow them to attach to malignant cells, and they can deliver the drug internally.

There are three main considerations which are need to be focused on designing a nanorobot to move through the body—navigation, power, and how the nanorobot will move through blood vessels. For directing the nanorobots to the cancerous cells, we can make use of ultrasonic signals which are emitted by the nanorobot. These ultrasonic waves are detected by ultrasonic sensors. Nanorobots can also be fitted with a small miniature camera assuming that the nanorobot is not tethered or designed to float passively through the bloodstream, and it will need a means of propulsion to get around the body. Because it may have to travel against the flow of blood, the propulsion system has to be relatively strong for its size. Another important consideration is the safety of the patient; the system must be able to move the nanorobot around without causing damage to the host.

For locomotion, we can mimic Paramecium. It moves through their environment using tiny tail-like limbs called cilia. By vibrating the cilia, the paramecium can swim in any direction. Similar to cilia are flagella, which are longer tail structures. Organisms whip flagella around in different ways to move around. The nanorobot would move around like a jet airplane. Miniaturized jet pumps could even use blood plasma to push the nanorobot forward, though, unlike the electromagnetic pump, there would need to be moving parts. For powering a nanorobot, we could use the patient's body heat to create power, but there would need to be a gradient of temperatures to manage it. Power generation would be a result of the Seebeck effect. Figure 3 shows the nanorobots moving inside the blood stream in the veins/capillaries to destroy the cancerous cells.

The Seebeck effect occurs when two conductors made of different metals are joined at two points that are kept at two different temperatures. The metal conductors

Fig. 3 Nanorobots moving inside the blood stream in the veins/capillaries to destroy the cancerous cells

become a thermocouple, meaning that they generate voltage when the junctures are at different temperatures. Since it is difficult to rely on temperature gradients within the body, it is unlikely that we will see many nanorobots using body heat for power. These are some of the most practical ideas which have been implemented in real life.

6 Advantages of the Modern Technology Employed

With the help of nanorobots, we can further understand the complexity of human body and brain. The development will further help in performing painless and noninvasive surgeries. Even the most complicated surgeries will be done with ease. Due to their microscopic features, they could surf through the brain cells and generate all the related information required for further studies. Scientist specially will be benefited from this nanotechnology application.

The best part is the nanobot which is so small that it is not visible with naked eyes, so they can be injected in a human body very easily. Days in future will be like this when a single shot will cure diseases. More than million people in this world are affected by this dreaded disease. Currently, there is no permanent vaccine or medicine available to cure the disease. The currently available drugs can increase the patient's life to a few years only, so the invention of this nanorobot will make the patients to get rid of the disease. As the nanorobot do not generate any harmful activities, there is no side effect. It operates at specific site only. The initial cost of development is only high but the manufacturing by batch processing reduces the cost.

7 Literature Survey

A large number of scientists, researchers, and students had worked on the curing of the cancer treatment using the nanorobots. Here follows a brief review of the same.

In the 1980s by Nobel Prize laureate Richard Smalley. Smalley has extended his vision to carbon nanotubes, discovered by Sumio Iijima, which he envisions as the next superinter connection for ultra-small electronics. The term nanotechnology has evolved to mean the manipulation of the elements to create unique and hopefully useful structures.

December 29, 1959: Richard Feynman gives the famous "There's Plenty of Room at the Bottom" talk. First use of the concepts of nanotechnology. Describes an individual atoms and molecules can be manipulated.

1974: Professor Norio Taniguchi defines nanotechnology as "the processing of, separation, consolidation, and deformation of materials by atom/molecule."

1980s: Dr. Eric Drexler publishes several scientific articles promoting nanoscale phenomena and devices.

1986: The book Engines of Creation: The Coming Era of Nanotechnology by Dr. Eric Drexler is published. He envisioned nanorobots as self-replicating. A first book on nanotechnology [5].

1981: Gerd Binnig and Heinrich Rohrer of IBM Zürich. Invented of the Scanning Tunneling Microscope (STM)By Used for imaging surfaces at the atomic level and identifying some properties (i.e., energy).

1985: Discovery of fullerenes (molecules composed entirely of carbon). They have many applications in materials science, electronics, and nanotechnology.

1991: Discovering Carbon nanotubes (cylindrical fullerenes) as direct result of the fullerenes. Exhibit high tensile strength, unique electrical properties, and efficient thermal conductivity. Their electrical properties make them ideal circuit components (i.e., transistors and ultra-capacitors). Recently, researched chemical and biomedical engineering have used carbon nanotubes as a vessel for delivering drugs into the body [5].

1991: Invented of atomic force microscope (AFM)—it has imaging, measuring, and manipulating matter at the nanoscale. It performs its functions by feeling the surface with mechanical probe. Since interaction with materials on the nanoscale, it is considered a nanorobot.

2000: United States national nanotechnology Initiative is founded to coordinate federal research and development in nanotechnology.

2000: The company nano factory collaboration is founded. Developing a research agenda for building a nano factory capable of building nanorobots for medical purposes.

Currently, DNA machines (nucleic acid robots) are being developed. Performs mechanical-like movements, such as switching, in response to certain stimuli (inputs).

Molecular size robots and machines paved the way for nanotechnology by creating smaller and smaller machines and robots. The applications of the nanorobotics are

more as: microrobotics, emerging drug delivery application, health care, biomedical application, cancer therapy, brain aneurysm, communication system, and new future nanotechnologies, etc.

The major development of nanomedicine molecular nanotechnology (MNT) or nanorobotics. Just as biotechnology extends the range and efficacy of treatment available from application of nanomaterials, the advent of molecule of nanotechnology will again expand enormously the effectiveness, precision, and speed of future medical treatments while at the same time significantly reducing their risk, cost, and invasiveness.

MNT will allow doctors to perform direct in vivo surgery of human cells. Nanomedicines can easily traverse the human body because nanorobots are so tiny. Scientists report that nanorobot is constructed of carbon atoms in a diamonded structure because of its inertproperties and strength. Glucose or natural body sugars and oxygen might be a source for propulsion, and it will have other biochemical or a molecular part depends on task.

A large potential applications for nanorobotics in medicine include early diagnosis and targeted drug delivery with treat mental medicine for cancer biomedical instrumentation, surgery, pharmacokinetics, monitoring of diabetes, and health care. In the future, medical technology is expected to nanorobots injected into the patient to perform treatment on a cellular level.

8 Main Objective of Nanorobotics in Medical Fields

The main objective of the nanorobotics in the medical fields is listed as follows.

- To help monitor the patient's body continuously and be able detect cancer and other diseases at early stages.
- To destroy cancerous cells without affecting the healthy cells.
- To reduce the time of recovery for people fighting against cancer and other diseases.
- To carry and deliver large amounts of anti-cancer drugs into cancerous cells without harming healthy cells, thus reducing the side effects related to the current therapies.
- To repair tissues, clean blood vessels and airways, transform our physiological capabilities, and even potentially counteract the aging process.
- The above-mentioned objective of our work can be achieved using the following steps one by one as follows …
- First, finding out the method of entry into the body for the nanorobot
- Finding means of propulsion for the nanorobot
- Finding means of maintaining a fixed position while operating
- Finding how to control of the device
- Finding the appropriate power source to nanorobot
- Finding means of locating substances to be eliminated by the nanorobot

- Finding means of doing the elimination the substance from the body
- Finally, continuously monitoring the body and giving feedback.

9 Design of Nanorobots for Medical Applications

Locomotion: There are a number of means available for active propulsion of our device.

Propeller: The very first Feynman prize in Nanotechnology was awarded to William McLellan for building an electric motor that fit within a cube 1/64th of an inch on a side. This is probably smaller than we would need for our preliminary microrobot. One or several of these motors could be used to power propellers that would push (or pull) the microrobot through the bloodstream. We would want to use a shrouded blade design so as to avoid damage to the surrounding tissues (and to the propellers) during the inevitable collisions.

Cilia/flagellae: In this scenario, we are using some sort of vibrating cilia (similar to those of a paramecium) to propel the device. A variation of this method would be to use a fin-shaped appendage. While this may have its attractions at the molecular level of operation, an electric motor/propeller combination would be more practical at the scale we are talking about.

Electromagnetic pump: This is a device with no moving parts that takes conductive fluid in at the front end and propels it out the back, in a manner similar to a ramjet, although with no minimum speed. It uses magnetic fields to do this. It would require high field strengths, which would be practical with high-capacity conductors. At the scale we are talking about, room (or body)-temperature ceramic superconductors are practical, making this a possibility.

Jet Pump: In this scenario, we use a pump (with moving parts) to propel blood plasma in one direction, imparting thrust in the opposite direction. This can either be done with mechanical pumps, or by means of steam propulsion, using jets of vaporized water/blood plasma.

Membrane propulsion: A rapidly vibrating membrane can be used to provide thrust, as follows: imagine a concave membrane sealing off a vacuum chamber, immersed in a fluid under pressure that is suddenly tightened. This would have the effect of pushing some of the fluid away from the membrane, producing thrust in the direction toward the membrane. The membrane would then be relaxed, causing the pressure of the fluid to push it concave again. This pressure would impart no momentum to the device, since it is balanced by the pressure on the other side of the device. At the macroscale, this thrust is not significant, but at the microscale, it is a practical means of propulsion.

Crawl along surface: Rather than have the device float in the blood, or in various fluids, the device could move along the walls of the circulatory system by means of appendages with specially designed tips, allowing for a firm grip without excessive damage to the tissue. It must be able to do this despite surges in the flow of blood

caused by the beating of the heart, and do it without tearing through a blood vessel or constantly being torn free and swept away.

Navigation: This information will be used to navigate close enough to the operations site that short-range sensors will be useful.

Ultrasonic: This technique can be used in either the active or the passive mode. In the active mode, an ultrasonic signal is beamed into the body, and either reflected back, received on the other side of the body, or a combination of both. The received signal is processed to obtain information about the material through which it has passed. This method is, of course, greatly similar to those used in conventional ultrasound techniques, although they can be enhanced greatly over the current state of the art.

In the passive mode, an ultrasonic signal of a very specific pattern is generated by the microrobot. By means of signal processing techniques, this signal can be tracked with great accuracy through the body, giving the precise location of the microrobot at any time. The signal can either be continuous or pulsed to save power, with the pulse rate increasing or being switched to continuous if necessary for more detailed position information.

In the passive mode, the ultrasonic signal would be generated by means of a signal applied to a piezoelectric membrane, a technology that has been well developed for at least a decade. This will allow us to generate ultrasonic signals of relatively high amplitude and great complexity.

NMR/MRI: This technique involves the application of a powerful magnetic field to the body and subsequent analysis of the way in which atoms within the body react to the field. It usually requires a prolonged period to obtain useful results, often several hours, and thus is not suited to real-time applications. While the performance can be increased greatly, the resolution is inherently low due to the difficulty of switching large magnetic fields quickly, and thus, while it may be suited in some cases to the original diagnosis, it is of only very limited use to us at present.

Radioactive dye: This technique is basically one of the illuminations. A radioactive fluid is introduced into the circulatory system, and its progress throughout the body is tracked by means of a fluoroscope or some other radiation-sensitive imaging system. The major advantage of this technique is that it follows the exact same path that our microrobot would take to reach the operations site. By sufficiently increasing the resolution of the imaging system, and obtaining enough data to generate a three-dimensional map of the route, it would provide valuable guidance information for the microrobot.

The active form of this technique would be to have a small amount of radioactive substance as part of the microrobot. This would allow its position to be tracked throughout the body at all times. Additionally, since the technique would not require the microrobot to use any power, or require a mechanism of any sort, it would greatly simplify the design of the microrobot. While there are risks from radiation, the amount of radioactive substance used would not be even a fraction of the amount used in radioactive dye diagnosis. Additionally, as advances in electronic sensors continue, the amount of radiation needed for tracking would steadily be reduced.

In fact, infrared sensing techniques are so advanced that we can fully shield the radioactive substance and merely track its heat directly.

X-ray: X-rays as a technique have their good points and bad points. On the plus side, they are powerful enough to be able to pass through tissue and show density changes in that tissue. This makes them very useful for locating cracks and breaks in hard, dense tissue such as bones and teeth. On the other hand, they go through soft tissue so much more easily that an X-ray scan designed to show breaks in bone goes right through soft tissue without showing much detail. On the other hand, a scan designed for soft tissue cannot get through if there is any bone blocking the path of the x-rays. Another problem with x-rays is that it is very difficult to generate a narrow beam, and even if one could be generated, using it to scan an area in fine detail would necessitate prolonged exposure. Consequently, x-rays are useful only for gross diagnosis, for which several of the techniques listed above are far better suited.

Radio/Microwave/Heat: Again, these techniques (really all the same technique) can be used in both passive and active modes. The passive mode for the techniques depends on various tissues in the body generating signals that can be detected and interpreted by external sensors. While the body does generate some very low frequency radio waves, the wavelength is so large that they are essentially useless for any sort of diagnostic purposes of the type we are interested in.

POWER: One major requirement for our microrobot is, of course, power. We have to be able to get sufficient power to the microrobot to allow it to perform all of its required operations. There are two possible paths we can take for this. The first is to obtain the power from a source within the body, either by having a self-contained power supply, or by getting power from the bloodstream. The second possibility is to have power supplied from a source external to the body.

Source within the body: There are a number of possible mechanisms for this scenario. The basic idea is that the microrobot would carry its power supply within itself. It would need enough power to move to the site of the operation, perform its functions, which might be very power intensive, and then exit the body. There are three basic scenarios for onboard power supplies.

Body heat: This method would use body heat to power the microrobot, in effect using the entire body as a power supply. The basic problem with this is that a power supply requires an energy gradient in order to function. In this case, we would need areas of different temperatures, so that we could set up a power flow between them. Since our microrobot would have to be mobile, and operate at full capacity in many different environments, this requirement would be difficult to fulfill.

Power from the bloodstream: There are three possibilities for this scenario. In the first case, the microrobot would have electrodes mounted on its outer casing that would combine with the electrolytes in the blood to form a battery. This would result in a low voltage, but it would last until the electrodes were used up. The disadvantage of this method is that in the case of a clot or arteriosclerosis, there might not be enough blood flow to sustain the required power levels. Also, if the electrodes were ever embedded in anything that blocked their access to the blood, power would drop to zero and stay there. This means that a backup would be required.

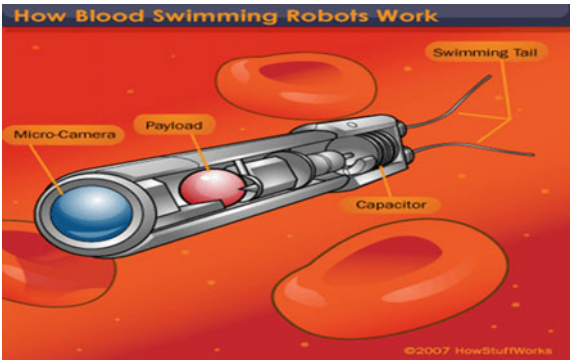


Fig. 4 Nano-machines in the blood vessel

The second way to get power from the bloodstream is by means of a fuel cell, or simply by burning blood chemicals. This is similar to a battery except that rather than obtaining power from the current flow between electrodes, we would obtain power by causing chemical reactions to take place at a controlled rate and obtaining power from this. This is much the same way that the body gets its own power by consuming fuel chemicals from the bloodstream. This has the same problem as the electrode method; it will stop working if access to the blood is blocked, or if the chemicals are not replenished. Figure 4 shows the nano-machines in the blood vessel doing some operation.

10 Possible Outcome of Our Research Work Undertaken

In the approach presented above, nanorobots perform similar tasks on detecting and acting upon medical targets demanding surgical intervention. Each nanorobot is programmed to move through the workspace being tele operated from the surgeons. The fluid flow pushes the concentration of the diffusing signal downstream. Consequently, a nanorobot passing more than a few microns from the source will not detect the signal, while it is still relatively near the source. As an example, the first nanorobot passing close a lymph node may on average detect the higher signal concentration within about 0.16 s. Thus, keeping their motion near the vessel wall, the signal detection happens after these have moved at most 10 mm past the source. Those passing within a few microns often detect the signal, which spreads a bit further upstream and away from the single tumor due to the slow fluid motion near the venule's wall and the cell's motion.

Thus, the present 3D simulation provides guidelines for nanorobot communication and activation control, as well as for sensor manufacturing design. Distinct performances were observed throughout a set of analyses obtained from the NCD software, where the nanorobots also use electromagnetic ultrasound transducers as the communication technique to interact dynamically with the 3D environment and to achieve a

more successful collective coordination. Figure 4 shows the virtual environment in our study, comprised a small venule which contains nanorobots, the red blood cells (RBCs) and a single tumor cell, which is the target area on the vessel wall. Here, the target area is overleaped by the RBCs. In the simulation, the nanorobots search for possible small cancer tumor into the workspace crowded by RBCs, transmitting back information for the surgeons.

11 Applications of Our Research Work Undertaken

There are a large number of applications of the work that we have undertaken as a review/survey paper for the research scholar's preliminary work on the chosen research topic. The applications are listed one by one as below with an in-depth explanation.

Nanorobots in Cancer Detection and Treatment: The current stages of medical technologies and therapy tools are used for the successful treatment of cancer. The important aspect to achieve a successful treatment is based on the improvement of efficient drug delivery to decrease the side effects from the chemotherapy. Nanorobots with embedded chemical biosensors are used for detecting the tumor cells in early stages of cancer development inside a patient's body.

Nanorobotics in Surgery: Surgical nanorobots are introduced into the human body through vascular systems and other cavities. Surgical nanorobots act as semi-autonomous on-site surgeon inside the human body and are programmed or directed by a human surgeon. This programmed surgical nanorobot performs various functions like searching for pathogens, and then diagnosis and correction of lesions by nano-manipulation synchronized by an onboard computer while conserving and contacting with the supervisory surgeon through coded ultrasound signals.

Diagnosis and Testing: Medical nanorobots are used for the purpose of diagnosis, testing and monitoring of microorganisms, tissues and cells in the blood stream. These nanorobots are capable of noting down the record and report some vital signs such as temperature, pressure, and immune system's parameters of different parts of the human body continuously.

Nanorobotics in Gene Therapy: Nanorobots are also applicable in treating genetic diseases, by relating the molecular structures of DNA and proteins in the cell. The modifications and irregularities in the DNA and protein sequences are then corrected (edited). The chromosomal replacement therapy is very efficient compared to the cell repair. An assembled repair vessel is inbuilt in the human body to perform the maintenance of genetics by floating inside the nucleus of a cell.

Nanodentistry: Nanodentistry is one of the topmost applications as nanorobots help in different processes involved in dentistry. These nanorobots are helpful in desensitizing tooth, oral anesthesia, straightening of irregular set of teeth and improvement of the teeth durability, major tooth repairs and improvement of appearance of teeth, etc.

12 Basic Constructional Features of a Nanobot

- Nanorobot has a C-nanotube body
- A bio-molecular n-motor that propels it and peptide limbs to orient itself
- Composed of biological elements such as DNA and proteins, genomes
- Hence, it can be easily removed from the body
- Sensors, molecular sorting rotors, fins and propellers
- <= 6 DOF
- Sensory capabilities to detect the target regions, obstacles
- C is the principal element comprising the bulk of a medical nanorobot.

13 Mathematical Modeling of a Simple 3-D NB

The motion dynamics of a n-robot in the fluid is assumed to be of a cubic polyl as the propulsion model by.

Propulsion model (3D):

$$x_k(t) = x_k(0) + p_0(t)\hat{e}_0 + p_1(t)\hat{e}_1 + p_2(t)\hat{e}_2 p_3(t)\hat{e}_3$$
$$y_k(t) = y_k(0) + q_0(t)\hat{f}_0 + q_1(t)\hat{f}_1 + q_2(t)\hat{f}_2 + q_3(t)\hat{f}_3$$
$$z_k(t) = z_k(0) + r_0(t)\hat{g}_0 + r_1(t)\hat{g}_1 + r_2(t)\hat{g}_2 + r_3(t)\hat{g}_3$$

These are the topics which are used in the modeling of the simple 3-dimensional nanobot.

- Quantum mechanics
- Fluid dynamics
- Hydraulics
- Thermal motions
- Friction, Re
- Fluid flow
- VLSI concepts
- System behavior
- Brownian motions
- Control strategy
- Theory of elasticity (Table 1).

Simulation parameters used for simulating the nanorobot.

- Physical
- Vein/Capillary/Vessel size
- Flow rate through vessel
- Differing diffusion coefficients
- Model

Table 1 Simulation parameters

Chemical signal	
Molecular weight	NOS = 58 kDa
Production rate	$\dot{Q} = 10^4$ molecules s^{-1}
Diffusion coefficient	$D = 100\ \mu m^2$ s–1
Background concentration	6×10^{-3} molecules μm^{-3}
Parameter	Nominal value
Average fluid velocity	$\vartheta = 1000\ \mu m\ s^{-1}$
Vessel diameter	$d = 30\ \mu m$
Workspace length	$L = 60\ \mu m$
Density of nanorobots	$L \times d$: 2 μm × 0.5 μm

- Mathematical chemical based computations
- A 3D environment including bloodstream particles, n-robots, and the proteomic signaling.

14 Simulation Tools

NanoHive-1 is a modular nanosys-simulator used for modeling the physical world at a nanometer scale. Purpose of the simulator is to act as a tool for the study, design, simulation, experimentation and development of nano- and biological entities.

14.1 Actual Buckling Mode—Shape Predicted Using Simulations

The component layout provides easy access to all of the components that can be included in a nanodevice and includes all of the levels of the nanodevice such as the:

- Nanosystems
- Molecular devices
- Device components
- Molecules
- Volumes
- Interfaces and connections
- Nanotubes
- DNA
- Dendritic polymers
- Neurons
- Ribosomes.

Navigating the Map (Simulations)

In each turn, you should decide what action the nanobot has to take.

Some possible actions:

Move To (L/R/U/D/Angle)

Collect From

Transfer To

Attack/kill

Idle

Build Needle

Build Block

Force Auto Destruction.

Figure 5 gives the X-simulation, whereas Fig. 6 gives the simulation motion in x-direction; this is followed by the component layout in Fig. 7. Next, Fig. 8 gives the functional simulation layout, whereas Fig. 9 gives the representation of predicted buckling mode shapes. Figure 10 gives a MNR navigation along the x-direction.

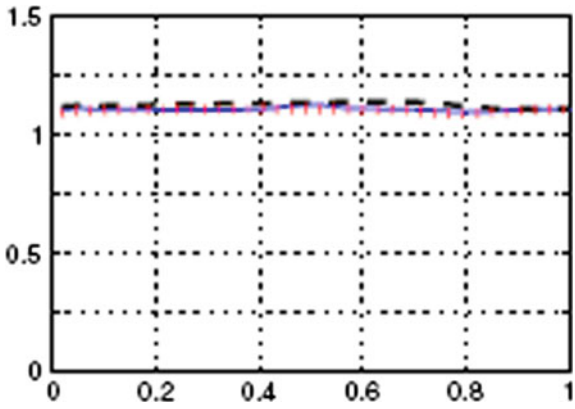


Fig. 5 *X*-simulation

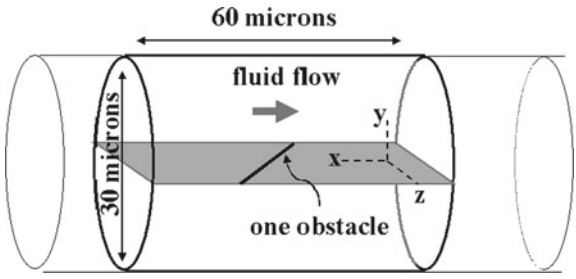


Fig. 6 Simulation motion in *x*-direction

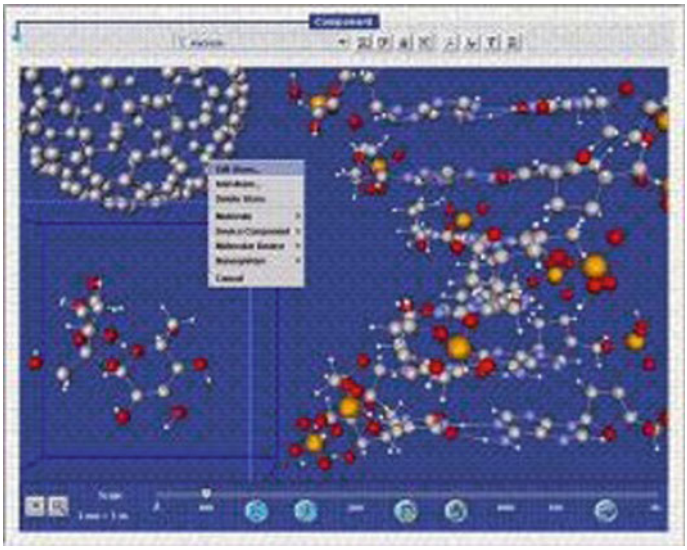


Fig. 7 Component layout

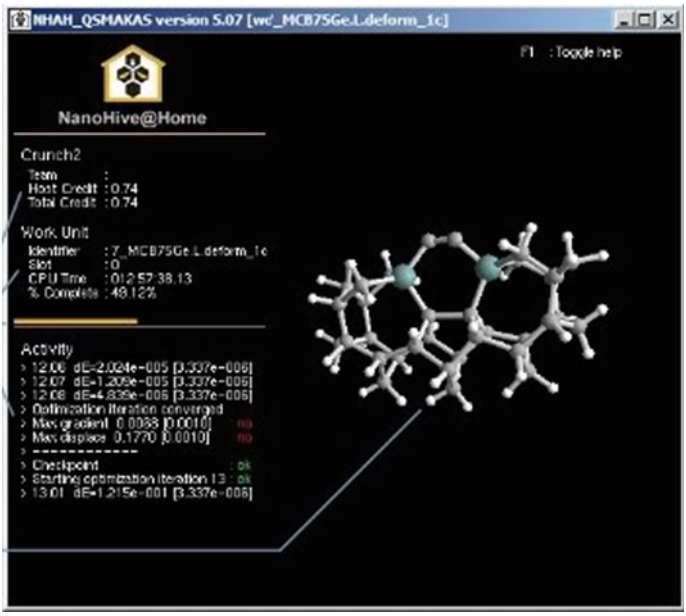


Fig. 8 Functional simulation layout

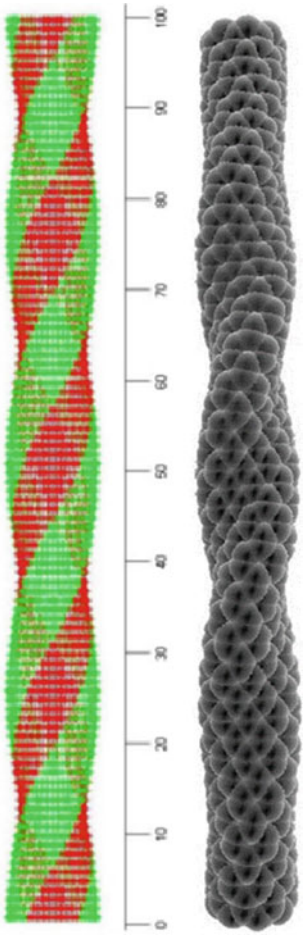
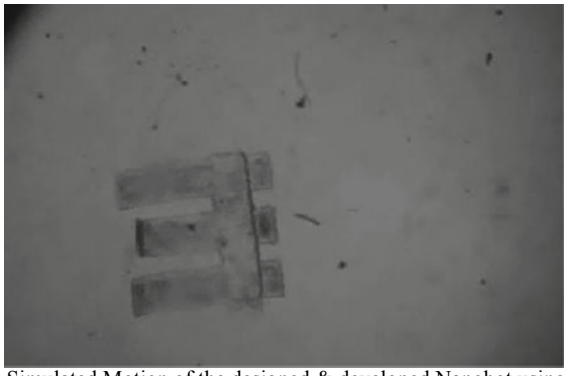


Fig. 9 Representation of predicted buckling mode-shapes

Figures 10 and 11 give the front view of the nano-hive simulation tool and the result of a simulation designed to test various distributed computing mechanisms (Fig. 12).

15 Conclusions

In this paper, we have presented a brief review of the nanorobots that were being used in the biomedical engineering or curing of the deadliest disease in the world, i.e., the cancer. Nature has created nanostructures for billenia. Biological systems are an existing proof of molecular nanotechnology. Rather than keep our eyes fixed on the far future, let us start now by creating some actual working devices that will allow us to cure some of the most deadly ailments known, as well as advance our



Simulated Motion of the designed & developed Nanobot using
Nano Hive-1 Ver 1.2.0 Beta-1 : Nano Systems Simulator
ToolMaterial : C Nano-tubes
Nano Hive-1 is a modular simulator used for
modeling the physical world at a nanometer scale
(adv : model the DNA / ribosomes)

Fig. 10 MNR navigation along the x-direction

Fig. 11 Front view of the nano-hive simulation tool

capabilities directly, rather than as the side effects of other technologies. There will be a day when eliminating cancer cells are mere an outpatient medical procedure and is just a review paper, which serves as a basis for all the students, faculties as a base for carrying out the research in this exciting field of nanorobotics. In the 2nd part of the paper, the mathematical modeling and the simulation of a nanobot using the nanohive software are being presented along with the simulation results.

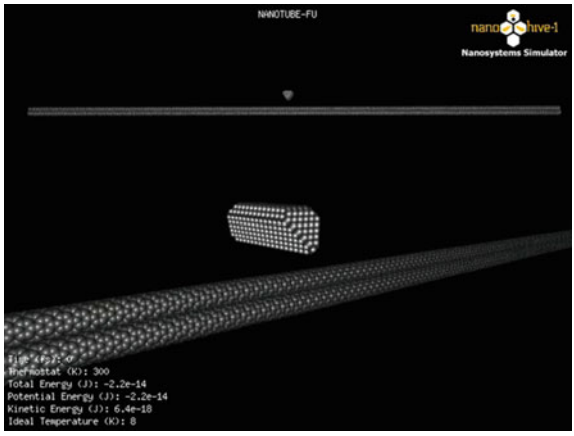


Fig. 12 Result of a simulation designed to test various distributed computing mechanisms. While just a test, it is still an interesting simulation to watch. A diamonded carbon "knife" is pushed down on the nanotubes with a 5 nN force. Will the knife cut through the nanotubes? The system comprises ~20,000 atoms and runs for 5.5 ps of sim time

16 Organization of the Research Article

The paper is organized as follows. The abstract is being presented in a concised manner in the beginning of the paper. A brief information about the matter presented in this paper is given in the introductory section. This is followed by the overview of the work in Sect. 2. The conventional methods of the cancer therapy treatments is given in Sect. 3 followed by the major drawbacks of the conventional methods used in the cancer therapy in Sect. 4. Section 5 presents the modern technology used to cure cancer therapy, whereas Sect. 6 gives the advantages of the modern technology employed. This is followed by an abridged literature survey, where a large number of research articles by various presenters are being portrayed in Sect. 7. The main objective of nanorobotics in medical fields is given in Sect. 8. Section 9 gives the design of nanorobots for medical applications. Next, the possible outcome of the research work undertaken is given in Sect. 10. Section 11 gives the applications of our research work undertaken followed by the basic constructional features of a Nanobot in Sect. 12. The mathematical modeling of a simple 3-D nanorobot is given in Sect. 13. The simulation tool used in the work and its overview are presented in Sect. 14. Finally, the conclusions are portrayed in Sect. 15 followed by an exhaustive number of references in the reference section.

17 Organization of the Research Article

The authors likes to thank the management of Acharya Institute of Technology—Bangalore, JSS Academy of Technical Education—Bangalore, Ibra College of Technology—OMAN & Dayananda Sagar College of Engg—Bangalore for having given all sorts of help related to the development of this research paper.

References

1. Cavalcanti A, Freitas RA Jr (2005) Nanorobotics control design: a collective behavior approach for medicine. IEEE Transactions on Nano BioScience 4(2):133–140
2. Hariharan R, Manohar J (2011) Nanorobotics as medicament (perfect solution for cancer). IEEE Emerging Trends in Robotics and Communication Technologies (INTERACT). ISBN: 978-1-4244–9006-6, January 2011
3. Cavalcanti A, Rosen L, Kretly LC, Rosenfeld M, Einav S (2004) Nanorobotics challenges in biomedical applications, design and control. IEEE ICECS International Conference on Electronics, Circuits and Systems, Tel-Aviv, Israel, pp 447–450, December 2004
4. https://www.nanorobotdesign.com
5. https://www.nanomedicine.com
6. https://www.www.foresight.org
7. Couvreur P, Vauthier C (2006) Nanotechnology: intelligent design to treat complex disease. Pharm Res 23(7):1417–1450
8. Montemagno C, Bachand G (1999) Constructing nanomechanical devices powered by biomolecular motors. Nanotechnology. 10(3):225–231
9. Jin S, Ye K (2007) Nanoparticle-mediated drug delivery and gene therapy. Biotechnol Prog 23(1):32–41

Switching Mechanism of Internal Model Control-based PI Controller for Lag Dominating Processes

Ujjwal Manikya Nath, Chanchal Dey, and Rajani K. Mudi

Abstract A good numbers of industrial processes are found to be lag dominating in nature. To obtain desirable closed-loop response for such processes, internal model control (IMC) technique with switching behavior is reported here. Initial setting of the reported PI controller is obtained through SIMC technique [1]. Here, the proposed scheme is based on switching between *smooth control* and *tight control* methodologies of IMC tuning depending on the instantaneous process operating conditions. Superiority of the proposed technique is verified through simulation study on lag dominating process models with the help of MATLAB/SIMULINK toolbox.

Keywords Internal model control · PI controller · Smooth control · Tight control · Switching mechanism · Lag dominating process

1 Introduction

Presently, a number of industrial control loops employ internal model control (IMC) technique [2–5] based proportional-integral (PI)/ proportional-integral-derivative (PID) controllers due to their simple tuning guideline. Primarily, Rivera et al. [6] proposed IMC based controller designing scheme for first-order processes,

U. M. Nath (✉)
Depatment of Instrumentation Engineering, Jorhat Engineering College, Garmur, Jorhat, Assam 785007, India
e-mail: um.nath@yahoo.com

C. Dey
Instrumentation Engineering Section, Department of Applied Physics, University of Calcutta, Kolkata, West Bengal 700009, India
e-mail: cdaphy@caluniv.ac.in

R. K. Mudi
Department of Instrumentation and Electronics Engineering, Jadavpur University, Salt Lake Campus, Kolkata, West Bengal 700098, India
e-mail: rajanikanta.mudi@jadavpuruniversity.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_32

subsequently Chien [7], Morari and Zafiriou [8] extended the IMC strategy for other process models. During closed-loop operation, IMC controllers provide good set-point tracking, but load rejection response is quite sluggish for processes with lag dominating nature. To enhance their load rejection performance, various modifications have been reported in the literature [1, 9, 10].

To ensure acceptable disturbance rejection, researchers reported the *tight control* mode for IMC controller to ensure higher speed of response. However, in practice, smooth process response is relatively preferred during set-point tracking [11]. Hence, to obtain improved overall performance for IMC controller during closed-loop operation, smooth set-point tracking along with faster load recovery is preferred. For achieving such dynamic behavior, incorporating an adaptive IMC [12, 13] or auto-tuning feature [14–16] or switching scheme [17–19] with a conventional IMC controller may be considered.

Here, in the proposed work, switching behavior is introduced in IMC structure where initial setting of the controller parameters is computed through SIMC [1] tuning. Here, in the reported scheme, controller setting switches between *tight control* and *smooth control* depending on the current process operating conditions. To justify the effectiveness of our proposed scheme, performance evaluation is made through simulation study on a number of lag dominating process models reported in the leading literatures. For quantitative assessment, performance indices—percentage overshoot (%OS), integral absolute error (IAE), integral time absolute error (ITAE), and total variation in control action (TV)—are calculated separately during set-point tracking and load recovery phases.

2 Internal Model Control

Figure 1 shows the block diagram of an internal model control (IMC) scheme, and its equivalent feedback structure is shown in Fig. 2. From the inner loop of Fig. 2 we obtain the equivalent controller expression as given by

$$G_c(s) = \frac{q(s)}{1 - \tilde{G}_p(s)q(s)} \tag{1}$$

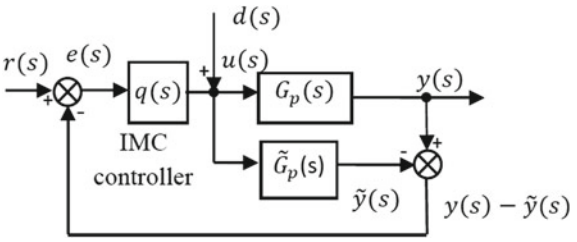


Fig. 1 Basic block diagram of internal model control

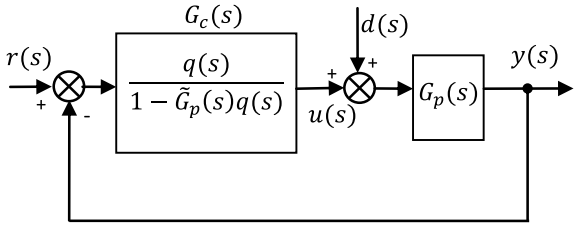


Fig. 2 Equivalent block diagram of IMC control scheme

Now, considering a lag dominating (i.e.,$\tau_P \gg \theta$) first-order plus dead time (FOPDT) process $\tilde{G}_p(s) = \frac{k_p}{\tau_p s+1} e^{-\theta s}$ where k_p is the open-loop gain, τ_p is the time constant, and θ is the dead time. Hence, IMC-PID controller can be formulated for the said lag dominating process by the following steps:

Step I: Dead time (θ) is factorized by first-order *Pade's approximation* $\tilde{G}_p(s) = \frac{k_p}{\tau_p s+1} \frac{-0.5\theta s+1}{0.5\theta s+1}$.

Step II: FOPDT model is factorized into two parts—inverting $\left(\tilde{G}_{p+}(s)\right)$ and non-inverting $\left(\tilde{G}_{p-}(s)\right)$, i.e.,

$$\tilde{G}_p(s) = \tilde{G}_{p+}(s)\tilde{G}_{p-}(s) = \frac{k_p}{(\tau_p s+1)(0.5\theta s+1)}(-0.5\theta s+1).$$

Step III: A first-order filter $f(s) = \frac{1}{\lambda s+1}$ is cascaded with the inverted part $\left(\tilde{G}_{p+}(s)\right)$ as obtained from *Step III*.

$$q(s) = \frac{(\tau_p s+1)(0.5\theta s+1)}{k_p} \frac{1}{\lambda s+1}.$$

Step IV: Substituting the relations obtained from *Steps I-IV* in Eq. (1) and comparing with the expression of non-interacting PID controller, $\left(k_c\left[\frac{\tau_I \tau_D s^2+\tau_I s+1}{\tau_I s}\right]\right)$, we can obtain the tuning parameters (k_c: proportional gain, τ_I: integral action time and τ_D: derivative time).

$$k_c = \frac{\tau_p + 0.5\theta}{k_p(\lambda + 0.5\theta)}, \tau_I = \tau_p + 0.5\theta, \tau_D = \frac{\tau_{p\theta}}{2\tau_p + \theta} \tag{2}$$

Equation (2) represents the conventional IMC-PID settings for FOPDT processes. With these settings (Eq. (2)), IMC based controller is capable to provide acceptable set-point tracking, but its load recovery is not satisfactory due to larger values of τ_I especially for lag dominating processes. From the literature survey [5–7, 9] it is found that that there are two alternatives to improve load rejection behavior, either the controller is to be tuned with a smaller value of integral time [1] or it should have a larger value of proportional gain [3].

As discussion is made in [18], larger value for λ and τ_I is desirable for smooth set-point tracking (*mode-I*) with lesser oscillation. On the contrary, during load change, smaller value for λ and τ_I is preferred for faster load recovery (*mode-II*). Hence, to ensure improved closed-loop responses during both set-point tracking and load rejection phases, switching between *mode-I* (*smooth control*) and *mode-II* (*tight control*) will be advantageous based on the instantaneous process operating conditions for providing appropriate values of λ and τ_I.

3 Proposed Switching Mechanism

Flowchart of the proposed online tuning scheme is shown in Fig. 3. Switching between *smooth control* and *tight control* mechanism with step excitation and pulsed nature load variation during closed-loop operation of a typical under-damped second-order process is depicted in Fig. 4. In the following section, detail about the operational scheme will be discussed with relevant timing diagram related to online switching between *smooth control* and *tight control*.

The switching mode control logic is designed based on two major signals, i.e., *state signal* and *triggering signal*. The *state* signal is derived based on the following logic:

'*If both the* $|e_N(k)|$ *and* $|\Delta e_N(k)|$ *are low* (L) *, then output is low* (L) *, otherwise output is high* (H)'.

Similarly, *triggering signal* is derived by the following logic:

'If output of state signal and $|\Delta r_N(k-1)|$ *are same* (*, i.e., low or high*) *,then output is high* (H) *, otherwise output is low* (L)'.

Now, the output of *triggering signal* switches from *low* (L) to *high* (H), and at the same time, output of *state signal* is *high* (H), then *mode-I* will be selected (*mode-I* corresponds to *smooth control* with higher values of τ_I and λ). On the contrary, during triggering if output of *state* is *low* (L), then *mode-II* will be selected (*mode-II* corresponds to *tight control* with smaller values of τ_I and λ). Selection of *mode-I* and *mode-II* is also separately shown by enlargement of the relevant portions of Fig. 4. Guidelines related to the selection of *mode-I, mode-II* and threshold values (Th_e, $Th_{\Delta e}$ and $Th_{\Delta r}$) are considered from [18].

4 4. Simulation Study

Efficacy of our proposed online switching mechanism for IMC controller is justified with various process models—first-order plus dead time (FOPDT), second-order plus dead time (SOPDT) and integrating plus dead time (IPDT) process. Here, it is to mention that though IMC controllers are realized based on the reduced FOPDT models, but its closed-loop performance is tested on actual higher-order model. For quantitative estimation, integral performance indices—%OS, IAE, ITAE, and total

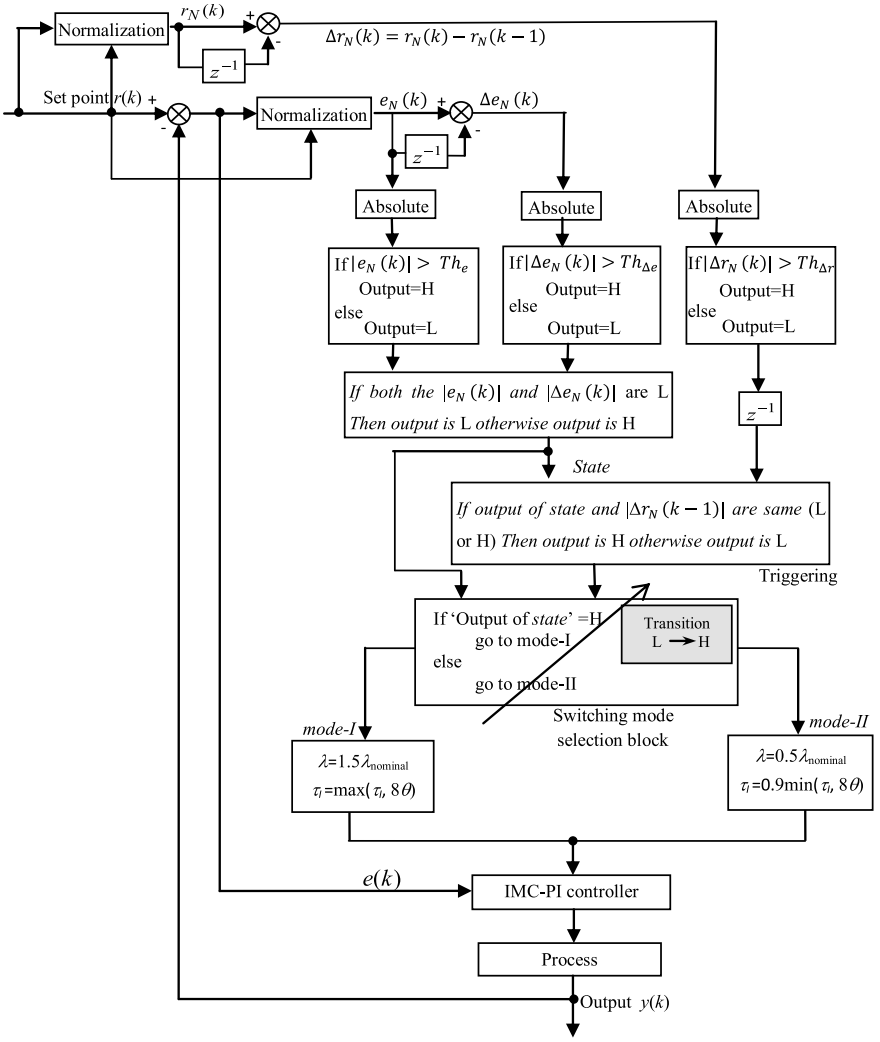


Fig. 3 Flowchart for online switching of IMC-PI controller

variation (TV) in control action—are computed. For simulation study, the following three process models are chosen.

Model I; $\tilde{G}_p(s) = \frac{3}{100s+1}e^{-10s}$ [3]

Model II: $\tilde{G}_p(s) = \frac{1}{(s+1)(0.2s+1)}$ [1]

Model III: $\tilde{G}_p(s) = \frac{1}{s}e^{-s}$ [20]

By incorporating our proposed switching method, we obtain separate set of values for λ and τ_I during set-point tracking, i.e., to ensure *smooth control* (*mode-I*) and during load recovery phase for *tight control* (*mode-II*). All the adjustable tuning

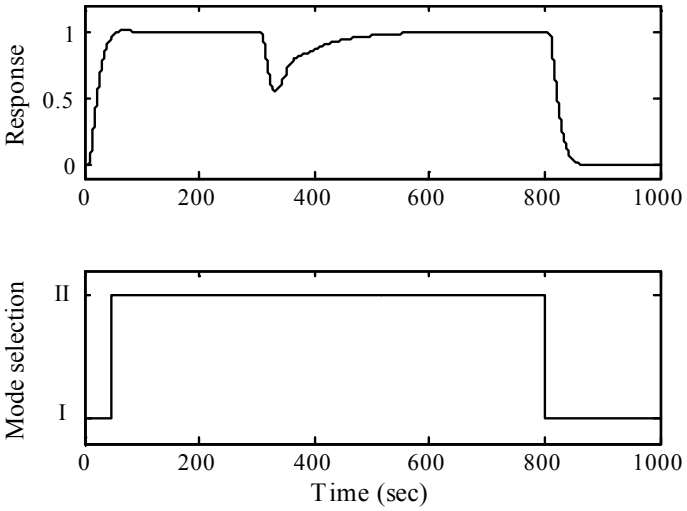


Fig. 4 Switching mechanism for a typical lag dominating process during set-point changes and load disturbance

parameters (λ and τ_I) for Model-I, II, and III are listed in Table 1, and corresponding performance indices are provided in Table 2. The closed-loop process responses for Model I-III are shown in Figs. 5, 6 and 7, respectively. From the performance indices (as listed in Table 2) as well as Figs. 5, 6 and 7, improved closed-loop responses are found by employing the reported online switching scheme in comparison with fixed SIMC tuning [1].

Table 1 Tuning parameters of controllers

Model	Controller	Operating condition	Controller parameters	
			λ	τ_I
I	SIMC [1]	Overall	10	80
	Proposed	Set-point tracking	15	100
		Load recovery	0.5	72
II	SIMC [1]	Overall	0.1	0.8
	Proposed	Set-point tracking	0.15	0.72
		Load recovery	0.05	1.1
III	SIMC [1]	Overall	1	8
	Proposed	Set-point tracking	1.5	100
		Load recovery	0.5	7.2

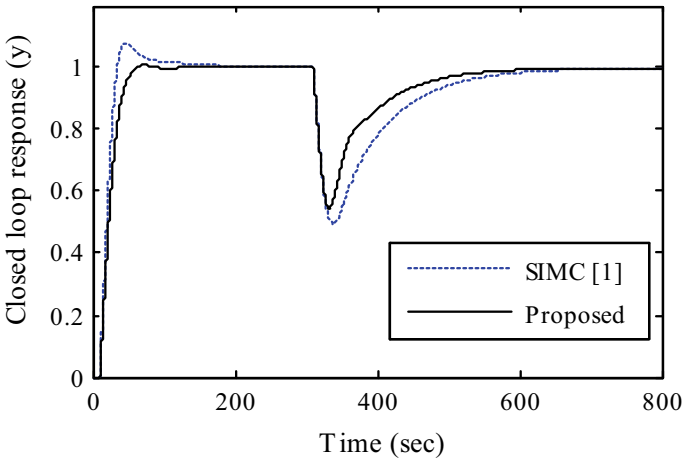


Fig. 5 Closed-loop response of Model I

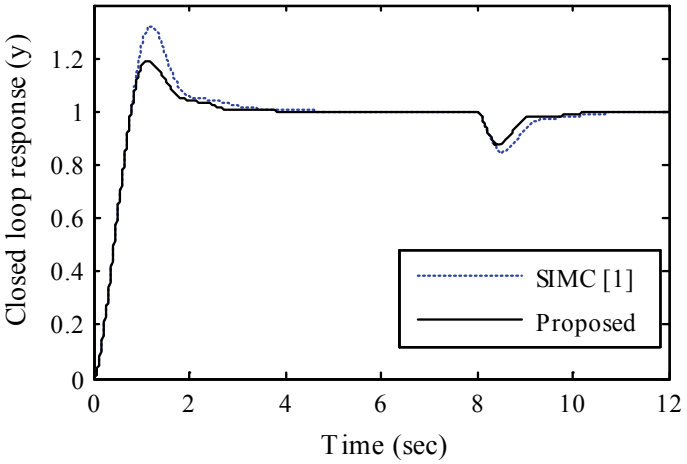


Fig. 6 Closed-loop response of Model II

5 Conclusion

In this work, an online switching for IMC controller is reported for lag dominating processes. Instead of considering fixed values of tuning parameters for IMC controllers, i.e. static closed-loop time constant (λ) and integral time (τ_I), dynamic nature is incorporated in selecting their values depending on the current process operating conditions. As a result, an overall performance enhancement during closed-loop response is observed by incorporating our proposed online switching scheme with the SIMC settings compared to fixed SIMC settings. In future, there is a scope

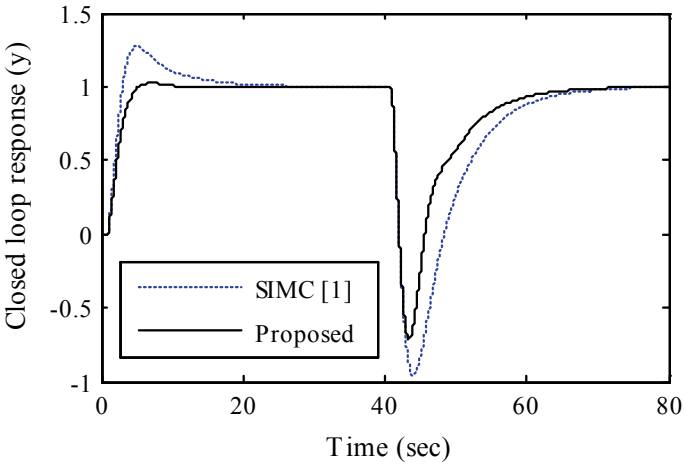


Fig. 7 Closed-loop response of Model III

Table 2 Performance indices of controllers

Model	Controller	Set point				Load rejection		
		%OS	IAE	ITAE	TV	IAE	ITAE	TV
I	SIMC [1]	8.1	24.88	645.3	3.53	47.84	1.89×10^4	1.49
	Proposed	0.8	25.17	391.7	4.50	32.38	1.23×10^4	1.85
II	SIMC [1]	32.1	0.715	0.574	4.89	0.144	128.9	2.55
	Proposed	19.2	0.598	0.371	4.40	0.097	0.861	2.79
III	SIMC [1]	27.6	3.919	20.03	1.218	15.98	766.20	1.55
	Proposed	1.8	2.563	4.826	1.36	10.79	508.80	1.844

for performance evaluation of the proposed technique in comparison with other existing well-known tuning relations through simulation study as well as real-time experimentation.

References

1. Skogestad S (2003) Simple analytic rules for model reduction and PID controller tuning. J Process Control 13(4):291–309
2. Nath UM, Dey C, Mudi RK (2019a) Designing of IMC-PID controller with second-order filtering for lag and delay dominating process based on suitable dead time approximation. Asia Pacific J Chem Eng. https://doi.org/10.1002/apj.2359
3. Ali A, Majhi S (2009) PI/PID controller design based on IMC and percentage overshoot specification to controller set point change. ISA Trans 48(1):10–15

4. Nath UM, Dey C, Mudi RK (2019b) "Designing of IMC-PID controller for higher-order process based on model reduction method and fractional coefficient filter with real-time verification. Chem Product Process Model. https://doi.org/10.1515/cppm-2019-0089
5. Seborg DE, Edgar TF, Mellichamp DA (2004) Process dynamics and control. Wiley, Singapore
6. Rivera DE, Skogested S, Morari M (1986) Internal model control. 4. PID controller design. Ind Eng Chem Process Design Develop 25(1):252–265
7. Chien IL (1988) IMC–PID controller design-an extension. IFAC Proc Series 6:147–152
8. Morari M, Zafiriou E (1989) Robust process control. Prentice- Hall, New Jersy
9. Shamsuzzoha M, Lee M (2007) IMC-PID controller design for imrproved disturbance rejection of time delayed processes. Ind Chem Eng Res 46(7):2077–2091
10. Shamsuzzoha M (2016) IMC based robust PID controller tuning for disturbance rejection. J Cent South Univ 23(3):581–597
11. Skogestad S (2006) Tuning for smooth PID control with acceptable disturbance rejection. Ind Eng Chem Res 45(23):7817–7822
12. Shafiq M (2005) Internal model control structure using adaptive inverse control strategy. ISA Tran 44(3):353–362
13. Rupp D, Guzzella L (2010) Adaptive internal model control with application to fueling control. Control Eng Practice 18(8):873–881
14. Nath UM, Dey C, Mudi RK (2019c) Fuzzy tuned model based control for level and temperature processes. Microsyst Technol 25(3):819–827
15. Nath UM, Datta S, Dey C (2015) Centralized auto-tuned IMC-PI controllers for industrial coupled tank process with stability analysis. In: IEEE 2nd international conference on recent trends in information systems (ReTIS) Kolkata, pp 296–301
16. Nath UM, Dey C, Mudi RK (2020a) Fuzzy rule-based auto-tuned internal model controller for real-time experimentation on temperature and level processes. Int J Auto Control 14(2):239–256
17. Liberzon D (2003) Switching in systems and control. Springer, USA, pp 3–15
18. Nath UM, Dey C, Mudi RK (2020b) A switching IMC-PID controller design for lag dominating processes with real-time implementation on a level control system. Mech Syst Control 48(3):171–182
19. Marchetti G, Barolo M, Jovanovic L, Zisser H, Seborg DE (2008) An improved PID switching control strategy for Type 1 diabetes. IEEE Trans Biomed Eng 55(3):857–865
20. Shamsuzzoha M, Skogestad S (2010) The setpoint overshoot method: a simple and fast closed-loop approach for PID tuning. J Process Control 20:1220–1234

Hybrid Model for Vendors Providing Materials to a Manufacturing Company

Prabjot Kaur

Abstract Selection of vendors in this paper is based on various multiple criteria like cost, quality, cycle time, service, and reputation. Due to multiple criteria the decision becomes vague and complex, so such types of situations are best handled by fuzzy sets. AI methods like artificial neural network (ANN) cope well with complexity and uncertainty as well. In this paper, various fuzzy AHP approaches like fuzzy extent analysis (FEA) and fuzzy geometric mean (FGM) method were used to find the priority weights [10] for the various vendor with respect to various criteria. From fuzzy AHP, the priority weights were passed through ANN for the selection of eligible vendors for the small firm. The advantage of combining the outputs of AHP with neural network model is that it refines, generalizes, and extracts information from data basis very efficiently and effectively. A numerical example explains the method adopted for vendor selection. Using performance measures like mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) showed that FEA with ANN gives better results than FGM with ANN.

1 Introduction

The vendor selection is a two phase process: "The first phase identifies the criteria for selection of vendor and secondly phase the methodology that uses these criteria for evaluation and ranking of the vendors" [18]. Dickson [17] laid the foundation of vendor selection problem, stating that 23 criteria were important for its selection. While Chai et al. [20] gave 26 methods divided into three categories called "multi-criteria decision-making (MCDM), mathematical programming (MP), and artificial intelligence (AI) methods" for selection of vendors. Neural networks are one such AI method. Aouadni et al. [22] classified the "MCDM approaches into" two categories called single-criterion synthesis which include methods TOPSIS, AHP, etc.,

P. Kaur (✉)
Department of Mathematics, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: prabjotkaur@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_33

and outranking synthesis approach. The advantage of AHP is its advantage of organizing the problem in a hierarchical structure and measuring consistency in decision maker's acumen. The common factor between fuzzy AHP and ANN is its ability to tackleuncertainty and complexity. In this paper, neural network is applied on fuzzy AHP methods to select the best vendor from a group of vendors.

The first study on the application of neural network model in vendor selection problem was proposed by Siying et al. [21]. Albino et al. [1], Nassimbeni and Battain [2], Vahdani et al. [3] used various neuro-fuzzy based approaches for supplier selection problems. Kumar and Roy [15] used an integrated model for vendor selection using neural network, Lakshmanpriya et al. [4] and Tang et al. [5] applied the analytic hierarchy process and neural network process to select vendor in supply chain management. Harikannan et al. [13] used TOPSIS with ANN method for supplier selection problem. Kar [16] in his study applied a hybrid approach for the supplier selection problem using fuzzy sets, analytic hierarchy process, and neural networks to provide group decision support under consensus achievement. Tavana et al. [19] in their approach formulated the ANN approaches like ANFIS and multi-layer perceptron (MLP) in a supplier selection framework with partial information.

In this paper, fuzzy AHP-ANN method is used as a vendor-rating system where the attributes weights are found by a series of pair-wise comparisons and then used to construct a vendor evaluation and selection system. Neural networks can extract decision-making capability from databases [6] and also a good adaptive system. If one combines the outputs of a neural network model with AHP, it improves accuracy. The application of the approaches explained are explained by a numerical example.

The organization of the paper consists of various sections. Sect. 1 states the introduction of the problem. In Sect. 2, the basic concepts of fuzzy sets are stated. Section 3 includes the discussions related to the methodology of fuzzy AHP and ANN to be used in the problem. Section 4 gives a numerical example for illustration. Section 5 consists of results and discussion, and finally, in the last section, we have conclusions of the results.

2 Methodology

2.1 Analytical Hierarchy Process (AHP)

The analytic hierarchy process (AHP) is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology. It was developed by Saaty [7] in the 1970s and has been extensively studied and refined since then. AHP is a process for developing a numerical score to rank each decision alternative based on how well the alternative meets the decision maker's criteria.

2.2 Fuzzy Analytical Hierarchy Process

Fuzzy AHP deals with uncertainty in judgments by expressing the comparison ratios as fuzzy sets or fuzzy numbers. In this paper, we use triangular fuzzy numbers to represent data sets, and methods of fuzzy geometric method [8] and fuzzy synthetic extent analysis method [9] were used for the decision makers' comparison judgment and evaluate the final priority of different criteria.

In our approach of fuzzy AHP if we use triangular fuzzy numbers, the fuzzy number $\boldsymbol{u_{ij}}$ is r represented as a triplet (l_{ij}, m_{ij}, u_{ij}) where $\mathbf{m_{ij}}$, $\mathbf{l_{ij}}$, and $\mathbf{u_{ij}}$ are the middle, the lower, and the upper values, respectively. In AHP a consistency index is used to measure any inconsistency in each comparison matrix and the entire hierarchy. Consistency ratio (**CR**) is required to be less than 0.10 for acceptable results.

2.3 Fuzzy Geometric Mean Method [10]

In fuzzy geometric mean method:

Step1: Let $\mathbf{A = [a_{uv}]_{kxk}}$ is fuzzy reciprocal matrix. Evaluate the geometric row mean

$$\mathbf{r_u} = (\mathbf{a_{u1}} * \mathbf{a_{u2}} * \ldots * \mathbf{a_{uk}})^{\mathbf{1/k}} \tag{1}$$

Step2: We normalized the geometric row mean given by

$$\mathbf{w_u} = \mathbf{r_u} * (\mathbf{r_1} + \mathbf{r_2} + \cdots + \mathbf{r_k})^{-\mathbf{1}} \tag{2}$$

Step3: Using Eqs. (1) and (2), calculate the weights of each criteria and subcriteria under its main criteria.

2.4 Fuzzy Extent Analysis [9]

Let us consider $\mathbf{X = \{x_1, x_2, \ldots, x_n\}}$ is an object set, and $\mathbf{U = \{u_1, u_2, \ldots, u_m\}}$ be a goal set. Therefore, according to the method for each object, set m extent analysis for each goal $\mathbf{g_i}$ is performed. Therefore, $\mathbf{m}$ extent analysis values for each object can be obtained as follows:

Step 1: We calculate the value of fuzzy synthetic extent with respect to the ith object as

$$S_i = \sum_{j=1}^{m} M_{g_i}^{j} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^{j} \right]^{-1} \tag{3}$$

$$\text{where } \sum_{j=1}^{m} M_{g_i}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j \right) \tag{4}$$

The inverse of the vector is computed by using Eq. (3) as

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{g_i}^{j} \right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i} \right) \tag{5}$$

Step 2: The degree of possibility of $\mathbf{M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)}$ is defined as

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2}(d) = \begin{cases} 1, \text{ if } m_2 \geq m_1 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, \text{ if } l_1 \geq u_2 \\ 0, \text{ otherwise} \end{cases} \tag{6}$$

where d is the ordinate of the highest intersection point D between μ_{M_1} and μ_{M_2}.To compare $\mathbf{M_1}$ and $\mathbf{M_2}$, we need both the values of $\mathbf{V(M_1 \geq M_2)}$ and $\mathbf{V(M2 \geq M_1)}$.

Step 3: The degree of possibility for a convex fuzzy number to be greater than k convex fuzzy numbers $\mathbf{M_i (i = 1, 2, \ldots, k)}$ can be defined by

$$V(M \geq M_1, M_2, \ldots, M_k))] = \min V(M \geq M_i), \; i = 1, 2, 3, \ldots, k. \quad \text{Assume that } d(A_i) = \min V(S_i \geq S_k) \tag{7}$$

For $\mathbf{k = 1, 2, .., n; k \neq i}$. Then, the weight vector is given by

$$\mathbf{W} = (\mathbf{d}(\mathbf{A_1}), \mathbf{d}(\mathbf{A_2}), \ldots, \mathbf{d}(\mathbf{A_n}))^{\mathbf{T}} \tag{8}$$

where $\mathbf{A_i (i = 1,2, \ldots, n)}$ are n elements**Step 4**: We obtain the normalized weight vectors as

$$\mathbf{W} = (\mathbf{di}(\mathbf{A_1}), \mathbf{di}(\mathbf{A_2}), \ldots, \mathbf{dA_n})^{\mathbf{T}} \tag{9}$$

where $\mathbf{W}$ is a deterministic value.

2.5 Artificial Neural Network (ANN)

Artificial neural networks (**ANN**) or **connectionist systems** [14] are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules. The implemented model performs tasks faster than traditional system. Basically, ANN has three types of learning: learning with a teacher, learning without a teacher, and reinforcement learning. This paper is based on unsupervised learning that is learning without a teacher. In unsupervised learning, desired outputs of the network are not available.

The neural network model is as:-

From the figure, it is clear that there is no feedback loop from the environment to inform what the outputs should be or whether the outputs can be correct. The network must itself discover patterns, regularities, features from the input data, and relations for the input data over the outputs.

The ANN model comprises of one input layer, one hidden layers, and one output layer. The inputs with their weights, pass through the hidden layer for processing. By using the sigmoid function to the given data, the outputs of the hidden layer are found. These outputs finalize the output layer. The highest weight is considered as the best vendor for the selection process.

3 The Performance Measure for the Fuzzy AHP Methods

Evaluating the performance of imodels by using mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) between the experiential and predicted values of ranking the vendors:

1. **Mean square error(MSE)** $= \frac{1}{n}\sum_{1}^{n}(t_i - o_i)^2$
2. **iRMSE** $= \sqrt{\sum \frac{(t_i - o_i)^2}{n}}$
3. **iMAE** $= \frac{1}{n}\sum|(t_i - o_i)|$

4 Numerical Example [10]

This example is taken from a paper Kaur et al. [10] where a manufacturing company wants to select suitable materials vendor to provide materials for manufacturing of a new product. Three competing vendors *A*, *B*, and *C* were accepted for evaluation after initial screening. The appraisal is based on five criteria (Table 1): (1) cost (C), (2) quality (Q), (3) service (S), (4) cycle time (CT), and (5) reputation (R).

The steps of selection for a final vendor are given in the algorithm below.

Table 1 Data for the pair-wise comparison between the criteria [10]

Goal	C	Q	S	CT	R
C	1	1	1	4	1
Q	1	1	2	4	2
S	0.25	0.25	0.2	1	0.33
CT	1	0.5	1	5	3
R	1	1	0.33	3	1

4.1 Algorithm for Evaluation of Vendors

1. The vendor selection problem has been organized in a hierarchy. Tables 2, 3, and 4 taken from [10] represent evaluation of the vendors by using crisp and fuzzy data.
2. From the fuzzy decision matrix, fuzzy weight of each criterion was determined.

Table 2 Fuzzy data for pair-wise comparison between the main criteria [10]

Goal	C	Q	S	CT	R
C	(0 ,1, 2)	(0, 1, 2)	(0, 1, 2)	(3, 4, 5)	(0, 1, 2)
Q	(0, 1, 2)	(0, 1, 2)	(1, 2, 3)	(3, 4, 5)	(1, 2, 3)
S	(0, 1, 2)	(0.33, 0.5, 1)	(0, 1, 2)	(4, 5, 6)	(2, 3, 4)
CT	(0.2, 0.25, 0.33)	(0.2, 0.25, 0.33)	(0.167, 0.2, 0.25)	(0, 1, 2)	(0.25, 0.33, 0.5)
R	(0, 1, 2)	(0, 1, 2)	(0.25, 0.33, 0.5)	(2, 3, 4)	(0, 1, 2)

Table 3 Fuzzy pair-wise comparison for the various vendor w.r.to cost criteria [10]

Quality	Vendor A	Vendor B	Vendor C
Vendor A	(0, 1, 2)	(0.2, 0.25, 0.33)	(0.3, 0.5, 1)
Vendor B	(3, 4, 5)	(0, 1, 2)	(2, 3, 4)
Vendor C	(1, 2, 3)	(0.25, 0.33, 0.5)	(0, 1, 2)

Table 4 Global weights from fuzzy geometric mean method [10]

	Criteria weights	Vendor A	Vendor B	Vendor C	CR
Cost (C)	0.2321	0.1435	0.6134	0.2441	0.08
Quality (Q)	0.2904	0.0948	0.3447	0.5613	0.04
Service (S)	0.0531	0.2834	0.6453	0.0711	0.008
Cycle time (CT)	0.2525	0.3076	0.2270	0.4643	0.08
Reputation i(R)	0.1721	0.4746	0.4746	0.06298	0.006
	Global weights	0.2352	0.5986	0.5244	

3.The geometric mean method and fuzzy extent analysis calculated the local and global priorities in Tables 5, 6, and 7 using Eqs. (1)–(9), respectively. The calculated results are taken from [10].
4. Then, ANN as shown in Fig. 1 is used to predict the selection of an appropriate vendor.
5. The calculations of ANN [4, 11] are as follows: Firstly, the output values of hidden layer is calculated. Take extracted criteria weights from geometric mean and fuzzy extent analysis methods.
6. Compute the input value for input layer
$X_i = 1$/number of vendors.
7. Compute

Table 5 Global weights from fuzzy extent analysis method [10]

	Criteria weights	Vendor A	Vendor B	Vendor C	CR
Cost (C)	0.9037	0.3321	1	0.6904	0.08
Quality (Q)	0.9822	0.1311	0.8052	1	0.04
Service (S)	0.226	0.5873	1	0	0.008
Cycle time (CT)	1	0.3076	0.7967	1	0.08
Reputation (R)	0.8132	1	1	0	0.006
	Global weights	2.334	3.530	2.533	

Table 6 Output values for hidden layer using fuzzy geometric mean method

Criteria	Criteria weights	Input value X_i	$\sum X_i W_{ci}$	Output layer for hidden layer Y_{ci}
Quality	0.3221	0.333	0.4318679	0.606319617
On-time delivery (OTD)	0.2904	0.333	0.2967032	0.573636389
Systematic billing (SB)	0.0531	0.333	0.2530469	0.562926304
Safety (S)	0.2525	0.333	0.4522475	0.611173464
Job knowledge (JK)	0.1721	0.333	0.3719279	0.591924744

Table 7 Results for output layer using fuzzy geometric mean

Vendor	$Y_{c0} =$ 0.60632	$Y_{c1} =$ 0.5736	$Y_{c2} =$ 0.5629	$Y_{c3} =$ 0.6112	$Y_{c4} =$ 0.5919	$\sum Y_{ci} W_{ci}$	Y_{vi}
Vendor A	0.1435	0.0948	0.2834	0.3078	0.4746	0.96984513	0.7251
Vendor B	0.6134	0.3447	0.6453	0.2270	0.4746	1.55256915	0.8253
Vendor C	0.2441	0.5613	0.0711	0.4643	0.06298	1.12967242	0.7558

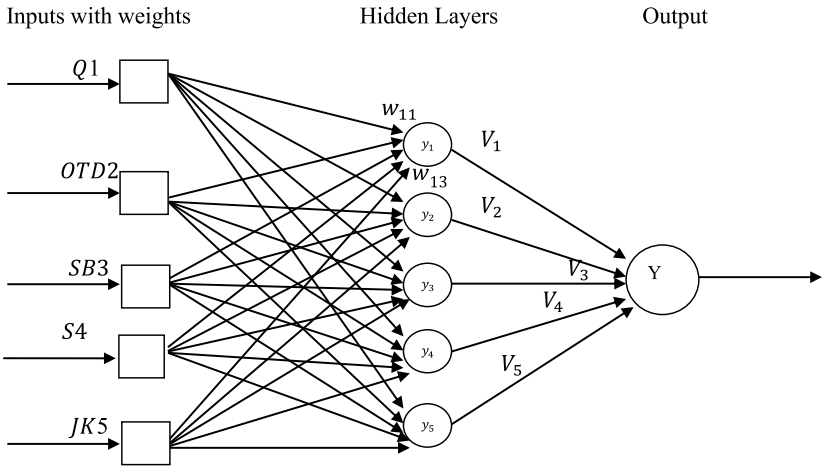


Fig. 1 A neural network model for VSP

$$\sum X_i W_{ci} = \sum_{i=0}^{n} \{\text{input value } (X_i) * \text{Weight on } i\text{th criteria}\} + \{\text{input value for all bias neuron} + \text{Weight for all bias neuron}\}$$

8. Calculate the output value for hidden layer.
Input value for output layer $(Y_{cii}) = 1/(1 + e^{-\alpha(\sum X_i W_{ci})})$
9. For matrix of output layer, write Y_{ci}, i.e., the computed value with respect to vendors
computed weights in matrix form.
10. Compute

$$\sum Y_{ci} W_{vi} = \sum_{i=0}^{n} \{i\text{th vendor weight} * i\text{th computed } Y_{ci}\} + \{\text{input value for all bias neuron} * \text{weight for all bias neuron}\}$$

11. Compute Y_{vi}, i.e., value for the output layer $= 1/(1 + e^{-\alpha(\sum Y_{ci} W_{vi})})$
12. Arrange the vendors according to Y_{vi} values calculated. The vendor with the largest value
is selected.
The above calculations of the algorithm are shown in Tables 4, 5, 6, 7, 8, 9, and 10 below.

Assume the intake for each of the bias neuron $= 1$
Assume the weight for each of the bias neuron $= 0.2 X_i =$ Intake value for input layer $= 1/3 = 0.333$, $\alpha = 1 W_{ci} =$ Weight of criteria $Y_{ci} =$ Output value for hidden layer $= 1/(1 + e^{-\alpha(\sum X_i W_{ci})})$ The result is: $V_2 > V_3 > V_1$ The result is: $V_2 > V_3 > V^1$

Table 8 Value for output values for hidden layer using fuzzy extent analysis method

Criteria	Criteria weights	Input value X_i	$\sum X_i W_{ci}$	Output layer for hidden layer Y_{ci}
Quality	0.9037	0.333	1.1027	0.7507
Systematic billing (SB)	0.9822	0.333	1.1812	0.7651
Service	0.2260	0.333	0.4257	0.6049
Cycle time	1	0.333	1.1999	0.7685
Reputation	0.8132	0.333	1.0123	0.7335

Table 9 Value for output layer using fuzzy extent analysis method

Vendor	$Y_{c0} =$ 0.7507	$Y_{c1} =$ 0.7651	$Y_{c2} =$ 0.6049	$Y_{c3} =$ 0.7685	$Y_{c4} =$ 0.7335	$\sum Y_{ci} W_{ci}$	Y_{vi}
Vendor A	0.3321	0.1311	0.5873	0.9588	1	2.3752	0.9149
Vendor B	1	0.8052	1	0.7967	1	3.5175	0.9711
Vendor C	0.6094	1	0	1	1	2.7214	0.9384

Table 10 Comparing the results of the two methods

	Fuzzy geometric mean (FGM)	Fuzzy extent analysis (FEA)
Vendor A	0.7251	0.9149
Vendor B	0.8253	0.9711
Vendor C	0.7558	0.9384
Rank	V2 > V3 > V1	V2 > V3 > V1

Note: we resemble *A*, *B*, *C* as 1, 2, and 3, respectively, in vendor.

4.2 Results and Discussion

We have considered three vendors with five criteria. Fuzzy AHP and ANN model have been proposed for the selection of a vendor. In this process, the weights [10] are calculated from fuzzy geometric mean method and fuzzy extent analysis and are fed into neural networks for selection of best vendor. The input is assumed to be the same for all neurons and depends on the number of vendors as shown in calculations above. From Table 9 calculations, we observe that vendor 2 has the largest value of 8253 in comparison with other vendors. We observe that the reputation and service (Table 2) of vendor 2 are excellent compared to vendor 3 and vendor 1. So, vendor 2 is best for selection based on the evaluation criteria. In any inevitable circumstances, the immediate vendors will be best selection according to that organization. Table 11 shows the performance accuracy measure MSE, RMSE,

Table 11 Performance measure of the two methods

	Fuzzy geometric mean method	Fuzzy extent analysis method
MSE	0.0189	0.002251
RMSE	0.1378	0.0474
MAE	0.1312	0.0415

and MAE for the two methods. We observe that fuzzy extent analysis gives better performance than fuzzy geometric mean method.

5 Conclusions

A hybrid model of vendor selection problem (VSP) with applications of fuzzy AHP and ANN was presented. The model extracts and exploits the advantages of the two methods. Fuzzy AHP is used to tackle uncertainty and imprecision in data based on approximations or pair-wise comparison, whereas ANN has good adaptive systems with tolerance of imprecise data and model-free structure.

In future scope, other fuzzy AHP methods can be used for comparative studies for a wider perspective of the problem. Neural network has a structure which is flexible enough to represent and integrate information. Thus, other learning rules can also be implemented in the model.

References

1. Albino, V., Garavelli, A.C., Gorgoglione, M.: Fuzzy logic in vendor rating: a comparison between a fuzzy logic system and a neural network. Fuzzy Econ. Rev. **3**(2), 25–47 (1998)
2. Nassimbeni, G., Battain, F.: Evaluation of supplier contribution to product development: Fuzzy and neuro-fuzzy based approaches. Int. J. Prod. Res. **41**(13), 2933–2956 (2003)
3. Vahdani, B., Iranmanesh, S.H., Mousavi, S.M., Abdollahzade, M.: A locally linear neuro-fuzzy model for supplier selection in cosmetics industry. Appl. Mathe. Model. **36**(10), 4714–4727 (2012)
4. Lakshmanpriya, C., Sangeetha, N., Lavanpriya, C.: Vendor selection in manufacturing industry using AHP and ANN. SIJ Trans Indust Financ Business Manag **1**(1), 29–34 (2013)
5. Tang, S.H., Hakim, N., Khaksar, W., Ariffin, M. K.A., Sulaiman, S., Pah, P.S.: A hybrid method using analytic hierarchical process and artificial neural network for supplier selection. Int. J. Inn. Manage. Technol. **4**(1), 109–111 (2013)
6. Wei, S., Zhang, J., Li, Z.: A supplier-selecting system using a neural network. In: IEEE International Conference on Intelligent Processing Systems (Cat. No. 97TH8335), vol. 1, pp. 468–471. IEEE (1997)
7. Saaty, T.L.: The Analytic Hierarchy Process. Mc Graw Hill, New York, (1980)
8. Buckley, J.J.: Fuzzy hierarchical analysis Fuzzy Sets Syst. **17**, 233–247 (1982)
9. Chang, D.T.: Applications of the extent analysis method on fuzzy AHP. European J. Operational Res. **95**(3), 649–655 (1996)

10. Kaur, P., Verma, R., Mahanti, N.C.: Selection of vendor using analytical hierarchy process based on fuzzy preference programming. Opsearch **47**(1), 16–34 (2010)
11. Kumar, J., Roy, N.: A hybrid method for vendor selection using neural network. Int. J. Comput. Appl. **11**(12), 35–40 (2010)
12. Sivanandam, S.N., Deepa, S.N.: Principle of Soft Computing. Wiley (2014)
13. Harikannan, N., Vijay, J., Mahalakshmi, N.: Decision making model for supplier evaluation and selection using MCDM methods. Bonfring Int. J. Ind. Eng. Manage. Sci. **4**(2), 76–82 (2014)
14. Simon, H.: Neural Networks: Comprehensive Foundations. Pearson Education (1997)
15. kumar, J., Roy, N.: A hybrid method for vendor selection using neural network. Int. J. Comput. Appl. **11**(12), 0975–8887 (2010)
16. Kar, A.K.: A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. J. Comput. Sci. (6), 23–33 (2015)
17. Dickson, G.W.: An analysis of vendor selection systems and decisions. J. of Purchasing **2**(1), 5–17 (1966)
18. Ordoobadi, S.M.: Development of a supplier selection model using fuzzy logic. In: Supply chain management. Int. J. **14**(4), 314–327 (2009)
19. Tavana, M., Fallahpour, A., Caprio, D.D., Santos-Arteaga, F.J.: A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. Expert. Syst. Appl. 61, 129–144 (2016)
20. Chai, J., Liu, J.N.K., Ngai, E.W.T. Application of decision-making techniques in supplier selection: A systematic review of literature. Expert. Syst. Appl. 40, 3872–3885b (2013)
21. Siying, W., Jinlong, Z., Zhicheng, l., A supplier selecting system using a neural network. In: IEEE Int. Conf. Intell. Process. syst., 28–31 (1997), Beijing, China
22. Aouadni, S., Aouadni, I., Rebaï, A. (2019) A systematic review on supplier selection and order allocation problems, J. Ind. Eng. Int. 15, 267–289 (1997)

Solar Powered Irrigation System

Russell Pinto, Carol Mathias, Nidhi Kokande, Mathew Thomas, and U. S. Pushpas

Abstract Agriculture and farming are not saturated to just farmers as there are variety of crops and plants that are produced in a large scale to fulfill the requirement of day-by-day growing population as there are various varieties of plants they all require specific environment to grow, and so, they all need proper fertilizers and water content. Farming is usually done by using manual methods that sometimes causes excess use of water or fertilizers which in turn results in wastage of water even sometimes if water is not supplied properly crops do not tend to grow at their full extend. Usually, pumps are used for irrigation, which requires ample amount of energy for large fields, and usually, diesel or electricity is used to run those pumps. As these manual methods consume a lot amount of energy and lead to problems like shortage of water and flooding some times because of manual controlling, we have chosen to select automatic controlling of irrigation powered by solar energy. Solar energy is a renewable source of energy, which is available to us and can be converted into usable form by using photovoltaic cells, so we have incorporated solar cells to save energy in our irrigation system.

Keywords Irrigation solar energy · Solar water pumping · Drip irrigation solar power · Motor

R. Pinto · C. Mathias · N. Kokande · M. Thomas · U. S. Pushpas (✉)
Department of Electronics and Telecommunication, Fr. Conceicao Rodrigues Institute of Technology, Vashi, Navi Mumbai, Maharashtra, India
e-mail: pushpaus@gmail.com

R. Pinto
e-mail: russellpinto996@gmail.com

C. Mathias
e-mail: Carolmathias06@gmail.com

N. Kokande
e-mail: nkokande@gmail.com

M. Thomas
e-mail: mathewjerrythomas@outlook.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_34

1 Introduction

The project took into consideration all the factors that are important and lead to less power consuming with higher efficiency of crop production that are necessary to fulfill the growing demand of food day by day. Energy efficient wireless sensor network provides better solution in this case [1–6]. The system works automatically as there is an Arduino controller to run the system. Multiple sensors are incorporated in the system to monitor the condition of soil and take necessary steps to fulfill the requirement.

We have incorporated GSM/GPRS module so that our system can be monitored remotely, so if the farmer is away from the field, he will be getting proper feedback of the system. There are many countries that do not receive proper rain, so they are unable to water their crops properly, but if they are having an automatic irrigation system which sense the requirement of the water in the soil, so they can save the excess amount of water that was wasted earlier in the times of water deficiency [7, 8].

As our system is powered by solar, there is no need to use diesel generators to power the system which in turn saves money and motivates the farmer [9, 10].

A. **Irrigation**: Irrigation system is considered as a framework that disseminates water to target territory. Irrigation system is intended for horticultural purposes. Productivity of the water system lies on the framework utilized. Since the prior occasions, human life is governed by agribusiness, and the Irrigation system is one and only of the apparatuses that expand farming. There are various dissimilar sorts of water system framework all over throughout the world, but these water systems are confronting several issues [11].
B. **Drip Irrigation**: The drip irrigation system is otherwise called trickle irrigation water system water is applied in the drops legitimately close to the base of the plant. Water is passed on through an arrangement of adaptable funnel lines, working at low weight, and is applied through plants through trickle spouts [12].
C. **Solar Energy**: Solar energy, brilliant light and warmth from the sun, Solar energy is the most limitless wellspring of energy on the planet. Sun-based power is not only an answer for the current energy crisis yet also an ecofriendly kind of energy. Photovoltaic age is a monetary technique for abuse of sun-based energy. Sun-based sheets are nowadays comprehensively used for running streetlights, for controlling water radiators and to meet neighborhood loads [13, 14]. The cost of sun situated sheets has been never-ending lessening that engage sits use in different divisions. One in all the uses of this development is used in water system frameworks for developing. Daylight-based energized water framework structure will be an appropriate alternative for farmers in the ebb and flow circumstance with energy crisis in India. This a green procedure for energy creation that gives free energy once a basic hypothesis is molded [15–20].

2 Literature Survey

With the ceaseless increment in petroleum product cost and decrease in top watt cost of sun-oriented cells because of large-scale manufacturing, the photovoltaic force is to turn out to be further efficient in future [21–25]. Studies have been completed on execution assessment, streamlining, measuring methods, productivity improvement and variables influencing framework execution, financial and natural parts of PV siphoning frameworks. The features of the examination examinations are introduced in this segment [26–34].

3 Block Diagram

Resulting are the main mechanisms used from which solar powered automatic irrigation system by means of Arduino and GSM has been fabricated (Figs. 1, 2, 3, 4, 5, 6 and 7; Table 1).

- Arduino UNO
- GSM SIM module
- Soil moisture sensor
- 12v relay
- BC 547 transistor
- Voltage Regulator-7805 and 7812
- Water level depth detection sensor
- Solar panel
- Battery
- PIR sensor.

PIR Sensor: A PIR movement sensor has an ability to recognize movement inside an exceptionally wide point. PIR means 'Latent Infrared' on the contrary side of the PIR is a supporting driver hardware and three pin connector. The pin out for the connector on the PIR movement sensor is GND, Out and 5 V gracefully. The module has three additional pins with a jumper between two of them. PIR additionally has two potentiometers, one for altering affectability of the sensor and other for choosing the time the yield signal remains high when movement is recognized. PIR sensor voltage ranges from 5–20 V with low force utilization of 65 mA. Yield is an advanced 3.3 V sign, and the detecting range is 120° inside 7 m. It is perfect to identify the development of a human inside or out of the scope of the sensor (Fig. 8).

IOT (Internet of Things): In this framework with the utilization of IoT, the focal unit speaks with the client, enlightens the present data concerning the field and furthermore takes orders from the client and in like manner plays out the ideal activities. Utilized IoT for the treatment of information from a server, which legitimately bring the information progressively from the sensors utilized for the observing reason.

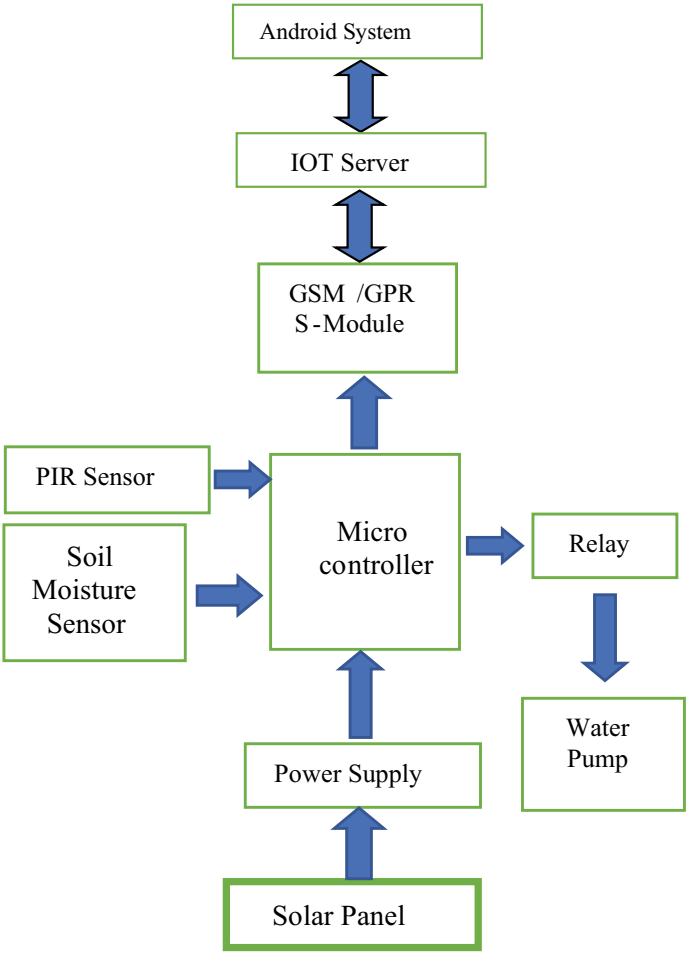


Fig. 1 Block diagram

Fig. 2 Arduino UNO board

Fig. 3 GSM/GPRS module

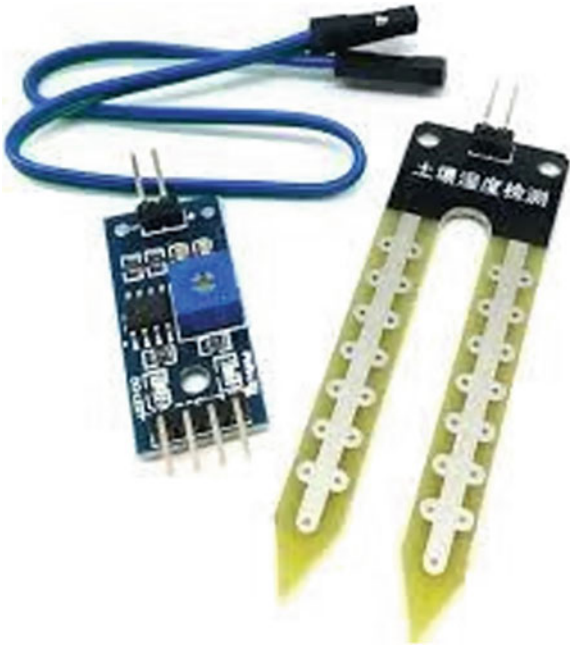


Fig. 4 Soil moisture sensor

4 Result and Conclusion

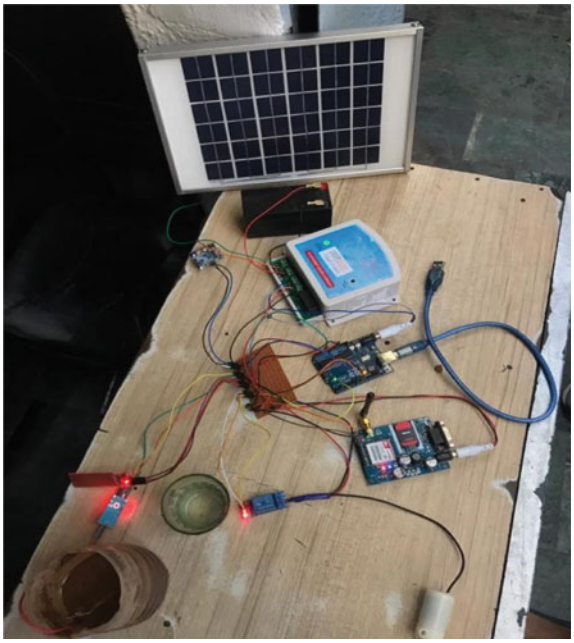


Fig. 5 12 V relay circuit

Fig. 6 Solar panel

Fig. 7 Lead acid rechargeable battery

The above picture is of the project implemented consisting of the solar panel, charge controller, rechargeable battery supplying the power to Arduino Atmel to which all the sensors like soil moisture sensor, water level detector sensor and PIR sensor are connected (Table 2).

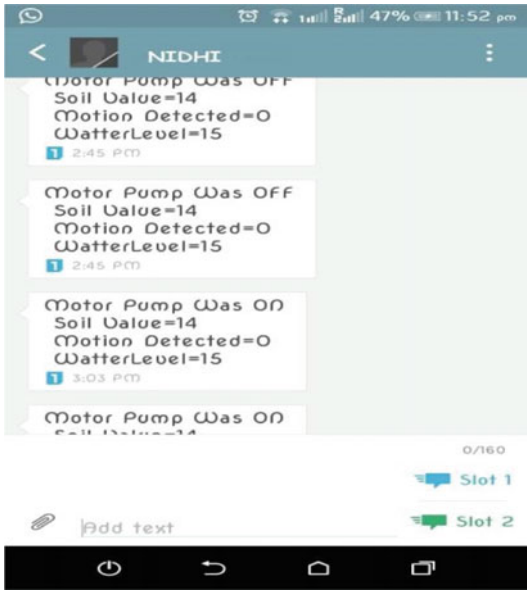
Table 1 Specification of Arduino UNO

Microcontroller	ATmega328
Operating voltage	5 V
Input voltage	7–12 V
Input voltage (limits)	6–20 V
Digital I/O Pins	14(of which 6 provide PWM output)
Analog Input Pins	6
DC Current per I/O Pin	40 mA
DC Current for 3.3 V Pin	50 mA
Flash Memory	32 KB (ATmega328) of which 0.5 KB used by bootloader
SRAM	2 KB (ATmega328)
EEPROM	1 KB (ATmega328)
Clock speed	16 MHz
Length	6.86 mm
Width	5.34 mm
Weight	25 g
Data Retain	20 year

Fig. 8 PIR sensor

Table 2 Water level sensor

Input volt	Low output	High output
5 V	10%	90%



The above screenshot is of the message that is obtained after regular interval as coded in the program. It gives the owner information about the status of the motor pump, soil moisture value, output of the motion detector sensor and the tank water level. This allows the user to have the information of all the sensors, motor and tank water level sitting away from the farm (Table 3).

The above table helps to read the status of the analog soil moisture sensor. The input voltage applied to the soil moisture sensor is 3.5 V. At this applied voltage, the sensor starts working. If the resulting output of the sensor, i.e., the conductivity between its plates is 667, then the motor starts working. If the sensor output is as low as 210, then the motor turns off (Table 4).

The above readings are of the digital PIR sensor. It is the motion sensor detector. Any motion within its range results in output as 1. It has potentiometer, with which we can set the sensitivity of the object to be determined. If no motion is detected, then the resultant output produced is 0.

The above table gives the reading of the water level sensor that determines the water level of the tank. Depending on the output of the sensor, the motor turns on and

Table 3 Soil moisture sensor

Input volt	Low output	High output
3.5 V	210	667

Table 4 PIR sensor

Input volt	Low output	High output
5 V	0	1

off. If the water level is below the 10% of the volume of tank, then the motor does not turn on irrespective of the status of the soil moisture sensor. If the water level of the tank is above 10% of the volume of the tank, provided the soil moisture sensor has low output, the motor turns on. So, this project aims toward water conservation and energy conservation with automated farming. It makes sure to make effective use of resources with the help of technology in order to make proper utilization of water and avoid wastage of electricity, which is a non-renewable source.

5 Future Scope

Automatic agriculture is the most suitable option considering today's condition as the decreasing interest of youth in agriculture. Controller-based soil monitoring will provide data of the soil condition with fertility and moisture levels of the soil for every month and can be compared to decide the usage of land for farming.

Agricultural drones are used to monitor fields and generate data accordingly. The farmers will get the bird's eye view to monitor the field and will be able to make a survey periodically with a larger view and less hazel. Agricultural drones will allow the farmer to locate the unhealthy crops and areas that are getting harmed by insects and causing damage to the crops which will lead to decrease in the production of crops and will cause loss to the farmer. There is a very large scope in this field, along with technological growth, there are multiple ways to incorporate technology with the very importance of our lives that is food.

References

1. Harishankar S, Kumar RS, Sudharsan KP, Vignesh U, Viveknath T (2014) Solar powered smart irrigation system. Adv Electron Electric Eng 4(4):341–346
2. Priyadarshi R, Soni SK, Sharma P (2019) An enhanced GEAR protocol for wireless sensor networks. Lect Notes Electric Eng. https://doi.org/10.1007/978-981-13-0776-8_27
3. Priyadarshi R, Gupta B (2019) Coverage area enhancement in wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-019-04674-y
4. Cuadros F, López-Rodrıguez F, Marcos A, Coello J (2004) A procedure to size solar-powered irrigation (photoirrigation) schemes. Sol Energy 76(4):465–473
5. Priyadarshi R, Singh L, Randheer, Singh A (2018) A novel HEED protocol for wireless sensor networks. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474286
6. Priyadarshi R, Yadav S, Bilyan D (2019) Performance and comparison analysis of MIEEP routing protocol over adapted LEACH protocol. In: Luhach AL, Hawari KBG, Mihai IC, Hsiung P-A, Mishra R B (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 237–245. https://doi.org/10.1007/978-981-13-6295-8_20
7. Priyadarshi R, Soni SK, Bhadu R, Nath V (2018) Performance analysis of diamond search algorithm over full search algorithm. Microsyst Technol 24(6):2529–2537. https://doi.org/10.1007/s00542-017-3625-0

8. Kelley LC, Gilbertson E, Sheikh A, Eppinger SD, Dubowsky S (2010) On the feasibility of solar-powered irrigation. Renew Sustain Energy Rev 14(9):2669–2682
9. Priyadarshi R, Rawat P, Nath V (2018) Energy dependent cluster formation in heterogeneous wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-018-4116-7
10. SatyaPrasanthYalla BR, Ramesh A (2013) Autonomous solar powered irrigation system. Int J Eng Res Appl (IJERA) 3(1):060–065
11. Priyadarshi R, Gupta B, Anurag A (2020a) Deployment techniques in wireless sensor networks: a survey, classification, challenges, and future research issues. J Supercomput. https://doi.org/10.1007/s11227-020-03166-5
12. Priyadarshi R, Singh L, Singh A, Thakur A (2018) SEEN: stable energy efficient network for wireless sensor network. In: 2018 5th international conference on signal processing and integrated networks, SPIN 2018. https://doi.org/10.1109/SPIN.2018.8474228
13. Abdelkerim AI, Sami Eusuf MMR, Salami MJE, Aibinu A, Eusuf MA (2013) Development of solar powered irrigation system. In: IOP conference series: materials science and engineering, vol 53, no. 1. IOP Publishing, p 012005
14. Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
15. Priyadarshi R, Bhardwaj A (2017) Node non-uniformity for energy effectual coordination in WSN. Int J Inf Technol Sec
16. Dhanne BS, Kedare S, Dhanne SS (2014) Modern solar powered irrigation system by using ARM. Int J Res Eng Technol 3(3)
17. Priyadarshi R, Yadav S, Bilyan D (2019) Performance analysis of adapted selection based protocol over LEACH protocol. In: Luhach AK, Hawari KBG, Mihai IC, Hsiung P-A, Mishra RB (eds) Smart computational strategies: theoretical and practical aspects. Springer, Singapore, pp 247–256. https://doi.org/10.1007/978-981-13-6295-8_21
18. Ganesan T, Elamvazuthi I, Shaari KZK, Vasant P (2013) Hypervolume-driven analytical programming for solar-powered irrigation system optimization. In: Nostradamus 2013: prediction, modeling and analysis of complex systems. Springer, Heidelberg, pp 147–154
19. Priyadarshi R, Gupta B, Anurag A (2020b) Wireless Sensor Networks Deployment: A Result Oriented Analysis. Wireless Pers Commun. https://doi.org/10.1007/s11277-020-07255-9
20. Rawat P, Chauhan S, Priyadarshi R (2020) Energy-efficient clusterhead selection scheme in heterogeneous wireless sensor network. J Circ Syst Comput. https://doi.org/10.1142/S0218126620502047
21. Deveci O, Mert O, Ozgur HU, Zafer O (2015) Design and development of a low-cost solar powered drip irrigation system using systems modeling language. J Cleaner Prod 102:529–544
22. Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol. https://doi.org/10.1007/s00542-020-04874-x
23. Dursun M, Ozden S (2012) Application of solar powered automatic water pumping in Turkey. Int J Comput Electri Eng 4(2):161
24. Randheer SSK, Kumar S, Priyadarshi R (2020) Energy-aware clustering in wireless sensor networks lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_39
25. Kumar S, Soni Randheer SK, Priyadarshi R (2020) Performance analysis of novel energy aware routing in wireless sensor network lecture notes in electrical engineering. https://doi.org/10.1007/978-981-15-2854-5_44
26. Dursun M, Özden S (2014) An efficient improved photovoltaic irrigation system with artificial neural network based modelling of soil moisture distribution–a case study in Turkey computers and electronics in agriculture. 102:120–126
27. Priyadarshi R, Tripathi H, Bhardwaj A, Thakur A (2018) Performance metric analysis of modified LEACH routing protocol in wireless sensor network. Int J Eng Technol. https://doi.org/10.14419/ijet.v7i1.5.9146

28. Priyadarshi R, Singh L, Kumar S, Sharma I (2018) A hexagonal network division approach for reducing energy hole issue in WSN. Euro J Pure Appl Mathe
29. Kumar M, Reddy KS, Adake RV, Rao CVKN (2015) Solar powered micro-irrigation system for small holders of dryland agriculture in India. Agric Water Manag 158:112–119
30. Anurag A, Priyadarshi R, Goel A, Gupta B (2020) 2-D coverage optimization in WSN using a novel variant of particle swarm optimisation. In: 2020 7th international conference on signal processing and integrated networks, SPIN 2020. https://doi.org/10.1109/SPIN48934.2020.9070978
31. Eker B (2005) Solar powered water pumping systems. Trakia J Sci 3(7):7–11
32. Priyadarshi R, Singh L, Sharma I, Kumar S (2018) Energy efficient leach routing in wireless sensor network. Int J Pure Appl Math 118(20):135–142
33. Balaji VR, Sudha M (2016) Solar powered auto irrigation system. Int J Emerg Technol Comput Sci Electron (IJETCSE) 20(2):203–206
34. Codreanu N, Gaudentiu V, Ciprian I (2014) Solar powered wireless multi-sensor device for an irrigation system. In: Proceedings of the 2014 37th international spring seminar on electronics technology. IEEE, pp442–447

A Novel Algorithm for Dynamic Control and Modeling of Myoelectric Signals for Prosthetics

Anudruti Singha and R. K. Mugelan

Abstract In present-day scenario, major efforts have been put into the development of human prosthetic models from electromyographic (EMG) signals to achieve more natural control for re-habitation of humans. Myoelectric control of prosthetic arms provides light for providing naturalistic movements. The objective of this article is to reconstruct the input myoelectric signal by removing the noise, mitigating the electrical and mechanical disturbances present in the EMG signal. To meet the above objective, a novel algorithm called Anudruti algorithm is proposed which makes use of Fourier transform spectral analysis and Kalman filter. The proposed algorithm smoothens the input signal to the prosthetic model, thereby enhancing the performance of the system even with worst input conditions. The algorithm is mathematically modeled, and the simulated results are obtained in support of the model. The Anudurti algorithm is put under various test conditions from best to worst, and the results are found to be more promising than the conventional system. The test results show that the system achieves its maximum efficiency of 94% and to a minimum of 73%.

Keywords Prosthetic model · EMG · Myoelectric signal · Kalman filter

1 Introduction

Electromyogram (EMG) signal is an electrical signal that is produced due to muscle contractions [1]. This EMG signal is having a major field of application such as identification of patients with motion imbalance, muscle functioning, and a major part in the development of prosthetic which aids the physically disabled peoples. The signal always has major noise issues in it which can be resolved in certain ways. The study proposes a Kalman filter technique to decode joint angles in *x*- and *y*-planes

A. Singha (✉) · R. K. Mugelan
School of Electronics Engineering (SENSE), Vellore Institute of Engineering and Technology, Vellore, India
e-mail: anudruti.singha2017@vitstudent.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_35

(θ_x and θ_y) from surface EMG signal [2]. The Kalman filter is first trained using training data to set its parameters, which are subsequently used to decode the EMG signals of the testing dataset to identify the movement direction. The paper helped us to understand how to model the angles and analyze the stability using the Kalman filter [3]. The ideology and the parameter selection such as force, voltage, and angle are taken from the papers.

As with all research reports and surveys on medical prosthetics explains that the prosthetics are an insignificant role toward human rehabilitation and growth in everyday life. From the surveys in the near past, it is evident that there are undoubtedly significant issues yet to be addressed, such as.

- Noise in EMG [4] signals due to different external and internal factors
- Electrical disturbances, i.e- majorly voltage errors due to improper placement of EMG leads, circuit problems, temperature, etc.
- Mechanical disturbances majorly instability due to high or low resistance, improper working of sensors which further lead to electrical instability.

These key issues lead to improper functioning of the prosthetic and leading to high prone damage in shorter periods which impacts the instability or fear of falling. It also tends to generate fatigue and reduced mobility, such as the complexity of the nonlinear system, unavailability of an accurate and precise mathematical model.

This article aims to provide a better quality of service to the user with reduced complexity. The objective is met by conditioning [5, 6] the input signal to the control system computational algorithm with an advanced white line Kalman filter altogether leading to a novel algorithm termed as 'Anudruti' algorithm. This novel algorithm is designed to reduce the system complexity and to maintain the stability of the system as well as to mitigate the imbalance in the system on its own without using any feedback. Hence, complete integration can be obtained by a joined, coordinated style of the hand optimization that considers the target not solely the intrinsic performance of every single scheme. Despite the intensive efforts being created with entirely different techniques mentioned [7], finding an appropriate algorithm that permits the decoding of arm movement exploitation surface myogram with high efficiency was hardly achieved. A recent study [8] that used the Kalman filter with a myogram signal to rewrite the movement resulted in poor performance.

2 System Modeling

The Anudruti algorithm works with signal conditioning followed by the input to the system, and the flow is explained as follows. The input EMG signal to the prosthetic arm fluctuates leading to a time-invariant input to the system which puts the system's stability under test. The EMG signal obtained from the prosthetic arm (user) is conditioned as shown in Fig. 1, by sampling according to our requirement followed by frequency-domain conversion by fast Fourier transform. The inverse Fourier transformed EMG signal is passed through a band-pass filter designed according to our

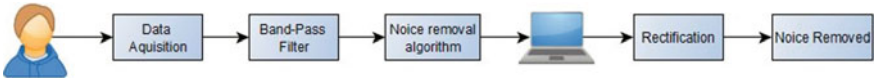


Fig.1 Signal conditioning in Anudruti algorithm

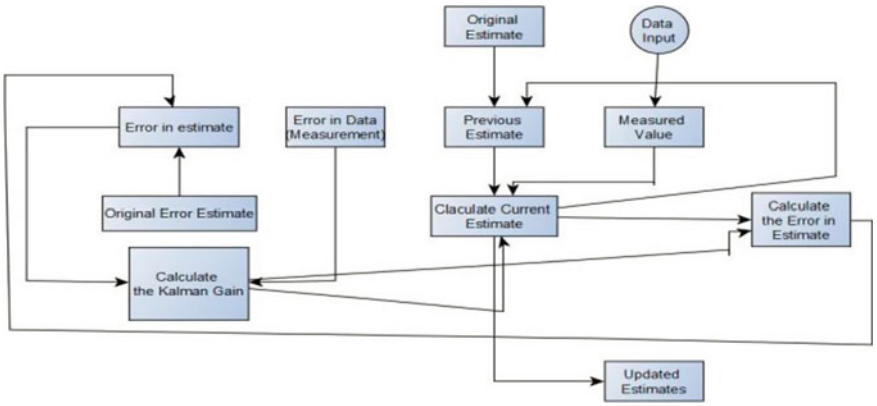


Fig. 2 Kalman filtering in Anudruti algorithm

required cut-off frequency. Maximum and minimum contraction voltages (MAV) are extracted by convolving the filtered signal with the original signal.

The conditioned EMG signals are passed through the Kalman filter as shown in Fig. 2, for the statistical analysis to find the most optimum averaging factor for each following voltage levels of the conditioned EMG signal. The processed signal is passed through motor control and computed through computational algorithm, i.e., transfer function of the system. The output of the system is analyzed for the stability condition in frequency domain for the performance analysis.

3 Mathematical Modeling

Let $V_a(t)$ be the input EMG signal, $(\theta_m = A)$ the armature output angle, AR the armature resistance, AL the armature inductance, $V_b(t)$, the armature (output) voltage and $I_a(t)$ the output armature current. Equations. (1), (2), (3), (4) and (5) show how the input EMG signal controls the output armature voltage.

$$I_a = I_b = \frac{\mathrm{d}V_b}{\mathrm{d}t} \tag{1}$$

$$V_b(t) = K_b \frac{\mathrm{d}A}{\mathrm{d}t} \tag{2}$$

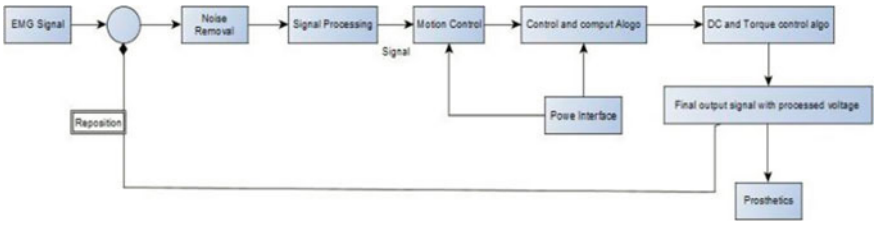


Fig. 3 Block diagram of system model Anudruti algorithm

$$V_b(s) = K_b s A \tag{3}$$

$$V_a = A_R I_a + A_L \frac{\mathrm{d}i_a}{\mathrm{d}t} + V_b(t) \tag{4}$$

$$V_a(t) = A_R \frac{\mathrm{d}V_b}{\mathrm{d}t} + A_L \frac{\mathrm{d}^2 V_b}{\mathrm{d}t^2} + V_b(t) \tag{5}$$

The system model of anudruti algorithm is shown in Fig. 3.

The conditioned signal is passed through the Kalman filter to minimize the fluctuation as shown below.

Kalman filter takes input from a time instant with particular time intervals and helps us to calculate how much of the prediction and measurement is required to be imparted in the next set of input in next interval; the Kalman filter minimizes the error and helps us in predicting the most accurate value as in Eq. (6),

$$\frac{E_{\mathrm{EST}}}{E_{\mathrm{EST}} + E_{\mathrm{MEA}}} \tag{6}$$

The Kalman range always be $0 < \mathrm{KG} < 1$.

Where KG is Kalman gain, E_{EST} error in the estimate, E_{MEA} error in the measurement, EST is the estimated value and MEA is the measured value.

$$\mathrm{EST}_t = \mathrm{EST}_{t-1} + \mathrm{KG}\left[\mathrm{MEA} - \mathrm{EST}_{t-1}\right] \tag{7}$$

$$E_{\mathrm{EST}(t)} = [1 - \mathrm{KG}]\left(E_{\mathrm{EST}(t-1)}\right) \tag{8}$$

If the KG is high then E_{EST} is larger than E_{MEA}, therefore, the difference (MEA—EST_{t-1}) is considered to be larger which means, when the next level of voltage is given, the response time to reach the desired output is reduced. Hence, there are minimal data variations in voltage.

If KG is low makes the E_{MEA} be larger which makes the difference (MEA—EST_{t-1}) to have a minimal adjustment. However, if the E_{MEA} is significant, the E_{EST}

will consider the actual values of multiple voltage points for better performance. In all cases, it is evident that the $E_{\mathrm{EST}(t-1)}$ is reduced as time tends to infinity.

The initial state consists of a state matrix(X_0) and a process covariance matrix (P_0); the state matrix contains the voltage values. The process covariance matrix (P_0) is keeping track of the error; as it goes through the process, the current state is updated as the previous state. The X_{kp} and P_{kp} are the new state predictions where V_k is the control variable matrix, the predicted state noise matrix is W_k and process noise covariance matrix is Q_k.

The measured values are added with X_{kp} and P_{kp} which is affected by noise Z_k. The Kalman gain KG decides the percentage of measured value (MEA) and the percentage of predicted value (EST) which will be used to predict the new state. The Kalman gain takes the process covariance matrix (P_0) and the vector R, which is sensor noise covariance matrix to predict the new state. The updated covariance error matrix predicts the new output and new error, and the process is repeated to get the minimal error.

The signal flow through the Kalman filter is shown below,

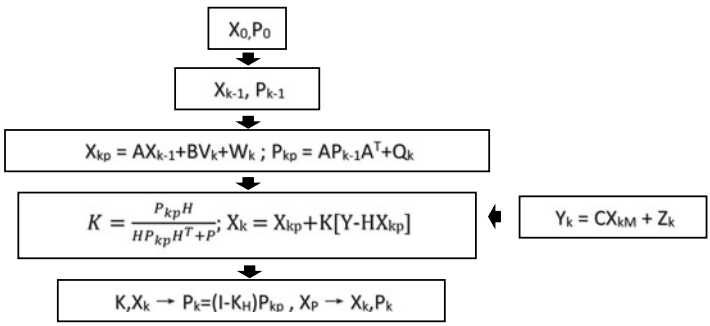


Table 1 shows the reduction in the error estimate with increasing time intervals using the Kalman filter.

Table 1 Theoretical calculation

	MEA (mv)	E_{MEA}	EST (mv)	$E_{\mathrm{EST}(t-1)}$	KG	$E_{EST(t)}$
t–1			68	2		
T	74	4	71.33		0.34	1.33
t + 1	71	4	70.50		0.27	1.00
t + 2	70	4	69.40		0.24	0.8
t + 3	73	4	72		0.17	0.66

4 Result Analysis and Discussion

Seven datasets of information were recorded with time intervals of 2, 5 and 10 min. The electromyogram activity from the left arm is recorded at the same time intervals. The initial four datasets recorded had electromyogram signals from striated muscle alongside the arm during a two-dimensional plane within the *x*- and *y*-axis directions as shown in Fig. 4.

The point of interest of the signal is the amplitude, which has a range between 0 and 24 millivolt (peak to peak), and the frequency of an EMG signal is between 0 and 520 Hz. However, the usable energy of the EMG signal lies in the range of 45–160 Hz. To avoid the distortion of the input EMG signal, a signal conditioning block is incorporated by using the proposed Anudruti algorithm. The EMG signal is domain shifted with fast Fourier transform as shown in Fig. 5. The band-pass filter is designed according to our required cut-off frequency window which ranges in between 5 and 10 Hz. Test samples are filtered out using the above window range get the inverse Fourier transform of the required signal as shown in Fig. 6.

The obtained signal is computed for the best-fit value through convolution by using mean average value (MAV) theorem and grouping together for future inheritance by *K*-clustering algorithm of the processed signal as shown in Fig. 7. The final comparative signal gives us the output with min contractions, max relaxation and

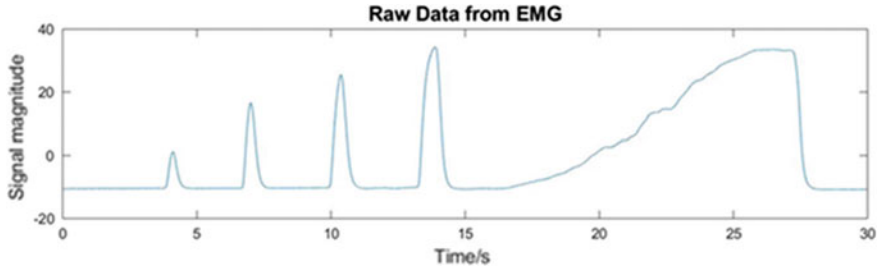


Fig. 4 Raw EMG data

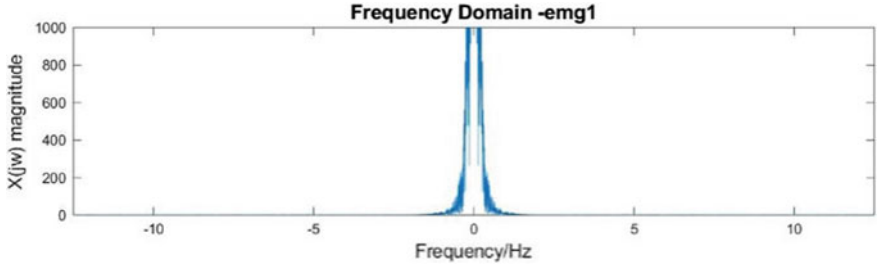


Fig. 5 FFT and shift

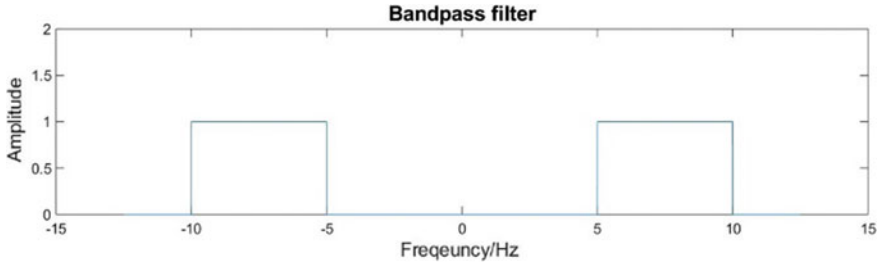


Fig. 6 Band-pass filer design

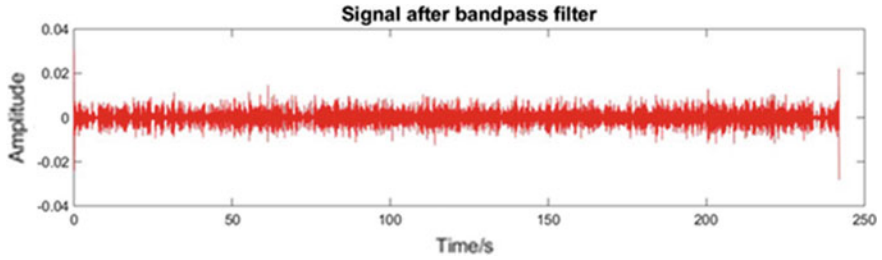


Fig. 7 EMG filter signal after passing through a band-pass filter

the exact comparative response. The resultant processed signal will be noise-free, interference-mitigated and distortion-less EMG signal as shown in Fig. 8.

The processed signal from the Kalman filter is given as the input to the system using the maximum contraction and minimum contraction values of the signal. The performance of the system is analyzed to measure the error and stability of the system. The data points in Fig. 9 and Table 2 shows the correlation between the actual outputs with the desired output under various circumstance ranging from the worst to best.

Here, in the dimensional format of cross-correlation coefficient, the system is able to achieve 70% of the desired output in the worst scenario, 81% in medium and 93% in the best scenario cases, respectively.

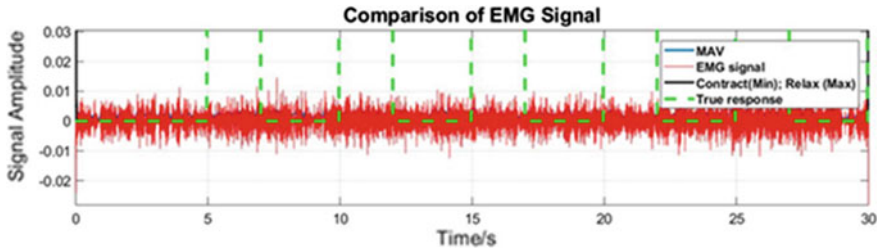


Fig. 8 Steady comparative EMG signal with max and min contractions and most fit region

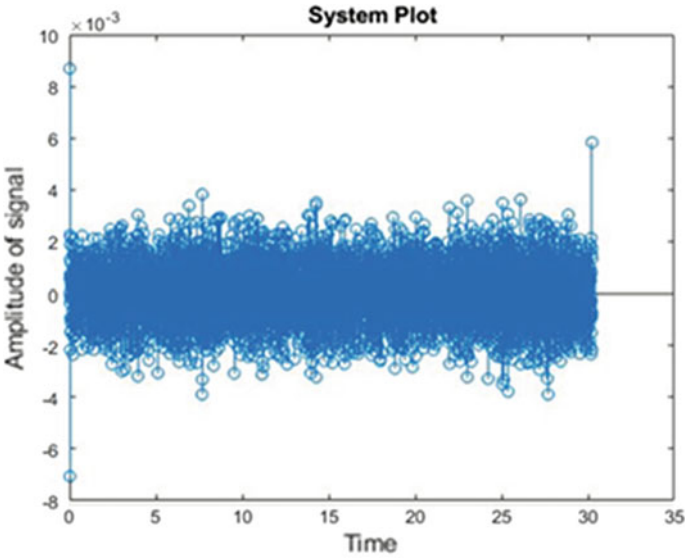


Fig. 9 Correlation coefficient data points

Table 2 Correlation tests

Worst scenario		Medium scenario		Best scenario	
1.0000	−0.7032	1.0000	−0.8154	1.0000	−0.9356
−0.7032	1.0000	−0.8154	1.0000	−0.9356	1.0000

5 Mathematical Analysis

The transfer function of the prosthetic arm using Anudruti algorithm is given in Eq. (9)

$$\frac{A(s)}{V_a(s)} = \frac{1}{A_L K_b s^3 + A_R K_b s^2 + K_b s} \tag{9}$$

Here, gear ratio is not added as there is no load in the entire process, and the processed EMG signal is used just to control the armature.

Taking ideal case $A_L = 0$

$$\frac{A(s)}{V_a(s)} = \frac{1}{s(A_R K_b s + K_b)} \tag{10}$$

For analysis, A_R is set to be 2.7 Ω and K_b as 1.3

$$G(s) = \frac{A(s)}{V_a(s)} = \frac{1}{s(3.51s + 1.3)} \tag{11}$$

The stability analysis of Eq. (11) using root locus is shown in Fig. 10. The graphical analysis shows two poles one at 0 and another at −0.412, with two asymptotes as 90° and 270°, respectively. The centroid falls at −0.206, and the breakaway point exists at −0.185, with an equation of $3.51s^2 + 1.3\ s + K$.*K*With $K > 0$.

The bode plot of the Eq. (11) under frequency-domain analysis is shown in Fig. 11. The analysis provides the gain margin to be infinite, and the gain crossover frequency also tends to infinite. Phase margin value is better and found to be 38.0169 and phase crossover frequency value of 0.4738 Hz. Any system whose gain crossover frequency tends to infinity is gaining its maximum stability as time tends to infinity. From the above results and discussion, it is evident that the proposed Anudruti algorithm

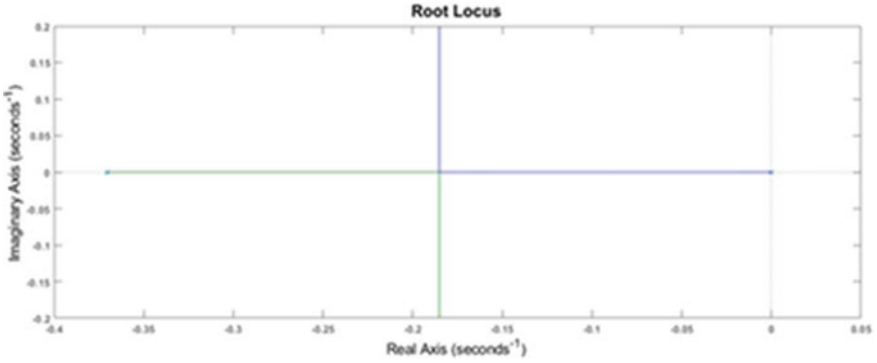


Fig. 10 Root locus

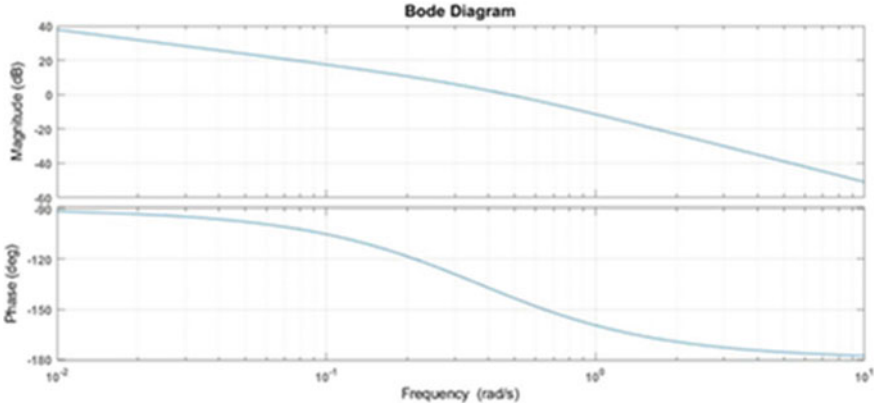


Fig. 11 Bode plot

improves the stability of the system even when the input signal to the system is distorted.

6 Conclusion

The mathematical model and the simulated results give an evident proof of the enhanced performance of the system when subjected to the proposed algorithm. The Anudruti algorithm is imparted with the existing conventional prosthetic model and simulated under various scenarios to test the performance. The frequency-domain analysis of the Anudruti algorithm imparted system shows the enhanced stability of the system under all scenarios giving an average efficiency of 81%.

References

1. Geethanjali P, Krishna Mohan Y, Bhaskar P (2013) A low-cost EMG-EOG signal conditioning system for brain computer interface applications. Int J Eng Technol (IJET) 5:2268–2269
2. ElMohandes H, Audi JMC, Ruff R, Eldawlatly S, Hoffmann KP (2018) Decoding arm kinematics from EMG signals using Kalman filter. In: International conference on advances in bioprocess engineering and technology, April 23–25, pp. 96
3. Triwiyanto, Wahyunggora O, Nugroho HA, Herianto (2018) Evaluating the linear regression of Kalman Filter model elbow joint angle estimation using electromyography signal. In: International conference on science and applied science (ICSAS), pp 5–7
4. Ngo TB, Le HL, Nguyen TH (2009) Survey of Kalman filters and their application in signal processing, In: International conference on artificial intelligence and computational intelligence, pp 2–4
5. Ramesh Babu P, Anandanatarajan A Signals and systems. Scitech publications(INDIA), pp 281–356.
6. Ramesh Babu P Digital signal processing, 4th edn. Scitech Publications(INDIA), pp 5.6–5.16.
7. Hotson G, Smith RJ, Rouse AG, Schieber MH, Thakor NV, Wester BA (2015) High precision neural decoding of complex movement trajectories using recursive Bayesian estimation with dynamic movement primitives. In: IEEE robotics and automation letters. Preprint Version, pp 3–4
8. Sarasola-Sanz A, Irastorza-Landa N, Shiman F, Lopez-Larraz, Martin Spuler E, Birbaumer N, Ramos-Murguialday A (2015) EMG-based multi-joint kinematics decoding forrobot-aided rehabilitation therapies. In: IEEE International Conference on Rehabilitation Robotics (ICORR), pp 230–231
9. Lu X, Wang L, Wang H, Wang X (2016) Kalman filtering for delayed singular systems with multiplicative noise. IEEE/CAA J Automat Sinica 3(1)
10. Tsyganova JV, Kulikova MV (2017) SVD-based Kalman filter derivative computation. IEEE Trans Automatic Control 62(9)
11. Spanias JA, Perreault EJ, Hargrove LJ, (2016) Detection of and compensation for EMG disturbances for powered lower limb prosthesis control. IEEE Trans Neural Syst Rehabili Eng 24(2)
12. Guo W, Sheng X, Liu H, Zhu X (2017) Toward an enhanced human–machine interface for upper-limb prosthesis control with combined EMG and NIRS signals. In: IEEE Trans Hum-Mach Syst 47(4)

Minimization of Drain-End Leakage of a U-Shaped Gated Tunnel FET for Low Standby Power (LSTP) Application

Suman Das, Avik Chattopadhyay, and Suchismita Tewari

Abstract In this paper, for the first time, the transfer characteristic of a '*U*'-shaped gated tunnel FET (TFET) has been thoroughly investigated considering the real-time adverse effects of gate-to-drain direct tunneling current and gate-induced drain leakage (GIDL) effect using SILVACO ATLAS device simulator. Clearly, these leakage phenomena degrade the device performance, especially for low standby power (LSTP) operation. Hence, for the first time, a novel design modification has been proposed in terms of the optimization of the oxide thickness (T_{GD}) of right vertical arm of the U-shaped gate, in order to mitigate the aforementioned problem. It has been found that when the T_{GD} value is increased to 7 nm from the equivalent oxide thickness (EOT) value of 1.6 nm, the ultimate device becomes optimized in terms of the performance matrices like, I_{OFF}, SS_{min}, I_{ON}/I_{OFF}, etc. Moreover, 43% reduction in delay and almost 11 decades of OFF-state power reduction have been obtained for the optimized device than that of the device having $T_{GD} = 1.6$ nm, for gate length of 70 nm.

Keywords U-shaped gate · Tunnel FET · Direct tunneling · GIDL

S. Das (✉)
Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, Sikkim, India
e-mail: das.suman194@gmail.com

A. Chattopadhyay · S. Tewari
Department of Radio Physics and Electronics, University of Calcutta, Kolkata, India
e-mail: avikjoy@yahoo.com

S. Tewari
e-mail: suchi_tewari@yahoo.co.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_36

1 Introduction

Over the last few decades, a continuous device miniaturization, in search for introducing more and more functionality in a single chip, while maintaining its compactness and easy portability has pushed the individual devices to work under extremely low power supply. In this regard, tunnel FETs have long outperformed the conventional MOSFETs, with its sub 60 mV/decade subthreshold swing (SS) leading to ultimately high I_{ON}/I_{OFF} ratio [1–3]. But, beyond the subthreshold region, when it comes to super-threshold region, the TFET devices have failed to achieve the high ON-current (I_{ON}) value and that limits the application of TFET devices in analog and mixed signal domain. In order to achieve high ON-current while maintaining its superior subthreshold performance, a number of schemes have been proposed, starting from innovative device architectures, viz. elevated drain and elevated source structures [4, 5], '*U*'-shaped gate [6], vertical TFETs [7], etc., to using alternate materials, like Ge, SiGe [8], etc. Interestingly, in all the proposed schemes [6, 8], the transfer characteristics for TFET devices have been generated, without considering the two main real-time leakage current components, i.e., (1) gate-to-drain direct tunneling or gate-to-drain leakage, and (2) current due to gate-induced drain leakage (GIDL) effect. Hence, the prediction of the apparent high device performance matrices, in those published papers [6, 8], is questionable in real-time domain.

In this paper, we have considered a '*U*'-shaped gate-based TFET structure and for the first time have done a thorough analysis of its transfer characteristic, considering the two aforementioned real-time adverse effects, i.e., gate-to-drain leakage and GIDL, and have observed that the unwanted leakage has a serious detrimental effect on the transfer characteristics of the device, especially in the subthreshold region, resulting inferior device performance, than that of the case without considering the leakage components. Based on this, for the first time, we have proposed a design modification, in terms of the optimization of the thickness (T_{GD}) of right vertical arm of the '*U*'-shaped gate oxide, to obtain the desired performance in the subthreshold region, without affecting the super-threshold region performance, resulting the most high-performance optimized device for low standby power (LSTP) operation. The projected device is termed as '*U*'-shaped optimized elevated source drain-TFET (UOESD-TFET). We have restricted the material to Si only, in order to keep the device cost-effective.

2 Device Structure and Simulation Framework

Figure 1a shows a skeleton of the projected '*U*'-shaped optimized elevated source drain-TFET (UOESD-TFET) device structure, which is an amalgamation of the reported devices, elevated drain-TFET (ED-TFET) [4], and elevated source-TFET (ES-TFET) [5]. The use of '*L*'-shaped gate, used in ED-TFET has achieved the reduction in ambipolar phenomenon [4] and that in ES-TFET has achieved an enhancement

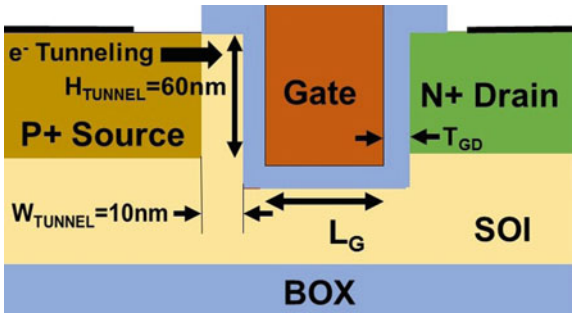


Fig. 1 Skeleton of the projected *U*-shaped optimized elevated source drain-TFET (UOESD-TFET) having $L_G = 70$ nm

of the source-to-channel tunneling area and hence an improvement in tunnel current [5]. In this UOESD-TFET device, both the features have been utilized, in the form of using a '*U*'-shaped gate, in order to get high channel current with reduced ambipolar effect. The step-by-step fabrication process flows of the above-mentioned ED-TFET and ES-TFET on silicon-on-insulator (SOI) substrate are already demonstrated in [4] and [5], respectively. In the UOESD-TFET device, the gate length is taken as 70 nm, with an EOT of 1.6 nm, following the ITRS roadmap [9], The device consists of a p^+ source (10^{20} acceptors/cm^3), an n^+ drain (10^{20} donors/cm^3), and an intrinsic channel (10^{17} donors/cm^3) regions. The values of different device parameters taken are consistent with that of the reported ED- and ES-TFET devices [4, 5].

Device simulator SILVACO Atlas [10] has been used for the simulation purpose. The non-local BTBT model [11] has been invoked. This, based on the actual potential profile of the tunneling barrier instead of assuming a maximum or an average field throughout the barrier, can provide a more refined form of tunnel probability for accurate calculation of indirect tunneling current density. CVT Lombardi model and the position-dependent Bohm quantum potential (BQP) model are employed to incorporate, respectively, the carrier mobility and the carrier confinement with the variations of local potential on the scale of the electron wave functions [12]. Also, the band gap narrowing (BGN) model together with the F-D statistics for carrier distribution is enabled at 300 K of lattice temperature to account high doping in source/drain regions [13]. Shockley–Read–Hall (SRH) and Auger models [14] are triggered on for calculating carrier lifetime.

3 Model Calibration

For model calibration purpose, the gate metal work function is adjusted, and the diffusion length for carriers in silicon is fine-tuned, as it largely affects the OFF-state leakage current, besides being impacted by the PAT. The ON-state current is adjusted primarily by varying tunnel masses for carriers in non-local BTBT process [11] and by fitting a couple of adjustable quantum parameters of BQP model. A few parameters related to CVT Lombardi model have also been used as fitting parameters for slight

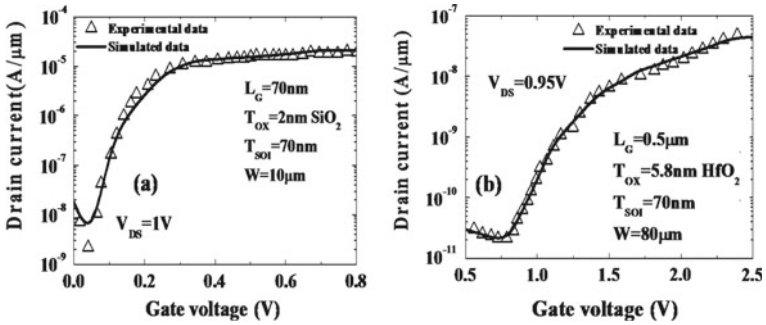


Fig. 2 Validation of simulated characteristics with the experimental data, (**a**) for ED-TFET [3] and (**b**) for ES-TFET [4]

adjustment of carrier mobility in the channel region. A good agreement between experimental [4, 5] and simulated characteristics can clearly be evidenced from Fig. 2a, b for ED- and ES-TFET, and this claims the authenticity of our simulation scheme.

4 Results and Discussion

Figure 3a shows the transfer characteristics of the ideal ESD-TFET device, without considering any type of leakage current, i.e., only the channel current (I_{channel}) is present. The inset of Fig. 3a is the device schematic, wherein the I_{channel} current path has been indicated by the black arrows. Figure 3b shows the corresponding band diagram of the ESD-TFET device to demonstrate its ON-state and OFF-state behavior. It is clearly shown that in the OFF-state, when $V_{\text{GS}} = 0$ V and $V_{\text{DS}} = 1$ V, the conduction band edge (E_C) in the channel region remains well above the valence band edge (E_V) in the source region, and hence, almost no current or negligible amount of current is passing through the device; but when V_{GS} starts to increase and ultimately reaches to 1 V, considered as ON-state, the E_C in the channel region starts to bend more and ultimately goes down below E_V in the source region, creating an efficient electron tunneling from source to channel, which gives rise to sudden increase in channel current. Figure 4a–g shows the channel current, gate-to-drain leakage current ($I_{\text{gate-leakage}}$), and the resultant or actual device current, which is nothing but the sum of the above two components and is termed as the resultant current ($I_{\text{resultant}}$), as a function of gate-to-source voltage (V_{GS}) with the gate-to-drain oxide thickness (T_{GD}) as a parameter. Interestingly, it has been observed, from Fig. 4a, for T_{GD}= 1.6 nm, that although I_{channel} maintains a low off-current and a high $I_{\text{ON}}/I_{\text{OFF}}$ ratio, the plot of $I_{\text{resultant}}$, which is the actual device current, considering the real-time adverse effect of gate-to-drain leakage component, is telling another story. When T_{GD} is 1.6 nm, the $I_{\text{gate-leakage}}$ remains much higher than I_{channel} throughout

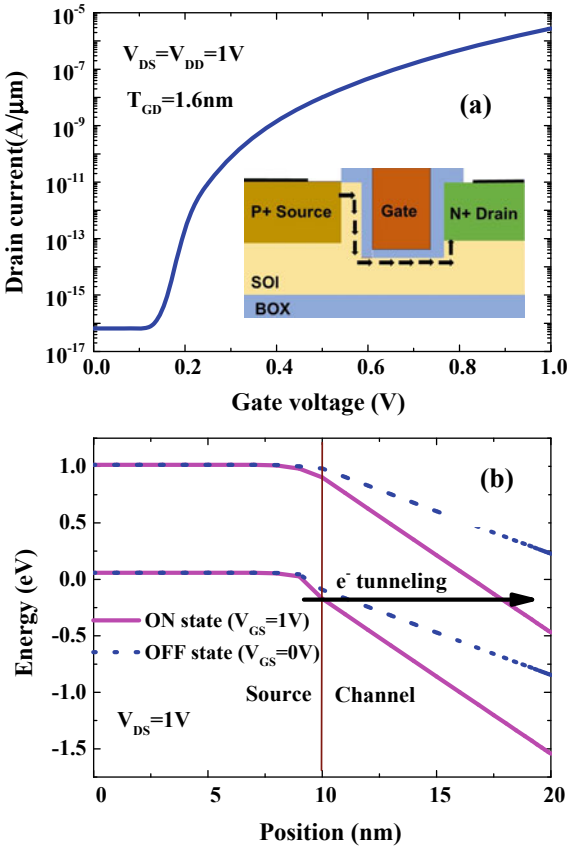


Fig. 3 **a** Ideal current–voltage characteristics for the ESD-TFET device [inset shows the path of $I_{channel}$], **b** Band diagrams of the ESD-TFET device showing the ON-state and OFF-state behavior of the device

the lower V_{GS} range up to $V_{GS} = 1$ V, then drops suddenly and ultimately increases again to a value, which is much lower than the $I_{channel}$. This may be explained by the fact, that when $V_{DS} = 1$ V and $V_{GS} = 0$ V, a potential difference and hence an electric field have been created from drain to gate, which is strong enough to make the electrons to tunnel through the oxide (T_{GD}) from the gate metal to drain. The strength of this field starts to decrease as V_{GS} starts to increase, and as soon as V_{GS} becomes equal to V_{DS}, i.e., 1 V, the potential difference becomes zero, and a sudden drop of $I_{gate\text{-}leakage}$ takes place at $V_{GS} = V_{DS} = 1$ V. Further increase in V_{GS}, changes the direction of electric field and electrons starts to tunnel from drain to gate metal through the same oxide. But, as this tunneling is from semiconductor to metal, the magnitude of the current in opposite direction is significantly lower than that of $I_{channel}$, and thus, beyond $V_{GS} = 1$ V, $I_{channel}$ dominates. As a result, below $V_{GS} = 1$ V, $I_{resultant}$ (comprising $I_{gate\text{-}leakage}$, and $I_{channel}$) coincides with $I_{gate\text{-}leakage}$, and above $V_{GS} = 1$ V, it coincides with $I_{channel}$, respectively. This leads to higher I_{OFF}-value than the I_{ON}-value for the same device, and that is highly undesirable. Here comes the motivation of altering the device design to minimize this $I_{gate\text{-}leakage}$, and this is

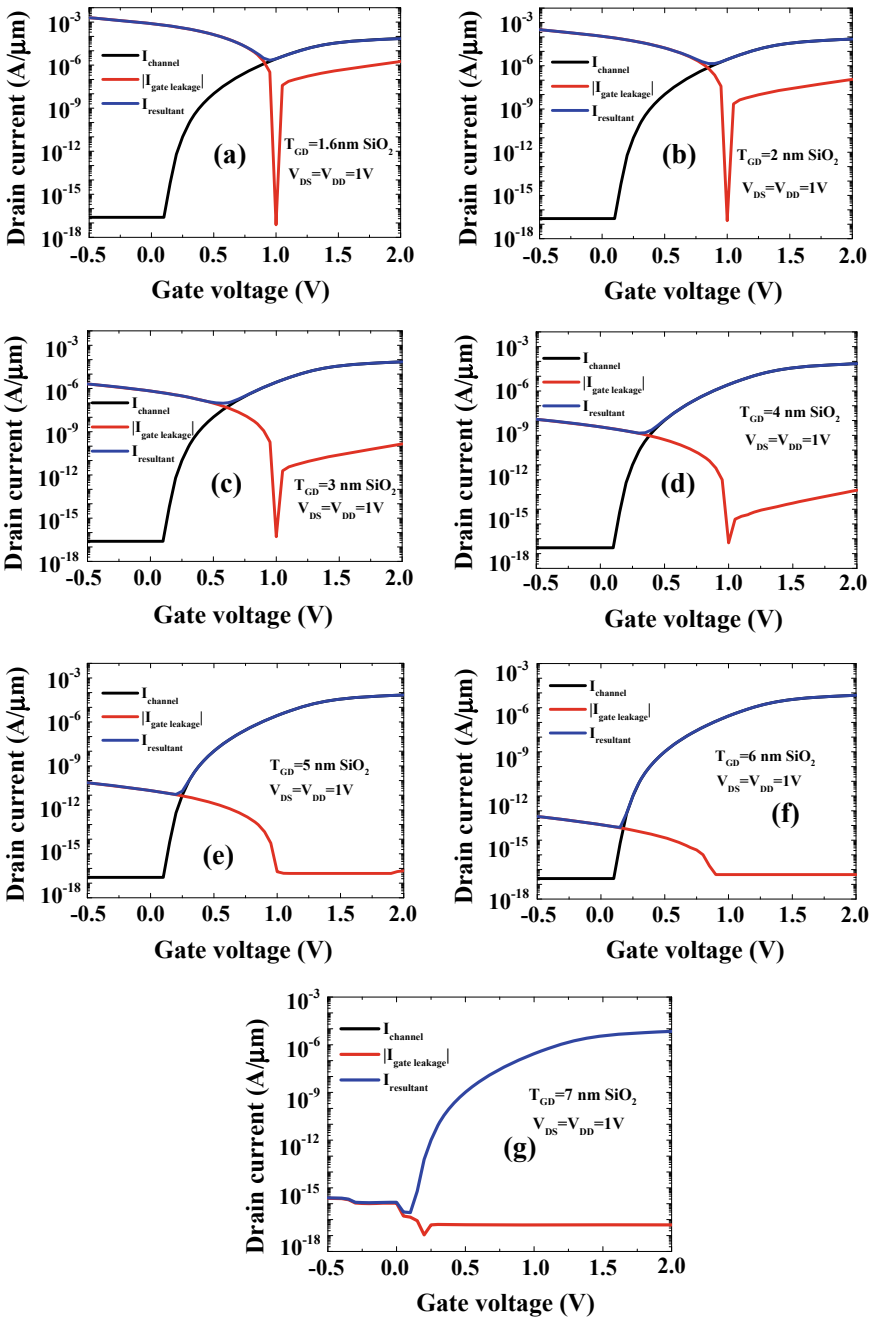


Fig. 4 **a–g** Plot of the channel current ($I_{channel}$), gate-to-drain leakage current ($I_{gate\text{-}leakage}$), and the resultant or actual device current [comprising $I_{channel}$ and $I_{gate\text{-}leakage}$] with respect to gate-to-source voltage (VGS) with the gate-to-drain oxide thickness (TGD) as the parameter

done in terms of increasing the value of T_{GD}. From Fig. 4a–g, it is clearly seen that as T_{GD} is increasing from 1.6 nm, the $I_{gate\text{-}leakage}$ components become lower and lower than that of the immediate previous device having thinner T_{GD}-value and goes down below 1 pA/μm, which (for all practical purposes) may be considered as the lower limit of the measurable value for current [15], when $T_{GD} = 6$ nm. Figure 5a, b) shows the band diagrams for the device having $T_{GD} = 1.6$ nm and 7 nm, respectively. It is clearly witnessed that the increased value of T_{GD} has prevented the penetration of electrons from metal gate to drain through tunneling. Moreover, no band bending in the channel region ensures the fact that no gate-induced drain leakage (GIDL) effect is taking place either, irrespective of the value of T_{GD}. This gives a clear indication about the choice of drain doping while doing the innovative design, and that is for our case is 10^{20} donors/cm^3, as stated earlier. So, to calculate $I_{resultant}$, only channel current and gate-to-drain leakage component have been considered, as current due to GIDL (I_{GIDL}) has been found to be zero over here. Figure 6 shows the actual device transfer characteristics, i.e., the plot of $I_{resultant}$ which is now termed as drain current, as a function of V_{GS} with T_{GD} as a parameter. Figure 6 clearly shows

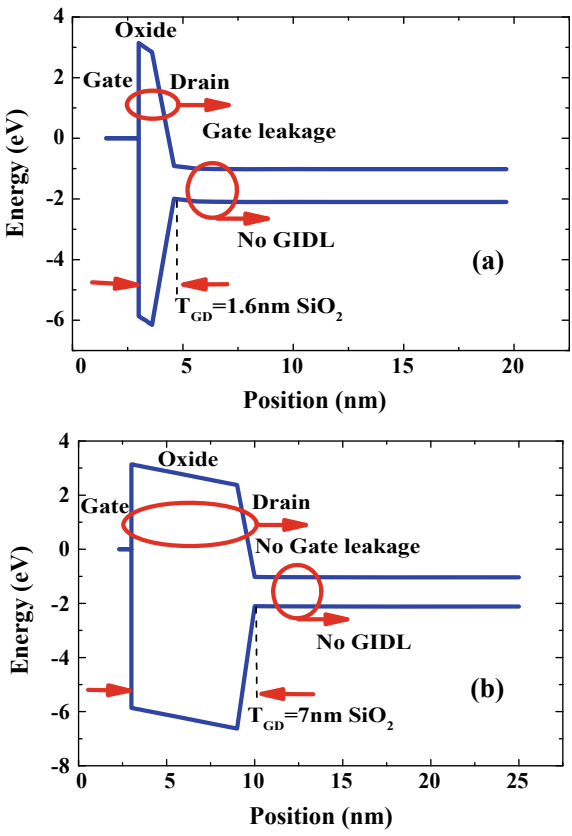


Fig. 5 Band diagrams indicating the presence and absence of different leakage components for the T_{GD} of **a** 1.6 nm and **b** 7 nm, respectively

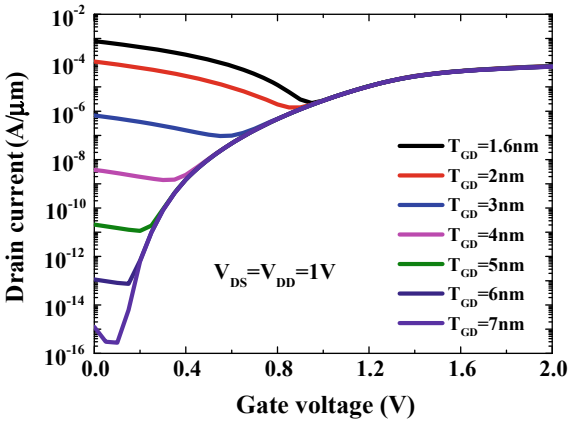


Fig. 6 Actual device transfer characteristics for different values of T_{GD}

that the subthreshold region of the actual device characteristics is getting enhanced with the gradual increase in T_{GD} value. Figure 7 shows the minimum subthreshold swing SS_{min} as a function of drain current for the UOESD-TFETs for three different T_{GD}-values (viz. 5-, 6-, and 7-nm, respectively).

After critically scrutinizing the plot, it is clear that sub-60 mV/decade value of SS_{min} is maintained for above 5 decades of drain current, only in the case of the device having $T_{GD} = 7$ nm. Thus, although the OFF-current drops down below the minimum measurable value, i.e., 1 pA/μm, for the device having $T_{GD} = 6$ nm, but the SS_{min} has failed to maintain its sub-60 mV/decade for 5 decades of drain current, which is one of the key requirements for utilizing any TFET structure in LSTP applications [15]. Different performance parameters have also been tabulated in Table 1 for the UOESD-TFETs for seven different T_{GD} values. It is clearly observed that with the increase in T_{GD}-value, I_{OFF} and hence, the OFF-state leakage power is decreasing by substantial amount. Moreover, a significant improvement in I_{ON}/I_{OFF} ratio is

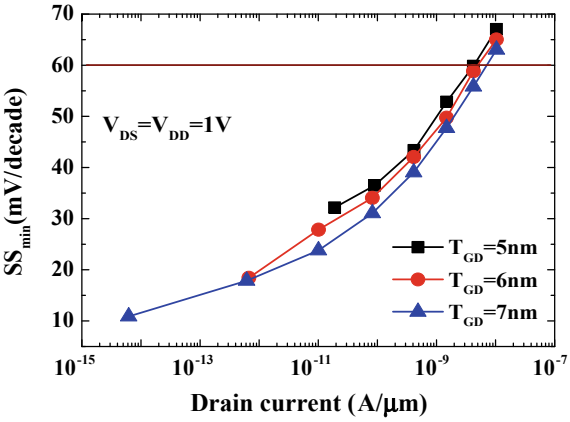


Fig. 7 Plot of minimum subthreshold swing SS_{min} as a function of drain current for the UOESD-TFETs for $T_{GD} = 5$, 6, and 7 nm

Table 1 Comparison of different performance parameters for devices with seven different T_{GD} values

T_{GD}(nm) SiO_2	I_{ON} (A/μm)	I_{OFF} (A/μm)	I_{ON}/I_{OFF}	SS_{min} (mV/decade)	$P_{leakage}$ (W/μm)	Delay (ns)
1.6	2.74×10^{-6}	7.7×10^{-4}	3.5×10^{-3}	132	7.7×10^{-4}	0.93
2	2.74×10^{-6}	1.1×10^{-4}	0.025	130	1.1×10^{-4}	0.83
3	2.74×10^{-6}	6.7×10^{-7}	4.1	103	6.7×10^{-7}	0.69
4	2.74×10^{-6}	3.7×10^{-9}	7.4×10^{2}	62	3.7×10^{-9}	0.62
5	2.74×10^{-6}	2.1×10^{-11}	1.3×10^{5}	32	2.1×10^{-11}	0.58
6	2.74×10^{-6}	1.1×10^{-13}	2.5×10^{7}	18	1.1×10^{-13}	0.55
7	2.74×10^{-6}	1.2×10^{-15}	2.3×10^{9}	10	1.2×10^{-15}	0.53

Table 2 Comparison of our work with Wang et al

	I_{ON}(A/μm)	I_{OFF}(A/μm)	I_{ON}/I_{OFF}	SS_{min}(mV/decade)
Wang et al	3.9×10^{-7}	5.33×10^{-16}	7.3×10^{8}	58
Our work	2.74×10^{-6}	1.2×10^{-15}	2.3×10^{9}	28

observed, which crosses the value of 10^5, which is another benchmark for a high-performance TFET device for LSTP operation [15], at $T_{GD} = 5$ nm. This I_{ON}/I_{OFF} has attained a value of 2.3×10^9 in case of the device with $T_{GD} = 7$ nm, claiming the superiority of the aforementioned device. In addition, SS_{min} and delay are also decreasing by considerable amount. In fact, 92.47% reduction in SS_{min} and 43% reduction in delay could be achieved, for the device with T_{GD}= 7 nm with respect to that of the device having $T_{GD} = 1.6$ nm (EOT value), as is witnessed from Table 2. Hence, from the point of view of LSTP applications, the device with $T_{GD} = 7$ nm is observed to satisfy the three most important criteria, viz. $I_{OFF} < 1$ pA/μm, sub-60 mV/decade SS_{min} for 5 decades of drain current, and $I_{ON}/I_{OFF} > 10^5$, together. Hence, UOESD-TFET with $T_{GD} = 7$ nm is considered as the ultimate optimized device. A comparison of our work with ref [6] is shown in Table 2.

5 Conclusion

In this paper, for the first time, a detailed investigation has performed on a UOESD-TFET device, considering the real-time adverse leakage phenomena, followed by proposing a novel design modification in terms of the optimization in the value of T_{GD}. It has been observed that for $T_{GD} = 7$ nm, our projected UOESD-TFET device

has produced a remarkably enhanced subthreshold characteristics, without affecting the super-threshold region performance, and thus becomes the optimized for LSTP operation.

References

1. Long P, Huang JZ, Povolotskyi M, Charles DVJ, Klimeck TKG, Rodwell MJW, Calhoun BH (2016) A tunnel FET design for high-current, 120 mV operation. In: IEEE international electron devices meeting (IEDM). San Francisco, CA, pp. 30.2.1–30.2.4
2. 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
3. Krishnamohan T, Kim D, Raghunathan S, Saraswat K (2008) Double-gate strained-Ge heterostructure tunneling FET (TFET) with record high drive currents and $\ll$60mV/dec subthreshold slope. In: IEEE International Electron Devices Meeting
4. Choi WY, Park BG, Lee JD, Liu TJK (2007) Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Device Lett 28(8):743–745
5. Kim SW, Kim JH, King Liu TJ, Choi WY, Park BG (2016) Demonstration of L shaped tunnel field effect transistors. IEEE Trans Electron Device 63(4):1774–1777
6. Chen S, Wang S, Liu H, Li W, Wang C, Wang X (2017) Symmetric U shaped Gate Tunnel Field Effect Transistor. IEEE Trans Electron Devices 64(3):1343–1349
7. Bhuwalka KK, Sedlmaier S, Ludsteck AK, Tolksdorf C, Schulze J, Eisele I (2004) Vertical tunnel field-effect transistor. IEEE Trans Electron Devices 51(2):279–282
8. Wang W, Wang PF, Zhang CM, Lin X, Liu XY, Sun QQ, Zhou P, Zhang DW (2014) Design of U-shape channel tunnel FETs With SiGe source regions. IEEE Trans Electron Devices 61(1):193–197
9. International Technology Roadmap for Semiconductors (2005)
10. ATLAS User's Manual (2015) A Device Simulation Package. SILVACO Int, Santa Clara, CA
11. Shen C, Yang LT, Toh E, Heng C, Samudra GS, Yeo Y (2009) A new robust non-local algorithm for band-to-band tunneling simulation and its application to Tunnel-FET. In: International Symposium on VLSI Technology, Systems, and Applications, Hsinchu, pp. 113–114
12. Iannaccone G, Curatola G, Fiori G (2004) Effective Bohm quantum potential for device simulation based on drift-diffusion and energy transport. SISPAD.
13. Slotboom JW (1977) The PN product in silicon. Solid State Electron 20:279–283
14. Selberherr S (1984) Analysis and Simulation of Semiconductor Devices. Springer-Verlag, Wien, New York
15. Ionesue AM, Riel H (2011) Tunnel field effect transistors as energy efficient electronics switches. Nature 479:329–337

Optimization, Design and Analysis of a MEMS Microphone with PVDF as a Structural Layer for Cochlear Implant Applications

Sudha R. Karbari, Shashank Jain, Shivam Gaur, M. Uttara Kumari, and G. Shireesha

Abstract The cochlear implants are the most advanced technology for hearing aid impairments. It consists of a microphone or sensor, speech processor, stimulator, and the electrodes. The speech processor in the implant uses the MEMS microphones to pick the signals and process. Here we have proposed a model of MEMS microphone with PVDF as a structural layer and aids in converting the received sound signal into electrical impulses. The proposed device modeling overcomes the main bottlenecks of CIs in terms of footprint, volume, mass, and stimulation signal. A simplified analytical model of PVDF membrane for a microphone application is optimized and developed. The PVDF membrane with its piezoelectric phenomenon acts as a frequency separator for a given acoustic pressure, and it undergoes vibration and accumulation of static charges on the surface of the material that is captured using appropriate signal conditioning circuits with charge amplifiers and filters. First, the geometry with optimized parameters using mathematical modeling was built using COMSOL Multiphysics. The optimization was carried out using predetermined values of thickness and width as the design parameters to form a structure for the dominant resonant frequency. The optimized model built-in geometry of COMSOL is analyzed to the next process that included meshing, setting the environment for the membrane to act as an acoustic model using appropriate multiphysics and perform the study using time-dependent variables and frequency domain. The membrane is analyzed for stress distribution along the walls and the center. Similar studies are carried out for PVDF membrane with a patch of PZT with optimized dimensions. A comparative analysis of the membrane for its acoustic property is to serve as a membrane in a MEMS microphone. A detailed analysis for a PVDF membrane is in terms of resonant frequency, stress, maximum displacement for the input acoustic pressure. This approach of simulating the optimized design for a structure to be a membrane in a

S. R. Karbari · S. Jain (✉) · S. Gaur · M. Uttara Kumari · G. Shireesha
Department of Electronic and Communication Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India
e-mail: shashankj492@gmail.com

Department of Physics, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India
e-mail: karbari.rsudha@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_37

MEMS acoustic sensor is validated, and thus, demonstrating it can be reliably used as a piezoelectric membrane in a MEMS process of a product.

Keywords Analytical Model · COMSOL Multiphysics · MEMS microphone · Cochlear Implants

1 Introduction

Microphones are mechanical, acoustic, electrical sensors, and transducers which convert the signals from acoustic form into electrical form. These types of sensors are utilized in various numbers of electronic applications. MEMS-based microphones are dominant in the market due to the reasons such as their compact dimensions and size, a very good SNR ratio, faster response, very good sensitivity, and their stability for longer durations. In the past decade, the MEMS microphones have acquired a significant market share in various consumer applications like hearing aids and mobile handsets. With the advent of newer technologies and consumer demand, the quality of audio recording is expected to be high along with microphones that have the ability to suppress maximum amount of noise. Hence, for this reason, MEMS microphones are extensively used. The microphone picks up the acoustic waves and forwards to the speech processor where the processor decomposes to eight overlapping sub-bands and transmitted to an electrode that is surgically implanted into the cochlea of a deaf person in such a way that they can stimulate the appropriate region in the cochlea for the frequency they are transmitting. Here the proposed model with piezoelectric layer of PVDF and PVDF + PZT increases the spatial frequency band with reduced footprint and mass less than 20 mg.

The microphone picks up the acoustic waves and forwards to the speech processor where the processor decomposes to eight overlapping sub-bands and transmitted to an electrode that is surgically implanted into the cochlea of a deaf person in such a way that they can stimulate the appropriate region in the cochlea for the frequency they are transmitting. Here the proposed model with piezoelectric layer of PVDF and PVDF + PZT increases the spatial frequency band with reduced footprint and mass less than 20 mg. The detection method for the commercially available MEMS microphone is by varying the capacitance and measuring it for the input acoustic vibrations. The electrodes micromachined using silicon technology for a capacitive-based includes the design that facilitates the changes in acoustic vibration to specific variations in capacitance. Along with the packaging, the dedicated read-out with analog or digital output has dimensions of few millimeters. The typical specifications of a MEMS microphone used in cochlear implants are as in Table 1.

Weigold et al. analog devices fabricated successfully a spherical moving membrane of polysilicon and SCS with perforations on the back plate to reduce the stress of an SOI wafers. A spring suspension diaphragm with suspensions at the **co**rners proposed **b**y analog devices leads to increase in the sensitivity of the device

Table 1 Specifications of MEMS microphones available in market [2]

Manufacturer	Model	Sensitivity [dBFS/1Pa]	Band width [kHz]	SNR [dBA]	MAX SPL [dB]	Current Supply [µA]	Size [mm^3]
Knowels	SPK0833LM4 HB	−26	63	4–6	122	500	4×3×1
AAC	SDM0401B-263-M02	−26	60.5	18–24	No data	550	4×3×1
Analog devices	ADMP521	−26	65	2.5–3	120	900	4×3×1
Goer Tek	SD040T263-01/02	−26	38	−33	No data	600	4.7×3.8×1.3
STMicroelectronics	MP34DB01	−26	62.8	0.106	120	650	4×3×1

by 4% from the existing MEMS microphones that are a part of cochlear implants. [1].

Leoppert and drive [1] proposed a polysilicon membrane with 0.6 mm diameter and 4 µm thickness contributing to a chip size of 1 mm^2 with a frequency response of 12 kHz and SNR of 58 dB with improved sensitivity. The device tested successfully with a peak resonance frequency of around 14 kHz and bias voltage of 12 V. Designs from Infineon technologies improved the SNR to 66 dB and a sensitivity of 38 dBV/Pa at 1 kHz for a 1.1 mm membrane diameter. Dehe reported the employment of springs as corrugation rings leading to improving the sensitivity compared to the standard clamped membrane.

A combination of surface and bulk micromachining technologies to minimize the residual stress which preloads the membrane and affects the sensitivity of the MEMS microphone during etching process for both the technologies employed. The release of structure through sacrificial etching plays a major role in controlling the process parameters and affects the crystal lattice structure. These stresses can be reduced by the use of interlayers and annealing steps [2].

STMicroelectronics microphone design found in iPhones is fabricated with a polysilicon membrane with four corners anchored provided SNR values ranging from 61 to 65 dB. MEMS microphones with rigid backplate and flexible membrane that deflects out of plane to reduce the induced stress. Metalized silicon nitride [1–3], the combination of silicon nitride and polysilicon [3–7] as membrane is reported. The membrane surface area ranges from 0.38 to 4.41 mm^2 to a maximum value [7]. In the present work, the various materials for a MEMS microphone with flexible membrane and an appropriate readout to reduce the area of the chip fabrication is explored.

The most commonly used materials for piezoelectricity include zinc oxide, aluminum nitrite, and lead zirconium titanate. Both ZnO and ALN are non-ferroelectric, whereas PZT is ferroelectric with its direction of polarization reversed by applying appropriate electric field. As the flexibility of the materials is concerned PVDF is an apt choice. The materials for the study are piezoelectric as the output

does not require dedicated electronics. The change in input acoustic vibration leads to changes in voltage for that specific frequencies along with stability.

The polyvinylidene fluoride (PVDF) material is a homopolymer of the compound 1,1-difluoroethane. The material shows a high insolubility property along with good electrical characteristics.

Early research on piezoelectric microphones exclusively with a plastic form based on bimorph diaphragm has been done in 1981. A microphone design for telephone handset which is operated by pressure is put forth which uses designs that are symmetrical. The noise-canceling units for pressure gradient are realized for first as well as second order, and they are compared with the then available microphones. Piezoelectric behavior is exhibited by certain plastics although it is observed that the PVDF material shows the highest activity to date as reported in the year 1969 by Kawai [8].

A piezoelectric microphone is realized and simulated in [9] by making use of SU-8 membranes and silicon as the materials separately. The PZT-5H or lead zirconium titanate is applied as the sensor material. The properties of standalone silicon and the SU-8 polymer membrane have been determined by subjecting them to piezoelectric analysis with PZT patches.

A demonstration of MEMS microphone is with piezoelectric Parylene-C [10]. It has been designed using a diaphragm of 6 mm across its diameter and 30 μm in its thickness. The Parylene-C material is considered due to its chemical inertness, mechanical characteristics and stability in high temperatures. It also has an added advantage of biocompatibility.

A multi-band directional MEMS microphone which is bio-inspired relying on the properties of hearing of the fly Ormia ochracea is put forth in [11]. The microphone is multi-user and is foundry fabricated works in the four bands of frequencies which are all within 10 kHz. It operates like a directional microphone with pressure gradient which yields a polar pattern of eight. It can also behave as a microphone that is omnidirectional based on the housing involved [12]. The simulations are carried out in order to determine the deflection of the membrane, the pull-in voltage, capacitance, and sensitivity of the microphone.

A piezoelectric transduction including a thin film of aluminum nitride composite as diaphragm in an array form is developed for aircraft noise source identification. This array aids in aircraft manufacturing needs of full-scale flight tests [13]. Parametric studies are carried out in terms of sensitivity and frequency response on the overall microphone. Thus, improving the sensitivity is 8 times greater than the conventional with lower output voltage [14]. The circular diaphragm optimization is for a point inflection in the sensitivity function with respect to radius [15]. An inexpensive high-sensitivity microphone with cross-hair-shaped PVDF film obtained by simple chemical etching is fabricated, and pressure amplification for area ratio along with a mix of the reduced capacitance of the top electrode that is shaped in a similar way is analyzed. The sensitivity gain thus obtained is near to the area ratio of 3.2 [16–19]. There are currently many ways to realize this process optimization for polymer-based piezoelectric sensors, but the optimization arrangements and design of experiments play a major role on its influence in piezoproperty of the material and the process for the fabrication of MEMS device with polymer as a structural

layer. Once the parameters are evaluated using FEA tool for its displacement based on the piezoproperty, conclusions are drawn on the final part geometric designs. This approach yields satisfying results when fabricated as devices. This study and analysis are conducted in 5 phases and analyze the properties of the PVDF membrane when a pressure of 1 Pa is applied at the center of the membrane alone and also with a PZT-5H patch placed in the middle. Phase 2 describes the structure of the microphone. Phase 3 concentrates on the simulation of the membranes involved. Phase 4 concludes the paper, and phase 5 gives an insight into the future work that can be carried out.

2 Materials and Methods of Fabrication

Polymer, polyvinylidene fluoride (PVDF), of long chains of the repeating monomer (—CH2—CF2—) *i*th hydrogen atoms positively charged and the fluorine atoms negatively charged. The flexible polymer PVDF piezoelectric layer is used as a structural layer with the process which is discussed in the earlier section. The previous studies concluded on the traditional structure and the process that is followed for the fabrication of the device. There are piezomaterials that serve as a membrane in the design, and our approach is to analyze the feasibility of PVDF, a flexible polymer as a structural layer the fabrication. Polyvinylidene fluoride (PVDF) has strong piezoelectric and ferroelectric properties. But the processing of piezoelectric material requires poling and stretching. The piezoelectric coefficient of poled and stretched thin films ranges around 6–7 pC/N [18]. To imprint the appropriate design onto PVDF flexible polymer layer, a photomask design with appropriate processing parameters can be fabricated [19].

Techniques to fabricate the nano-composites include solvent-casting, spin-coating, and hot-embossing [19]. Polyvinylidene fluoride exists in different phases of α, β, γ, and δ phases; among them, β phase is of greater importance because of polar crystalline electroactive nature. Thus, the nucleation of polar β phase in PVDF is an essential parameter for its application as structural layer in MEMS microphone for cochlear implant applications. In this context, complex membrane fabrication process is tailored to adhere thermal, mechanical, and electrical properties [20]. Piezoelectric sensors can measure vibrations on the basis of the direct effect and conversion of energy to electrical output for a function of resonant frequency. PVDF as piezoelectric sensors for vibration measurements owing to its advantages includes outstanding chemical resistance, high thermal stability, low permittivity, low acoustic impedances [21–24]. So, PVDF sensors have been applied to vibration measurements in a greater extent, which is the focus of this paper. In the present work, the feasibility of the prototype devices as implantable microphones as totally implantable cochlear implants is investigated [25].

Table 2 Material properties of PVDF and PZT

Material	PZT-5H	PVDF
Youngs modulus(GPa)	71	4–6
d_{31}(pC/N)	−274	18 −24
d_{32}(pC/N)	−274	2.5−3
d_{33}(pC/N)	593	−33
e_{33}(pC/N)	30.1	0.106

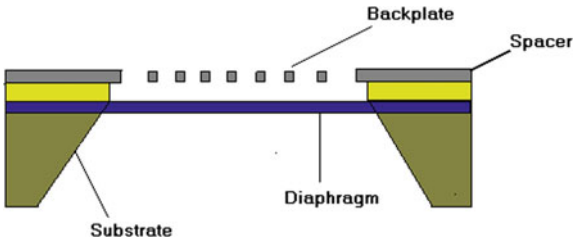


Fig. 1 Schematic representation of the melt pool and the corresponding geometric characteristics [16]

2.1 Microphone Structure

The designed sensor uses MEMS bulk micromachining technology with sacrificial etching and consists of a silicon capacitor with a merged two plates: one of the plates fixed and the other movable. The fixed surface of the designed sensor is conductive and is covered by electrode with perforated holes that allow the acoustic input to pass. Pressure is enforced on to the membrane, the vibration results in accumulation of charges on the PVDF membrane and captured along the patch of PZT-5H. The microphone developed using MEMS technology is an apt device to fit in a cochlea where the transduction principle is a coupled capacitance change between a fixed and a movable piezoelectric plate for the given input frequency. The microphone consists of a pair of gold electrodes which are used in order to sense the amount of voltage produced. An adhesive material such as chromium is used between the gold electrode and PZT-5H material. Here, the microphone membrane is simulated with PVDF as structural layer, and its feasibility for the application is validated using simulations from COMSOL Multiphysics (Table 2; Fig. 1).

2.2 Analysis and Simulation Results

The analytical part of the study is conducted using FEA tool COMSOL Multiphysics with structural mechanics as physics chosen and eigenfrequency as study. The results are interpreted using appropriate plots from the tool, and the conclusions are drawn. The parameters required for the simulation and the boundary conditions for the

material property of piezoelectricity, constitutive equations are

$$\mathrm{Di} = \mathrm{e}\sigma\ \mathrm{ij}\,\mathrm{Ej} + \mathrm{ddim}\,\sigma\,\mathrm{m} \tag{1}$$

$$\in \mathrm{k} = \mathrm{dc\ jk}\ \ \mathrm{Ej} + \mathrm{SEkm}\ \sigma\,\mathrm{m} \tag{2}$$

The piezoelectric materials used as a sensor from its direct effect are focused toward combining the two types of piezoelectric materials, PZT and PVDF. The ceramics are fabricated by mixing together proportional amounts of lead, zirconium and titanium oxide powders and heating the mixture at around 800–1000 °C. The materials isotropic in nature and by poling process become transversely isotropic in the plane normal to the poling direction, henceforth tailoring the piezoelectric properties of the layers in the device fabrication for it to be used as a membrane in cochlear implants. The characteristics of PZT and PVDF for setting the environment in COMSOL are in Table 1. Young's modulus of the PZT membrane is comparable to aluminum; whereas for PVDF, it is 1/12th of aluminum. The analysis of sandwitched two piezoelectric materials PVDF and PZT as a piezoelectric structural membrane is sandwiched, and its effect on the sensor design is simulated in accordance with the. Here in the present work, a combination of PVDF and PZT as a piezoelectric layer for acoustic device is proposed. The different results with the comparison of PVDF piezomembrane with PZT behaving as a patch and without the patch of PVDF-5H are compared (Fig. 2).

Various analyses in terms of stress, displacement, and polarization from COMSOL are interpreted. The dimensions and structure for the device with PVDF as a membrane that is used for both the studies are shown below. The dimensions of PVDF chosen are 1 mm × 1 mm device for it to perform its functionality of an acoustic membrane. A free tetrahedral mesh is constructed on the membrane, and the stress analysis is carried out (Table 3).

The four edges of the membrane are given as the fixed constraints for the structure. The frequencies with different modes of vibration along the materials are considered in working of the microphone as a sensor. The pressure from the sound signal is converted into the change in voltage which occurring at resonant frequencies with

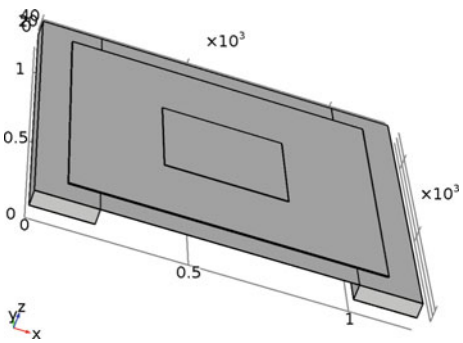


Fig. 2 Schematic geometry of MEMS microphone [16]

Table 3 Dimensions of the PVDF as a MEMS acoustic sensor

Material	Thickness	Area in μm²
Au	0.5	1000 x 1000
Cr	0.01	1000 x 1000
Si	10	1000 x 1000
PVDF	4	1000 x 1000
SiO2	1	400 x 400
Si Substrate	1	1200 x 1200

the maximum output voltage provided. As there are 6 modes of operation for the resonant frequencies, six plots and the maximum displacement occur at the first resonant mode. The eigenfrequencies range from 6955.1 to 26,766 Hz. The variation of the displacement of the membrane is observed along the length with a value of 19 μm at the center of the membrane.

The electric potential conveys the charge conversion for the piezoelectric material including the various changes in the frequencies. The optimized structural dimensions provided with the conversion efficiency are a validation for the piezoelectric materials generating charge and the values ranging from 2.4 to 5 V.

The von Misses stress for the membrane for all the 6 resonant frequencies along with the arc length from 0 along the vertex to 1000 micron is displayed. Maximum stress of 8.5 × 106 N/m^2 is for the frequency of 26.766 kHz.

The kinetic energy density in the given membrane is about 4100 J/m^3 at 500 μm. The acceleration along with the *Z*-component in Fig. 4 is found to be 5 × 10^5 m/s^2 at 500 μm. The change is experienced in electric polarization which is observed in some materials when mechanical stresses are applied on them. The material PVDF is a piezoelectric material and experiences piezoelectric effect only in the beta phase of crystallinity of the material. The piezoelectric polarization as obtained along the *Y*-direction is 7 × 10^{-5} C/m^2. The other modes obtain a very drastic and slow acceleration along the *Z*-component with the values ranging from 0 to 0.5 m/s^2.

The arc length is chosen across the cross-sectional area along the *X*-axis. The strain energy stored in an elastic material upon deformation is elastic strain energy. It is found to have a maximum value of 11 kJ/m^3 at 500 μm. Maximum displacement of about 19 μm is observed in the central area. The first modal frequency designates

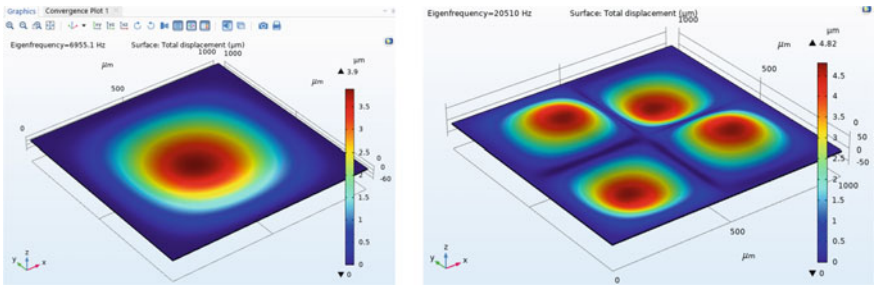


Fig. 3 Total displacement and resonant frequency

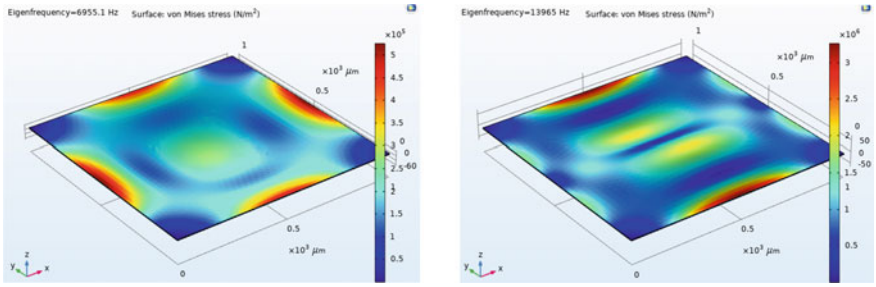


Fig. 4 Von Mises stress with PVDF as membrane

the dominant frequency for PVDF in beta phase which gives the d33 coefficients and the polarization direction is along with the *Z*-axis occurring at 6.9551 kHz.

3 Simulations with PZT Patch

3.1 Analysis and Simulation Result

To establish the relationship between the previously defined structural analysis for a PVDF membrane with input as pressure acoustic, the comparative study of PVDF with a patch is provided.

A patch PZT-5H is placed at the center of the membrane. The mesh is dense at the patch and gradually becomes less dense toward the membrane boundaries. The mesh statistics include average element quality of 0.2839, and the number of tetrahedral elements includes 3896. With PZT patch, there is a variation in the frequencies, with the first modal frequencies increasing from 6951 to 8445 Hz and the last modal frequency varying from 26,766 to 47,591 Hz. The increase in the frequency ranges from a PVDF membrane to a PVDF +PZT patch and concludes that the response of the system increases drastically. The experimental and analytical data confirm the piezoelectric property by the capacity of the material to convert the vibrations into voltage through surface charge density. This is facilitated by the polarization of the dipoles formed during the crystalline nature, processing of the material. The computational analysis considering the compliance and permittivity equations fed to FEA tool COMSOL Multiphysics for setting the parameters is to behave as a piezoelectric layer. For given resonant frequencies, variations in the output voltages are observed with 2 V to be maximum. For a material irrespective of its property of piezoelectricity, stress and strain analysis is a very crucial parameter that helps in tailoring the device parameters at fabrication level. von Mises stress for eigenfrequencies is plotted. The values of displacement for the diaphragm with PZT-5H at the center result in reduction of the displacement from micron to 1.2 nm. It is

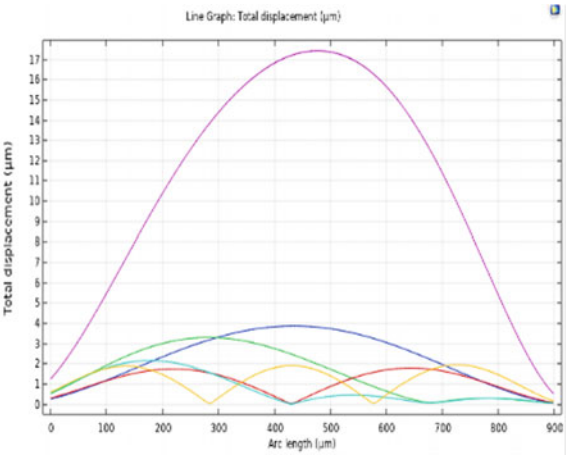


Fig. 5 Von Mises stress along the arc length

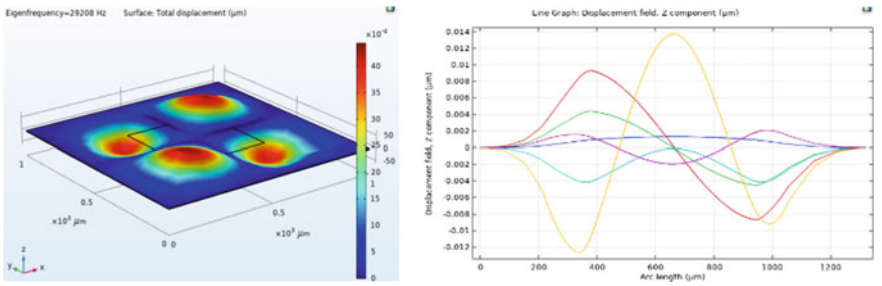


Fig. 6 Total displacement and von Mises stress with PZT as patch

found to be maximum at the center of the membrane. The displacement is found to be 1.3 nm near 500 μm range of the membrane.

von Mises stress variation along with the arc length of 1000 μm for the two modal frequencies is plotted, and they are maximum at 500 and 700 μm contributing to a value of 600 N/m^2 and 700 N/m^2, respectively. For the first modal frequency, the acceleration along *Z*-direction is maximum at 200 μm and attains its minimum value at 750 μm. Polarization is maximum at the center giving values of 4.5×10^{-8} C/m^2, 1.6×10^{-7} C/m^2, and 0.9×10^{-7} C/m^2 for *X*-, *Y*-, and *Z*-directions, respectively, that indicate the membrane exhibits significant amount of piezoelectricity between 400 and 500 μm. The beta coefficients can be improved by incorporating the dopants of appropriate material properties for a PVDF membrane.

Similarly, the beta coefficients along the Y-component are 1.6×10^{-7} C/m^2. This is for a dominant resonant as discussed in previous results. Similarly, the beta coefficients along the *Z*-component are 0.8×10^{-7} C/m^2. This is for a dominant resonant as discussed in the previous results. The refining and calibration using appropriate equations and the described approach can be used to build more detailed processing parameters during fabrication.

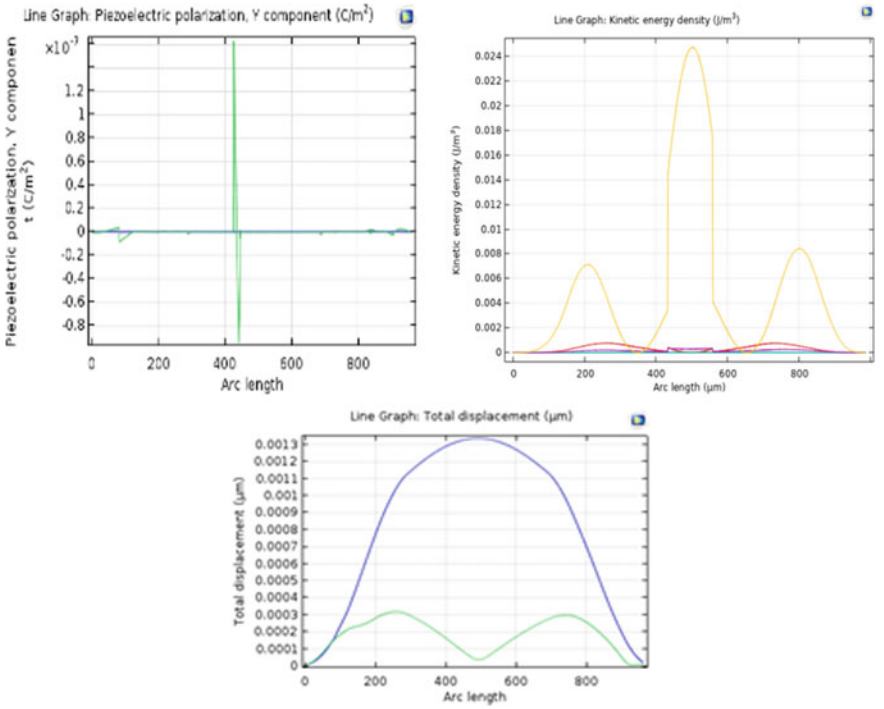


Fig. 7 Plot of polarization, acceleration, and displacement

3.2 Validation Strategy for the Proposed Methodology for Device Fabrication as MEMS Acoustic Sensor or Microphone

The results of the optimization process using COMSOL Multiphysics with the mathematical and analysis data in terms of resonant frequency, stress, displacement for the resonant frequency and electric potential indicate that the charge conversion capacity is increased and concentrated toward the center for the previous studies.

The behavior of the microphone with PVDF membrane and also with a PZT patch in the middle is simulated and analyzed. The results are observed for stress, displacement, and piezoelectric polarization. The feasibility of choosing polymers as the membrane materials for the fabrication of the MEMS microphones is analyzed. Optimization of the PVDF piezoelectric material as a membrane is performed, and the results are significantly promising in terms of the displacement, stress, and piezoelectric polarization. Hence, we can conclude that PVDF is a good material for a MEMS microphone.

References

1. Leinenbach, C., Van Teeffelen, K., Laermer, F., Seidel, H.: A new capacitive type MEMS microphone. In: 2010 IEEE 23rd International Conference on Micro Electro Mechanical Systems (MEMS), pp. 659–662, Wanchai, Hong Kong, China (2010)
2. Grubb JA (1982) Piezoelectric plastic microphones. Electro-component. Sci. Technol. 9:197–203
3. Loeppert, P.V.: The first commercialized MEMS microphone. In: Solid-State Sensors, Actuators, and Microsystems Workshop, vol. 2006 (2006)
4. Dehé, A., Wurzer, M., Füldner, M., Krumbein, U.: Design of a poly silicon MEMS microphone for high signal-to-noise ratio. In: 2013 Proceedings of the European Solid-State Device Research Conference (ESSDERC), pp. 292–295. Bucharest, Romania (2013)
5. Dixon-Warren, S.: Survey of MEMS microphone technology 2016, July 2018, http://www.chipworks.com/about-chipworks/overview/blog/survey-of-mems-microphone-technology
6. AN4426 application note tutorial for MEMS microphones." Tech. Rep., Tech. Rep., ST Microelectronics (2017)
7. A MEMS condenser microphone based acoustic receiver for totally implantable cochlear implants, Department of Otorhinolaryngology, Head and Neck Surgery
8. Muralidhar, Y.C., Somesh, B.S., Neethu, K.N., Yeshashwini, L.R., Veda S.N., Pinjare, S.L.: Design And Simulation Of Silicon And Polymer Based Piezoelectric MEMS Microphone. In: International Conference on Emerging Trends in Communication, Control, Signal Processing and Computing Applications (C2SPCA), pp. 1–6 (2013)
9. Kim, J.Y.H., Liu, Y., Scianmarello, N., Tai, Y.C.: Piezoelectric Parylene-C Mems Microphone. In: Transducers & Eurosensors XXVII: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems, pp. 39–42 (2013)
10. Bauer, R., Zhang, Y., Jackson, J.C., Whitmer, W.M., Brimijoin, W.M., Akeroyd, M., Uttamchandani, D., Windmill, J.F.C: Housing influence on multi-band directional MEMS microphones inspired by Ormiaochracea. IEEE Sensors, pp. 1–3 (2016)
11. Cezary, M., Wojciech, Z., Andrzej, N.: "Parameterized model of MEMS microphone in COMSOL multiphysics. In: 14th International Conference The Experience of Designing and Application of CAD systems in Microelectronics (CADSM), pp. 428–431 (2017)
12. Matthew D. Williams, Benjamin A. Griffin, Tiffany N Reagan, James R. Underbrink and Mark Sheplak, "An AlN MEMS Piezoelectric Microphone for Aeroacoustic Applications", Journal of Micromechanical Systems, pp. 270–283 (2012)
13. Schrag, G., Reutter, T., Wachutka, G.: Finite element simulations combined with statistical analysis of variance: exploring the potential of piezoelectric MEMS microphones. In: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems, pp. 1452–1455 (2013)
14. Song, K., Zhan, W., Xu, W., Lee, Y.-L.: Scaling analysis of capacitive MEMS microphones considering residual stress. In: 11th IEEE Annual International Conference on Nano/Micro Engineered and Molecular Systems (NEMS), pp. 184–187 (2016)
15. Xu, J., Headings, L.M., Dapino, M.J.: High sensitivity polyvinylidene fluoride microphone based on area ratio amplification and minimal capacitance. IEEE Sensors J. 2839–2847 (2015)
16. Liu, W., Han, M., Sun, X., Zhang, H.: Fabrication of spiral-shaped PVDF cantilever based vibration energy harvester. In: The 9th IEEE International Conference on Nano/Micro Engineered and Molecular Systems (NEMS), pp. 458–461. Waikiki Beach, HI https://doi.org/10.1109/nems.2014.6908849
17. Han, H., Nakagawa, Y., Takai, Y., Kikuchi, K., Tsuchitani, S.: PVDF film micro fabrication for the robotics skin sensor having flexibility and high sensitivity. In: 2011 Fifth International Conference on Sensing Technology, pp. 603–606. Palmerston North https://doi.org/10.1109/icsenst.2011.6137052
18. Kim, H., Torres, F., Mingyue, L., Yirong, L., Tzu-Liang B.T.: Fabrication and Characterization of 3D Printed BaTiO3/PVDF Nanocomposites. J Compos Mater **52**(2), 197–206 (2018). https://doi.org/10.1177/0021998317704709

19. Kang, G.D., Cao, Y.M.: Application and modification of poly(vinylidene fluoride) (PVDF) membranes—A review. J. Membr. Sci. **463**, 145–165 (2014)
20. Lang SB, Muensit S (2006) Review of some lesser-known applications of piezoelectric and pyroelectric polymers. Appl Phys A 85:125–134
21. Jia, Y.F., Chen, X.J., Sh, Q., Ni, L.H., Ju, L.I.: Dependence of the impact response of polyvinylidene fluoride sensors on their supporting materials elasticity. Sensors **13**, 8669–8678 (2013)
22. Park S, Guan X, Kim Y, Creighton F, Wei E, Kymissis IJ, Nakajima HH, Olson ES (2018) PVDF-based piezoelectric microphone for sound detection inside the cochlea: toward totally implantable cochlear implants. Trends Hearing 22:2331216518774450
23. Daniel, Z., Karl-Ludwig, K.: Classification and simulation method for piezoelectric PVDF sensors
24. Weigold, J.W., Brosnihan, T.J., Bergeron, J., Zhang, X.: A MEMS condenser microphone for consumer applications. In: 19th IEEE International Conference on Micro Electro Mechanical Systems, pp. 86–89. Istanbul, Turkey (2006). View at Publisher. View at Google Scholar
25. Goto, M., Iguchi, Y., Ono, K., et al.: High-performance condenser microphone with single-crystalline silicon diaphragm and backplate. IEEE Sensors J. **7**(1). 4–10 (2007). View at Publisher.·View at Google Scholar.·View at Scopus

Magnetic Resonance Imaging Classification Methods: A Review

Nikhil Pateria, Dilip Kumar, and Sunil Kumar

Abstract Magnetic resonance imaging (MRI) is one of the most important medical diagnosis methods in the field of computer-aided detection of medical images. The MRI images help to find the presence of abnormal cells or tissues, referred as tumors. Prior to classification, preprocessing of MRI images is performed. These preprocessing operations help to reduce undesired distortions and select only relevant features for further analysis, making the classification technique more accurate and efficient. In this paper, we present various image classification techniques used over MRI images and their performance.

Keywords Magnetic resonance imaging (MRI) · Image processing · Supervised image classification · Unsupervised image classification · Artificial neural networks (ANN)

1 1. Introduction

A. *Medical Resonance Imaging*

MRI is a diagnosis tool for capturing images of internal body parts and tissues in detail [1]. These images are used for the diagnosis of various diseases. MRI image processing is the most challenging and innovative field. There are other techniques available for medical imaging, namely mammography, CT scan, and x-ray, but the MRI imaging outperforms these techniques in quality and detailing. It is best suited

N. Pateria (✉) · D. Kumar · S. Kumar
Department of Computer Science & Engineering, NIT Jamshedpur, Jamshedpur, India
e-mail: nik.sati29@gmail.com

D. Kumar
e-mail: dilip.cse@nitjsr.ac.in

S. Kumar
e-mail: 2018rscs016@nitjsr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_38

for soft tissue abnormality detection. MRI scan is useful in diagnosing brain tumors, brain injuries, diseases like Alzheimer's, epilepsy, arthritis, bone infection, breast cancer, prostate, etc. [2]. Traditionally, human inspection was the only method used for the classification of MRI images. This method was impractical as MRI images contain noise due to improper operator assistance, and this noise leads to inaccuracies in classification of images. The MRI image data is large and complex. For appropriate diagnosis of MRI images, we have to preprocess these images and extract only relevant features by applying various feature reduction and feature selection techniques. In some emergency cases, diagnosis with wrong result or delay in accurate diagnosis may be fatal. MR images play an important role in human brain research [3]. MR imaging comes under clinical radiology and is one of the most effective medical imaging techniques. The classification of MR images can be done by using supervised techniques as well as unsupervised techniques.

B. *MRI Image Processing*

In MRI image processing, mainly the following steps are essential:

- Image acquisition
- Image preprocessing
- Image feature extraction
- Image feature reduction
- Image classification

Image acquisition [4, 5] is the first step; the MRI image will be obtained through MRI scanner, and researchers may acquire MRI images from publically available databases or from the radiologists.

Image preprocessing involves two steps; the first is noise removal and the second is image enhancement [6]. The process of noise removal is carried out by applying filters such as Laplacian filter [4] and median filter. This process is also known as image smoothing, as it removes unwanted distortions and irrelevant features. Image enhancement improves the relevant characteristics of image by modifying the image attributes making it more suitable for further analysis.

Image feature extraction is a major step involved in MRI image processing. Feature extraction helps to find relevant information contained in image so that process of classification becomes easier.

In most of the cases, the image input data becomes very large to process, and there is always large amount of redundant data [7], so before processing, this data is converted into a reduced form of image features. These image features are known as feature vectors [8]. An image feature defines a function to specify various properties of an object. The features may be general or domain specific features [9]. So, by feature extraction, input data is converted to set of feature vectors. Some widely used methods for feature extraction are discrete wavelet transforms (DWT), zoning, Fourier descriptors, gradient feature, spline curve approximation, etc. For MR images, DWT is preferred as it is a powerful tool for feature extraction, and it obtains wavelet coefficients from MR images to form feature vectors. DWT provides localized frequency [10] information, hence making classification efficient.

Image feature reduction is done after feature extraction; the feature vector dataset has to be minimized from large to most relevant features [11]. Feature reduction is applied to reduce overall processing time as well as cost of the whole system, because only reduced and relevant data will be processed. The most widely used feature reduction method is principal component analysis (PCA). PCA obtains the linear lower dimensional representation of data such that the reconstructed data has its variance preserved. PCA transforms the input feature data into a lower dimensional feature data using large eigenvectors of correlation matrix. So, PCA makes it easy for classification algorithm to perform efficiently.

The process of grouping images, regions or pixels composing the images into multiple classes on the basis of some similarities in characteristics is known as image classification [1]. It is the final step of MRI image processing. Image classification can be broadly done in two ways:

- Supervised Classification
- Unsupervised Classification

The supervised classification of images can be done at image level, region or object level and at pixel level. When supervised classification is done at image level, the classifier is trained with various image categories, and then, it is able to classify the images which were not present in the training data. Figure 1 illustrates the supervised classification at image level. The training set is divided into five categories. The classifier goes in training using the dataset and becomes ready to classify the unseen images. At pixel level, the classifier categorizes the pixels of unknown category into some known category. Here, the classifier is trained on training sets consisting of pixel categories. Figure 2 illustrates the supervised classification at pixel level. Here,

Fig. 1 Supervised classification at image level [1]

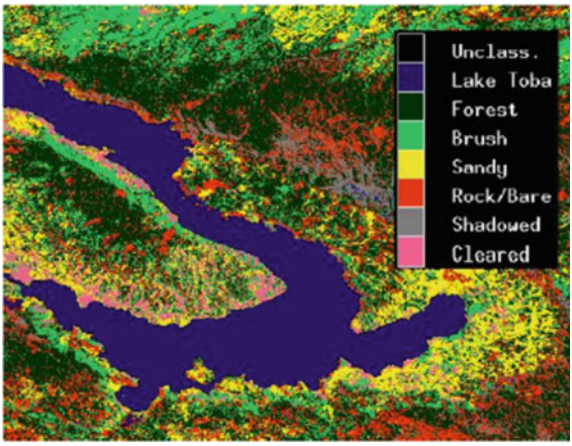


Fig. 2 Supervised classification at pixel level [1]

the pixels are categorized in seven classes of land cover areas. At object level, it finds out the regions inside the image differentiated by some characteristics and hence represents various objects in the image defined by regions. On the basis of these regions, supervised classification is then performed.

The unsupervised classification is also known as clustering [1]. At image level, the classification is done on the images of unknown categories, the set of images with common characteristics defines a class or cluster. Figure 3 illustrates unsupervised image-level classification where the group of twenty images forms five different clusters. The images of same semantics belong to same cluster. The unsupervised pixel-level classification groups the image pixels on the basis of some common characteristics, forming uniform regions of images. Figure 4 illustrates unsupervised classification at pixel level, known as segmentation. Pixels of same color and texture are categorized in same cluster.

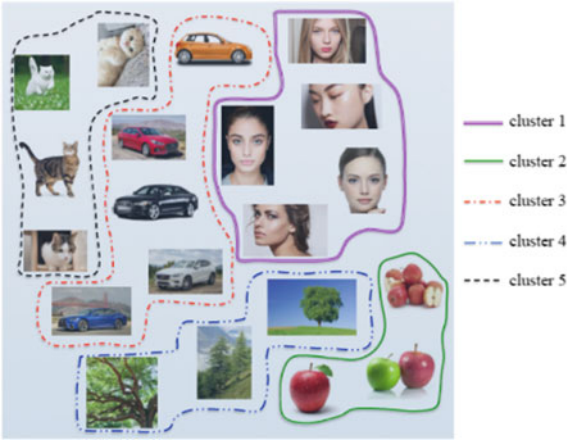


Fig. 3 Unsupervised classification at image level [1]

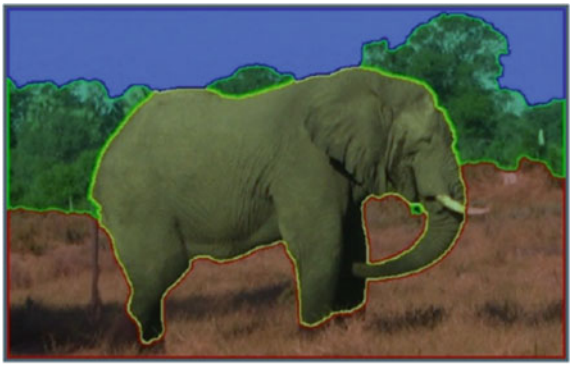


Fig. 4 Unsupervised classification at pixel level [1]

2 MRI Image Classification Methods

The MRI image classification can be done in supervised manner as well as in unsupervised manner [12]. The process of classification of MR images is challenging and can be achieved various ways. The supervised techniques mainly used are artificial neural network (ANN) based on back-propagation neural network (BPNN), convolutional neural network (CNN), probabilistic neural network (PNN), and apart from ANN, there are support vector machine (SVM) and k-nearest neighbor (KNN) which are most preferred. Figure 5 illustrates the comparison of normal brain image and abnormal brain image. The second image clearly depicts brain tumor [13]. MRI image classification can be used in detection of large variety of diseases; Figure 6 illustrates various MRI patterns to identify diseases on the basis of white spot patterns seen in MRI images, like small strokes, multiple sclerosis, brain tumor, infection, and migraine, lupus, and B12 deficiency.

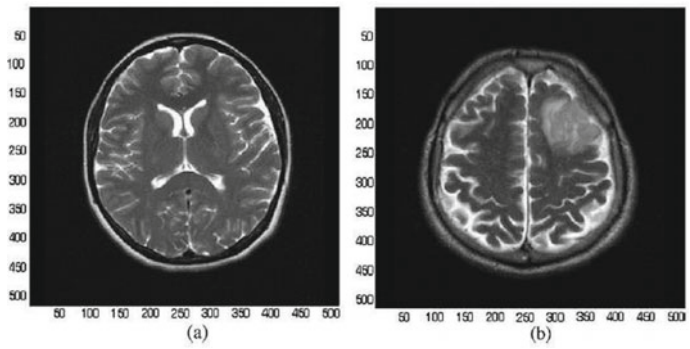


Fig. 5 **a** Normal brain MRI image, **b** Abnormal brain MRI image [13]

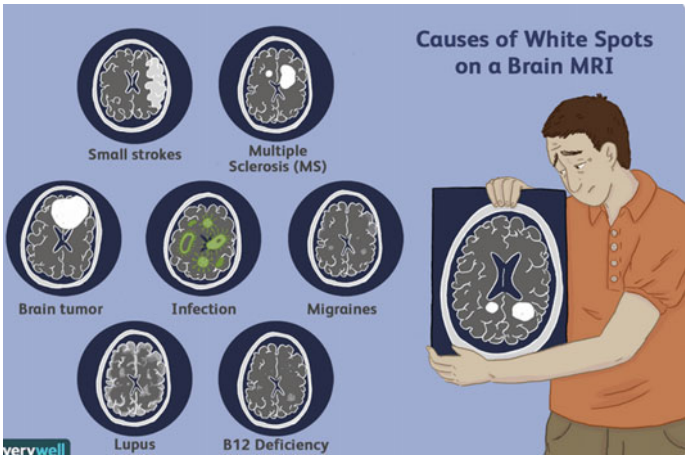


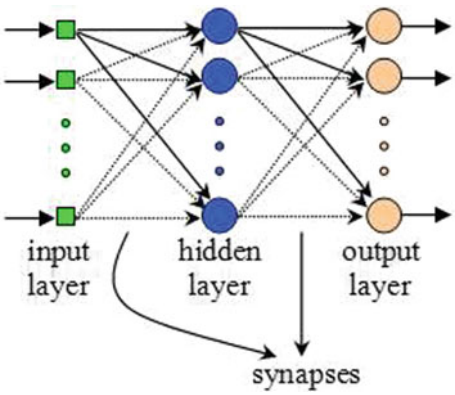
Fig. 6 Various disease diagnosis with the help of MRI image analysis

A. *Supervised Classification Methods*

1. Back-Propagation Neural Network

ANN-based classifiers are the most preferred for the classification of MR images. The ANN classifier consists of massively interconnected processing elements. The elements and layers are organized such that the whole architecture resembles the human brain. An ANN seems to be massively parallel distributed processor. Figure 7 illustrates a simple ANN having three layers with feed-forward connections. The most frequently used training algorithm is BPNN [10, 14, 12]. The signals are sent forward by the neurons, while the errors are sent backwards. The process continues until the ANN learns the training data. The BPNN is supervised learning, and its main aim is to reduce the error. The BPNNs are useful for brain tumor detection from MRI brain images.

Fig. 7 Simple ANN architecture



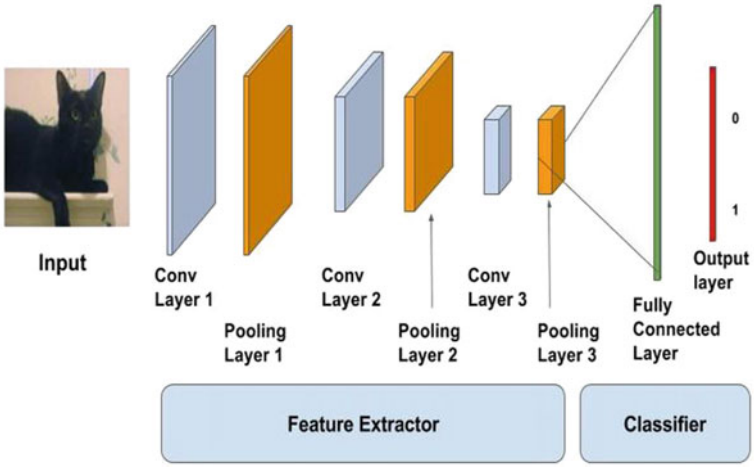


Fig. 8 Simple CNN architecture

2. Convolutional Neural Network

CNNs [15, 16] are regularized multilayer perceptron; these are fully connected neurons where neurons of two consecutive layers are completely interconnected. CNN performs regularization with the help of hierarchical data pattern and assembling complex patterns using smaller and simpler patterns. Figure 8 illustrates a simple CNN architecture consisting of input layer, a series of three combination of convolution and pooling layers, a fully connected layer and the output layer. A series of convolution layer and pooling layer can be used as per the requirement.

The input to CNN is informed of one-dimensional or two-dimensional matrixes of data. A fully connected layer acts as multilayer perceptron (MLP). CNN is mostly preferred for supervised classification of MRI images, as it comes under deep learning. CNN learnt architecture gives more meaningful and deep features using which more accurate results are possible. The CNN is further classified into some modified architectures as per the requirement [1] which are very useful for MRI image classification.

3. K-Nearest Neighbor Classification Method (KNN)

K-nearest neighbor is the simplest classification method for MR images. In this method, the K closest training vectors of feature vector X are determined to suitable distance metric. The metric use is Euclidian distance. KNN [2, 10] is a conventional classifier for supervised learning. The method consists of two phases, a training phase where the data points are scattered in n-dimensional space; the data points are labeled and designates their class. In testing phase, the unlabeled data is fed to algorithm, to generate the list of K-nearest data points to the unlabeled data. Hence, the class of data points which turns out to be nearest to the labeled data points designates the class of unlabeled data. KNN is simple but powerful classifier; the performance

of KNN depends upon the selection of K value and the distance metric used. The value of K should not be so small that the data points remain unlabeled and it may not be so big that may result in overlapping of various classes. KNN is beneficial in diagnosing brain tumors and other abnormalities in brain causing various diseases.

4. Probabilistic Neural Network (PNN)

Probabilistic Neural Network is a machine learning algorithm for the classification of MRI images into different classes of brain images. PNN is mainly used for brain tumor image classification. The PNN [4, 17] consists of three layers, namely input layer, pattern layer and competitive layer. The input vector pairs as well as target vector pairs are divided into a number of classes. After training, the network is able to classify the input vector into a specific class due to maximum probability of correctness of this class. PNN is mostly used in pattern recognition problems. PNN uses probability distribution function for the approximation of classes. It is much faster and accurate than multilayer perceptron networks.

5. Support Vector Machine (SVM)

SVM uses a nonlinear mapping function for converting input space into higher dimensions [14, 18]. After this, it separates the hyperplane on the basis of the distance from the closest point in training set. SVM is preferred for supervised classification of brain images. SVM can perform linear as well as nonlinear classification; it is widely used in biological sciences due its classification accuracy. SVM is preferred for pattern recognition of in MRI images, because of its computational efficiency and its ability to generalize the performance [3].

B. Unsupervised Classification Methods

1. Self-organizing Map (SOM)

SOM is an unsupervised method for classification of images [14]. In this method, the series of input is supplied and SOM learns itself to group the similar patterns together and hence produce similar outputs. The input data is mapped onto SOM module having two-dimensional grids. The mapping is done from the training data where the SOM modules are adjusted with respect to feature vectors. In the first step, SOM initializes the weight vectors; from that point, a sample vector is randomly selected. Now, the weight vectors of similar match are searched and mapped with the sample weight vector. As each weight vector has some location, its neighboring weights having similar characteristics will also be mapped. SOM applies competitive learning instead of error-correction learning. SOM is preferred for MRI image segmentation because of its unsupervised nature, as there are always new and critical patterns with variety of abnormalities in brain tissues. The self-organizing nature of this method improves the quality of diagnosis to a greater extent.

2. K-Means Clustering Method

K-means clustering method classifies the data points into multiple classes on the basis of the inherent distance between these points. For each cluster, the algorithm specifies K centroids [19]. These K centroids are then placed in different locations

and distant locations. In the next step, each point from a given set is associated with the nearest K centroid. The above step is repeated until no point is pending and the grouping is done. After this step, we get K new centroids of clusters, on the basis of these new centroids new mapping starts between the same data points and closest new centroid. This process continues and K centroids change their location step by step until no changes are seen or the centroids stops moving. The outcome is the formation of clusters around the centroids. The MRI data is too large to process, the K-means clustering method is beneficial, and it converges faster and forms appropriate clusters easily, defining various abnormalities in brain.

3 Literature Survey

Medhani Menikdiwela, Chuong Nguyen and Marnie Shaw (2018 IEEE) proposed a methodology for the detection of Alzheimer's disease using deep learning for the classification of MRI scans. The MRI scan dataset was applied to CNN, Inception and ResNet. The Inception network achieved the highest accuracy followed by ResNet and CNN.

Reema Mathew A and Dr. Babu Anto P (2017) proposed a method for tumor detection and classification. In this work, the MRI brain images were preprocessed using anisotropic diffusion filters, followed by discrete wavelet transform-based feature extraction. After feature extraction, SVM was used for classification and hence, detection of tumor. The classification accuracy for the proposed methodology was 86%.

Sathya Subramanian and Manvalan Radhakrishnan (2015) proposed a partial differential equation method for improving the quality of image before classification. For classification, neural network (NN) classifier was used with bee colony optimization (BCO). The experiment was conducted over 100 MRI brain images, consisting of normal and abnormal brain image samples. The classification accuracy in case of NN with BCO was 76% while in case of NN without BCO was up to 74%.

Waala Hussein Ibrahim, Ahmed Abdel Rhman Ahmad Osman and Yusra Ibrahim Mohamad (2013) proposed an approach for MRI brain image classification. The authors used BPNN as a classifier, and the experiment was conducted over 3×58 datasets of MRI images. The classification accuracy of 96.33% was achieved for the proposed method.

Dina Abdul Dahab, Sunny S.A. Ghoniemy and Gamal M. Selim (2012) proposed a modified PNN based on linear vector quantization (LVQ). The experiment was carried out on MRI brain tumor image dataset. The LVQ-based PNN achieved 100% classification accuracy. Also, the proposed method achieved 79% decrease in processing time as compared to conventional PNN classifier.

Shahla Najafi, Mehdi Chehel Amirani, and Zahira Sedhgi (2011), proposed a methodology with three classifiers: KNN, Parzen window and artificial neural network (ANN). The classification accuracy for ANN was 98.4%; for Parzen and KNN, it was 99.2%.

Mohd Fauzi Othman and Mohd Ariffanan Mohd Barsi (2011) used PNN classification technique for brain tumor detection. The proposed method used twenty MRI brain images for the training of PNN classifier and achieved accuracy ranging from 73 to 100% according to spread values smoothing factor.

El-Sayed Ahmed El-Dahshan, Tamer Hosny, Abdel-Badeeh M-Salem (2009) have developed a methodology for classification of MRI brain images. The brain image data was preprocessed with wavelet transforms and principal component analysis before classification. The proposal produced 90% classification accuracy rate for both k-nearest neighbor (KNN) and feed-forward back-propagation network (FPNN).

4 Conclusion

This paper studied the current trends in the classification of MRI images. A general overview of various supervised and unsupervised classification methods was discussed. The analysis was mainly focused on MRI brain images classification methods. The classification accuracy of MRI brain images played important role for the diagnosis of various diseases with accuracy. The field of medical image diagnosis is full of challenges as the techniques available lags in various aspects, so there is always a scope for the further enhancements in the performance of the existing methods. The survey will be useful for researchers and new learners who aim for the advancement in MRI image classification methods.

References

1. Amelio L, Amelio A Classification methods in image analysis with a special focus on medical analytics. In: Machine learning paradigms. Intelligent systems reference library, vol. 149. Springer, Cham
2. El-Dahshan ESA, Hosny T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. Digital signal processing 20 (2010). Elsevier, pp 433–441
3. Mathew AR, Anto PB (2017) Tumor detection and classification of MRI brain image using wavelet transform and SVM. In: 2017 (IEEE-Xplore) international conference on signal processing and communication (ICSPC). Coimbatore, 2017, pp 75–78
4. Ghotekar BB, Mahajan KJ (2016) MRI Brain image segmentation and classification: a review. IRJET 3:1170–1176
5. Fuse H, Oishi K, Maikusa N, Fukami T, Initiative JADN (2018) Detection of Alzheimer's disease with shape analysis of MRI Images. In: 2018 joint 10th international conference on soft computing and intelligent systems (SCIS) and 19th international symposium on advanced intelligent systems (ISIS). Toyama, Japan, 2018, pp 1031–1034
6. Sivasundari S, Siva Kumar R, Karnan M (2014) Review of MRI image classification techniques. IJRSCSE 1:21–28
7. Kumar G, Bhatia PK (2014) A detailed review of feature extraction in image processing system. ICACCT 2014, pp 06–12
8. Medjahed SA (2015) A comparative study of Feature Extraction methods in Images Classification. I J Image Graph Signal Process 16–23

9. Ryszard SC (2007) Image feature extraction techniques and their applications for CBIR and biometric systems. IJBBE 1:6–16
10. Dokare I, Kant N (2014) Performance analysis of SVM, k-NN and BPNN Classifiers for motor imagery. IJETT 10:19–23
11. Ibrahim WH, Osman AAA, Mohamed YI (2013) MRI brain image classification using neural networks. ICCEEE 2013, published in IEEE Xplore, pp 253–258
12. Fouzia A, Syed MSI, Naveed A, Naeem KJ (2019) Going deep in medical image analysis: concepts, methods, challenges and future directions. IEEE Access, pp 1–33
13. Wahid, F, Fayaz M, Shah, Abdul S (2016), An evaluation of automated tumor detection techniques of brain magnetic resonance imaging (MRI). Int J Bio-Sci Bio-Technol 8:265–278. https://doi.org/10.14257/ijbsbt.2016.8.2.25
14. Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and machine learning. Biomed Signal Process Control Elsevier, pp 86–92
15. Jadoon MM, Zhang Q, Haq IU, Butt S, Jadoon A (2017) Three-class mammograph classification based on descriptive CNN features. BioMed Res Int 11 pages
16. Menikdiwela M, Nguyen C, Shaw M (2018) Deep learning on brain cortical thickness data for disease classification. In: 2018 digital image computing: techniques and applications (DICTA). Canberra, Australia, 1–5
17. Dahab DA, Ghoniemy SSA, Selim GM (2012) Automated brain tumor detection and identification using image processing and probabilistic neural network techniques. I J Image Process Visual Commun 1:1–8
18. Gladis, VP Rathi P, Palani S (2012) Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis IJIST 2:131–146
19. Dhankar S, Tyagi S, Prasad TV (2010) Brain MRI segmentation using K-means algorithm. In: Proceedings 2010, https://doi.org/10.13140/RG.2.1.4979.0567
20. Garg RK, Kulshreshtha A (2017) A review of automated MRI image processing techniques employing segmentation and classification IJCST 5:117–120

Analysis of Electroencephalogram (EEG) Signals for Detection of Major Depressive Disorder (MDD) Using Feature Selection and Reduction Techniques

Shalini Mahato, Abhishek Roy, Akshay Verma, and Sanchita Paul

Abstract Extensive research has been done in identification of major depressive disorder (MDD) or depression using various classifiers and features. Many features are extracted from each channel. Hence, the accuracy of a model can be substantially improved with feature selection and reduction technique. In this paper, different selection technique, i.e., variance inflation factor (VIF) and min-redundancy-max-relevance (mRMR) and reduction technique principal component analysis (PCA) have been analyzed. Publically available dataset contributed by Wajid et al. ("A wavelet-based technique to predict treatment outcome for major depressive disorder") is used in the study. The dataset comprised of thirty-four MDD patients and thirty healthy subjects. Band power and alpha asymmetry are the features used in the study. The classifier used is logistic regression (LR). From the results, it can be concluded that the selection technique (VIF and mRMR) is more efficient than the reduction technique (PCA) for the given dataset. It can also be concluded that mRMR is more efficient than VIF for the given dataset.

Keywords Major depressive disorder (MDD) · Logistic regression (LR) · Variance inflation factor (VIF) · Min-redundancy-max-relevance (mRMR) · Principal component analysis (PCA)

S. Mahato · S. Paul
Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, India
e-mail: swarup.shalini@gmail.com

S. Paul
e-mail: sanchita07@gmail.com

A. Roy (✉) · A. Verma
Department of Mathematics and Computing, Birla Institute of Technology, Mesra, Ranchi 835215, India
e-mail: abhishekroy993@gmail.com

A. Verma
e-mail: akshay.verma07@yahoo.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_39

1 Introduction

MDD or depression is a mental illness which negatively influences the feeling, thinking and activity of a person. More than 0.3 billion people suffer with MDD worldwide. WHO envisages that by 2030 one of the major causes of disease burden will be of MDD patients [1, 2].

Person suffering from depression shows some special characteristics which unravel early recognition of people suffering from depression. Depression patients lack of interest in day to day chores of life and delightful activities. Additionally, four more symptoms like feeling unenergetic, worthless, guilty, restless, nervous, weight loss/weight gain, having suicidal thoughts, problem in concentrating, unable to take decision, etc., for a minimum period of two weeks serve as a basis for identification of MDD patient [3].

Presently, diagnosis of depression is mainly based on internationally recognized questionnaires like diagnostic and statistical manual of mental disorders-fifth edition (DSM-V), Hamilton depression rating scale (HAM-D) and Beck depression inventory (BDI) [4].

A set of standardized questionnaires as per DSM-V criteria forms the basis for detection of MDD. The method is of subjective nature and may lead to incorrect diagnosis. Presently, there is no scientific system for the explanation of depression.

Depression exhibits degradation in brain functioning. The change in brain activity is expected to be reflected in EEG. EEG being a safe, cost efficient and non-invasive technique, it can be used to detect brain functioning during depression and can be helpful in creating a bio-marker for depression. Considerable amount of research has been carried out utilizing various distinct EEG features and classifiers to enhance the classification accuracy for detection of MDD patients. Feature selection and reduction plays significant role as huge amount of features are being extracted from each of the channel.

Considerable research has been conducted to enhance the classification accuracy for detection of MDD patients. For detection of MDD, four nonlinear features (Lyapunov exponent (LE), Higuchi's fractal dimension (HFD), detrended fluctuation analysis (DFA), correlation dimension (CD)) and band power were extracted from each of the nineteen channels. This led to large feature set so selection was done using genetic algorithm (GA) along with LR classifier which gave an accuracy of 90%, GA along with K-nearest neighbor (KNN) classifier gave an accuracy of 80% and GA along with linear discriminant analysis (LDA) classifier provided an accuracy of 86.6% [5]. Accuracy of 80% was obtained using decision tree (DT) in which LDA was used for feature reduction, and GA was used for feature selection and band power as feature [6]. Feature selection for selecting wavelet features was done using receiver operating characteristic (ROC) along with LR to provide an accuracy of 87.5% [7].

The purpose of this paper is to compare the effectiveness of different feature selection techniques (VIF and mRMR) and feature reduction technique (PCA) using alpha, beta, delta, theta power and alpha asymmetry as feature and LR as classifier.

2 Materials and Methods

Subjects
The dataset of EEG signals for MDD and healthy subjects used in the paper is the dataset contributed by Wajid et al. [7]. It consisted of EEG signal of 30 healthy subjects (9 females and 21 males) and 34 MDD patients (age matched 17 females and 17 males). The average age of MDD patients is 40.3 years and standard deviation of 12.9 years. The average age of healthy subjects is 38.3 years and standard deviation of 15.6 years.

The inclusion criteria for MDD patient was based on DSM IV. Approval to above study [7] was given by Human Ethics committee of Hospital Universitii Sains Malaysia (HUSM), Malaysia.

Data Acquisition
Five minute eyes closed, EEG data was recorded using 19 electrodes as per the international 10–20 electrode system. According to the electrode position in 10–20 system, electrodes can be divided into different regions: central (C_3, C_4, C_z), frontal (Fp_1, Fp_2, F_3, F_4, F_7, F_8, F_z), parietal (P_3, P_4, P_z), occipital (O_1, O_2) and temporal (T_3, T_4, T_5, T_6). EEG signal was recorded with linked ear (LE) reference with frequency band 0.1–70 Hz at sampling frequency of 256 Hertz. Notch frequency of 50 Hz was used for line noise removal.

Preprocessing Degradation of EEG signal is caused by many artifacts like muscle activity, eye blink, movements, noise from electrical power line, etc. To prevent wrong EEG signal analysis, it is necessary to remove these unwanted signals arising due to these artifacts.

In order to clean the EEG data, a number of techniques have been applied. EEG signal was filtered at a cut off frequency of 0.5–32 Hz. Power line noise as well as muscle artifact signals are removed automatically by this process as they lie above 32 Hz. Eye blink and eye movement (vertical and horizontal) artifacts were removed by applying independent component analysis (ICA) [8]. Background noise was removed by re-referencing to common average reference (CAR) [9, 10].

3 Methodology

Feature Extraction
Band Power. Band power for alpha (8–13 Hertz), beta (13–30 Hertz), delta (0.5–4 Hertz) and theta (4–8 Hertz) was extracted from EEG signals. Power spectrum was computed using Welch periodogram.

Alpha Inter-hemispheric Asymmetry. Alpha inter-hemispheric asymmetry is computed as alpha power of left half of the brain minus the right half of the brain is termed as alpha inter-hemispheric asymmetry [11]. It is calculated as difference

between channels C_4–C_3, P_4–P_3, T_6–T_5, T_4–T_3, Fp_2–Fp_1, F_4–F_3, F_8–F_7 and O_2–O_1. Mathematically, it is represented as

$$\text{Alpha}_{\text{sym}} = \text{mean}(\log(\text{PW}_r) - \log(\text{PW}_l)) \quad (1)$$

where PW_r is the sum of alpha power of right channels and PW_l is the sum of power alpha power of left channel.

Feature Selection and Reduction

Large number of features are extracted from each of the channel. Inclusion of all the features in the classifier would lead to very high cost but if only important features are selected, there would be a significant reduction in cost as well as the efficiency may also increase. So, feature selection and reduction is used [12, 13].mRMR, PCA, LDA, GA and VIF are few of the feature selection and reduction techniques.

The block diagram representation of the study is shown below in Fig. 1.

Min-Redundancy-Max-Relevance (mRMR): It is a selection technique which is very often used in methods to accurately identify genes and phenotypes by narrowing down in the order of their relevance. It distinguishes the subsets of data that are important to the parameters used and is normally called maximum relevance. The

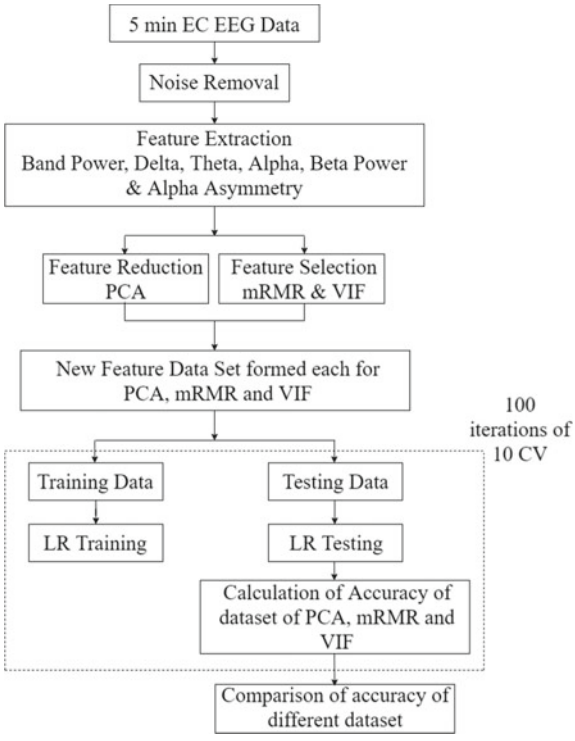


Fig. 1 Block diagram representation of the study where feature selection was done using min-redundancy-max-relevance (mRMR) and variance inflation factor (VIF) and feature reduction was done using principal component analysis (PCA)

other redundant subsets are removed. It is an attribute selection approach that selects features which are in high correlation with the class or predictor (output) and low correlation among themselves [1, 14, 15].

Mutual information $M(R, S)$ calculates the degree of likeness between two discrete random variables R and S:

$$M(R, S) = \sum_{r \int R} \sum_{s \int S} p(r, s) \log\left(\frac{p(r, s)}{p_1(r)\, p_2(s)}\right) \tag{2}$$

Here, $p(r, s)$ represents the joint probability distribution function of R and S, and $p_1(r)$ and $p_2(s)$ represent the marginal probability distribution functions corresponding to R and S accordingly.

The mRMR algorithm next uses the mutual information formula above between target variable label $t = t_1 + t_2 + t_{3.} + \cdots + t_n$ and the attribute i which will be designated as $M(T, F_i)$.

Let U be the attribute set we wish to choose, and $|U|$ be the cardinality of set U. Therefore, the necessary constrains for the feature subset to be best are:

(i) The minimum redundancy condition:

$$A = \frac{1}{|U|^2} \sum_{F_i, F_j \in U} M(F_i, F_j) \tag{3}$$

(ii) The maximum relevance condition:

$$B = \frac{1}{|U|} \sum_{F_i \in U} M(F_i, T) \tag{4}$$

The simplest way for the above combinations can be:

$$\text{Max}(B - A) \tag{5}$$

$$\text{Max}(B/A) \tag{6}$$

The mRMR uses the algorithm where it selects the first feature. Then, the characteristic i which satisfies the constraints selected at every step with the remaining selected attributes in the set U.

Let Ω be the feature subset of all data. So, let $\Omega_U = \Omega - U$ be the feature subset of every feature excluding those selected already. Then,

$$\min_{F_i \in \Omega_U} \frac{1}{|U|} \sum_{F_j \in U} M(F_i, F_j) \tag{7}$$

$$\max_{F_i \in \Omega_U} M(F_i, T) \tag{8}$$

Using Eqs. (7) and (8) in Eqs. (5), it will result into the equation:

$$\max_{F_i \in \Omega_U} \left[M(F_i, T) - \frac{1}{|U|} \sum_{F_j \in \Omega_U} M\left(F_i, F_j\right) \right] \tag{9}$$

is called mutual information difference (MID). MID has been used in this study.

Variance Inflation Factor (VIF): VIF is broadly used to measure of the degree of multi-collinearity of the kth independent variable with other variables in a simple regression model [16]. It is the ratio of the variance with numerous terms divided by the variance of the model with a single term. It scales the strictness of multi-collinearity in an ordinary least squares regression analysis. After applying it successfully, it gives an index that measures the increase of the variance of a regression coefficient that would have occurred because of collinearity.

Collinearity in a dataset occurs when two variables are highly correlated and it occurs that they contain similar information about the variance of the dataset. It increases the variance of regression parameters which could lead to the wrong prediction values [17].

Multi-collinearity emerges when more than two features of a dataset are highly correlated. Multi-collinearity can emerge even when isolated pairs of variables are not collinear. Multi-collinearity can inflate the variance among the features in a model. These inflated features create problems in regression because few variables contribute to little or no new information to the model [18].

VIF factor gives an index that could measure how much the variance of an estimated regression coefficient is changed because of collinearity.

VIF for the estimated coefficient c_n denoted by VIF_n is the factor by which variance is inflated.

Let x_n be the only feature used

$$-y_i = \beta_Q + \beta_n x_{\mathrm{in}} + \varepsilon \tag{10}$$

where

β_n is coefficient of feature x_{in}
β_Q is bias of the model
ε is error
y_i is response.

The variance of b_n can be shown as

$$-\mathrm{Var}\left(b_\mu\right)\min = \frac{\sigma^2}{\sum_{i=1}^{m} (x_{\mathrm{in}} - \bar{x}_{\mathrm{n}})^2} \tag{11}$$

whereσ^2 is variance of the entire dataset$\bar{x}_n$ is mean of features x_{in}And "min" represents the smallest variance.

Let there be a model with correlated features,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta x_{iu} + \beta_p - 1 x_{jp} - 1 + \varepsilon \tag{12}$$

where

y_i is response
β_n is coefficient of feature x_{in}
β_0 is bias of the model
ε is error.

Now, if some features are correlated with features x_n, then the variance of b_n is inflated.

$$\text{Var}(b_n) = \frac{\sigma^2}{\sum_{i=1}^{m}(x_{\text{in}} - \bar{x}_{\text{n}})^2} \times \frac{1}{1 - R_n^2} \tag{13}$$

where R_n^2 is the R^2-value obtained by regressing the nth feature on the remaining features. The greater the linear dependence the larger the R_n^2 value and from earlier equation larger value of R_n^2 means larger variance of b_n.

Now, to find exactly how large it is, the ratio of the two variances is calculated.

Now, dividing Eq. (13) by Eq. (11), we get

$$\frac{\text{Var}(b_n)}{\text{Var}(b_n)\text{min}} = \frac{\frac{\sigma^2}{\sum_{i=1}^{m}(x_{\text{in}} - \bar{x}_n)^2} x \frac{1}{1-R_n^2}}{\frac{\sigma^2}{\sum_{i=1}^{m}(x_{\text{in}} - \bar{x}_n)^2}} = \frac{1}{1 - R_n^2} \tag{14}$$

The above factor is called variance inflation factor for the nth feature.

$$\text{VIF}_n = \frac{1}{1 - R_n^2} \tag{15}$$

The variance inflation factor exists for each of the n features in a multiple regression model.VIF factors are always greater than 1. The VIF factor tells us what percentage of the variance is inflated for each coefficients present.

Principal Component Analysis (PCA): PCA is used for data reduction [19]. It is done by decreasing the number of dimensions which is having less information. Basically, this technique consists of two major steps (i) computation of covariance for the data matrix is done, (ii) eigenvectors and eigenvalue for the computed covariance matrix are calculated.

The computation for the eigenvector and eigenvalues is done using the following equation.

$$Eg_V \sum = \lambda Eg_V \tag{16}$$

where Eg_V, λ and $\sum$ correspond to eigenvectors, eigenvalues and covariance matrix, respectively.

Eigenvalues are scalar values which represent magnitude or scaling factor. Eigenvectors represent the principal component. First, the principal component is the principal component with highest eigenvalue and has largest variance. Eigenvectors are arranged on the basis of decreasing value of eigenvalues. Eigenvectors with largest eigenvalues are selected.

Classifier Model Logistic Regression: Logistic regression is a categorization algorithm usually applied to binary dependent variables [20]. Binary outcomes like Yes/No, 1/0, True/False are forecasted given a set of attributes. For the representation of binary or categorical output, dummy variables are used. Logistic regression is also the significant case of linear regression when the output variable is discrete, where we use log of odds as dependent variable [21, 22]. Basically, the function foretells the probability of occurrence of an event by fitting data to the logit function described below. It is also called sigmoid function that is rising quickly and maxing out at certain level [23]. The equation is given by:

$$\log(p) = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_N. \tag{17}$$

where p is the probability of the characteristic of interest or dependent variable.

The logit transformation is defined as the logged odds:

$$\mathrm{ODDS} = \frac{p}{1-p} = \frac{\text{probability of presesnce of characteristic}}{\text{probability of absence of characteristic}} \tag{18}$$

and

$$\log \mathrm{it}(p) = \log\left(\frac{p}{1-p}\right) \tag{19}$$

So, the logistic function can now be written as:

$$p(x) = 1/\left(1 + e^{-(\beta 0 + \beta 1)}\right) \tag{20}$$

Validation Cross validation is used to verify the models, prediction accuracy in new dataset. In this paper, we have used 100 iterations of 10 fold CV. Dataset is divided into 10 parts, of which nine parts are selected randomly as training data and one part as testing data. This process is repeated for fixed number of iterations.

The accuracy of classifier is defined as

$$\text{Accuracy} = \frac{\text{No. of testing data correctly classified}}{\text{Total no. of testing data}} \tag{21}$$

4 Results

The results based on the different selection and reduction technique for the features alpha, beta, delta, theta power and alpha asymmetry and combination of alpha power and alpha asymmetry when applied on LR classifier are shown below in Table 1 and Fig. 2 [24].

From the results in Table 1, it can be observed that there is always an improvement in accuracy by applying either selection techniques (VIF and mRMR) or reduction technique (PCA) as compared to when none of these methods are being applied.

From Fig. 2, it can also be found that selection techniques (VIF and mRMR) always outperform reduction technique (PCA) for the given dataset. From Fig. 2, it can also be found that mRMR works better than VIF.

From Table 1, it can be found that 88% of accuracy is the maximum accuracy achieved when alpha power was used as a feature, mRMR as feature selection technique and logistic regression as a classifier.

The result from Table 1 also shows the higher potential power of alpha band in detection of depression patients as compared to other bands, i.e., delta, theta, beta, alpha asymmetry considered in this study.

The comparison in Table 2 shows that the proposed work's accuracy is quite high as compared other works done in depression.

Table 1 Performance comparison of selection (VIF and mRMR) and reduction techniques (PCA)

	Alpha power (%)	Beta power (%)	Delta power (%)	Theta power (%)	Alpha asymm (%)	Alpha power + Alpha asymmetry (%)
Without selection/reduction techniques	85	76	70	83	58	60
With VIF	87	82	82	83	63	87
With MRMR	88	85	86	87	63	75
With VCF	83	78	79	86	62	92

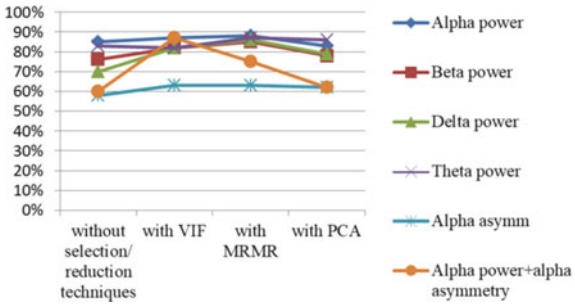


Fig. 2 Classification accuracy using selection (VIF and mRMR) and reduction techniques (PCA)

Table 2 Comparison of the proposed method with previous work in depression

Sl. No.	Ref.	Methodology	Analysis method	Features	Accuracy (%)
1	Proosed	LR as classifier and PCA, mRMR and VIF as feature reduction and selection technique	LR	Band power	88
2	[25]	Artificial neural network (ANN) was used as a classifier and band power as the feature	ANN	Band power	84
3	[26]	Features used were Higuchi's fractal dimension (HFD) and linear spectral asymmetry index (SASI) and analysis was done using t-test	t-test	HFD, SASI	85
4	[6]	DT was used as classifier along with LDA and GA with band power as feature	DT	Band power	80
5	[27]	Classification of normal, depression and schizophrenic individual was done using Lep-Ziv complexity as a feature and backpropagation artificial neural network as classifier	BP-ANN	Lep-Ziv complexity	60–80

5 Discussion

Thus, from the results, whenever selection technique, i.e., VIF or mRMR is applied or reduction technique PCA is applied there is always an improvement in accuracy as compared to when none of the selection or reduction technique is applied.

From among the selection technique, mRMR and VIF (except in case of combination of alpha power + alpha asymmetry), mRMR performs better than VIF for the given dataset. It can be also seen that selection techniques (VIF and mRMR) always outperform PCA for the given dataset.

6 Conclusion

The study reveals that selection techniques (VIF and mRMR) are better than PCA for the given dataset and when VIF and mRMR are used, accuracy mRMR works better than VIF for the given dataset. Thus, it can be concluded that it is better to use selection technique as compared to reduction technique when classification of depression patient is done using LR for the given dataset. The study also reveals the higher potential power of alpha band in detection of depression patients as compared to other bands, i.e., alpha, delta, theta, beta, alpha asymmetry considered in this study.

Thus, the paper finds important results which could be an essential part in improving the classification accuracy of depression patients.

References

1. World Health Organization.: Global Burden of Mental Disorders and the Need For a Comprehensive, Coordinated Response From Health and Social Sectors at The Country Level. Report by the Secretariat. EB 130/9 (2011)
2. World Health Organization (2017) Depression and Other Common Mental Disorders Global Health Estimates. WHO Document Production Services, Geneva, Switzerland
3. American Psychiatric Association.: Diagnostic and Statistical Manual of Mental Disorders, 4th edn, pp. 339–345. American Psychiatric Association Washington, DC (1994)
4. Cusin, C., Yang, H., Yeung, A., et al,: Rating scales for depression. In: Baer, L., Blais, M.A. (eds.) Handbook of Clinical Rating Scales and Assessment in Psychiatry and Mental Health, pp. 7–37. Current Clinical Psychiatry, Boston, USA (2010)
5. Hosseinifarda B, Moradia MH, Rostami R (2013) Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Comput. MethodsPrograms Biomed. 109:339–345
6. Mohammadi, M., et al.: Data mining EEG signals in depression for their diagnostic value. In: BMC Medical Informatics and Decision Making, pp. 108–123 (2015)
7. Mumtaz W, Xia L, Ali SSA et al (2017) A wavelet-based technique to predict treatment outcome for major depressive disorder. PLoS ONE 2017:1–30
8. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 9–21 (2004)
9. Gandhi, V.: Brain computer interfacing for assistive robotics. In: Electroencephalograms, Recurrent Quantum Neural Networks, and User-Centric Graphical Interfaces, 1st edn, pp. 21–29. Academic Press (2014)
10. Mahato, S., Paul, S.: Detection of major depressive disorder using linear and non-linear features from EEG signals. In: Microsystem Technologies, pp. 1–12 Springer (2018)
11. Mahato S, Paul S (2017) Electroencephalogram (EEG) signal analysis for diagnosis of major depressive disorder (MDD): a teview. In: Nath V, Mandal JK (eds) Proceeding of Nanoelectronics, Circuits and Communication Systems (NCCS 2017). Springer, Singapore, pp 323–336
12. Shroff, K.P., Maheta, H.: A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy. In: 2015 International Conference on Computer Communication and Informatics (ICCCI), pp. 4–6. IEEE (2015)
13. Sorzano, C.O.S., Vargas, J., Pascual, A.: A Survey Of Dimensionality Reduction Techniques, pp. 14–25. arXiv:1403.2877 (2014)

14. Nugraha, I., Ayuningtyas, C., Saptawati, P.: Performance analysis of relief and mRMR algorithm combination for selecting features in lupus Genome-Wide Association Study. In Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, pp. 2–5. IEEE (2011)
15. Zhang Z et al (2011) MRMR optimized classification for automatic glaucoma diagnosis. Ann. Int. Conf. IEEE Eng. Med. Biol. Soc. 2011:6228–6231
16. Brien, R.M.O.: A Caution regarding rules of thumb for variance inflation factors. Q Quantity **41**, 673–690 (2007)
17. Dormann CF, Elith J, Bacher S, Buchmann S, Carl G, Carré G et al (2012) Collinearity: A Review of Methods to Deal With it and a Simulation Study Evaluating Their Performance. Dept of Computational Landscape Ecology, Helmholtz Centre for Environmental Research-UFZ, pp 28–30
18. Jamal, I.: Daoud-IOP conference,"Multicollinearity and regression analysis. J Phys. 3–5 (2017)
19. Jolliffe IT (2002) Principal component analysis, series: Springer series in statistics, 2nd edn. Springer, New York, USA, pp 1–147
20. Kutner, M.,Nachtsheim, C., Neter, J.: Applied Linear Regression Models Fourth Edition, pp. 214–324. Published by McGraw Hill Irwin (2004)
21. Yusuff, H., Mohamad, N., Ngah, U.K.,. Yahaya, A.S.: Breast Cancer Analysis Using Logistic Regression, pp. 1200–1201. Published in ArpaPress (2012)
22. de Jongh, P.J., de Jongh, E., Pienaar, M., Gordon, G.H., Oberholzer, M., Santana,: The Impact of Pre-Selected Variance Inflation Factor Thresholds on the Stability and Predictive Power of Logistic Regression Models in Credit Scoring, pp. 18–20. Published in Journal Orion (2014)
23. Ranganathan, P., Pramesh, S.C., Aggarwal, R.: Common Pitfalls in Statistical Analysis: Logistic Regression, pp. 11–16. US National Library of Medicine (2017)
24. Vanaja, S., Rameshkumar, K.: Performance analysis of classification algorithms on medical diagnoses-a survey. J. Comput. Sci. 36–37 (2014)
25. Mantri, S., Agrawal, P., Patil, D., Wadhai, V.: Non invasive EEG signal processing framework for real time depression analysis. SAI Intelligent Systems Conference, pp. 518–521 (2015)
26. Bachmann, M., Lass, J., Suhhova, A., Hinrikus, H.: Spectral asymmetry and Higuchi's fractal dimension measures of depression electroencephalogram. Comput. Mathe. Methods Med
27. Fan, F., Li, Y., Qiu, Y., Zhu, Y.: Use of ANN and complexity measures in cognitive EEG discrimination. In: Proceedings of 27th Annual International Conference, IEEE-EMBS of Engineering in Medicine and Biology Society, pp. 4638–4464 (2005)

Calculation of Torque Ripple and Derating Factor of a Symmetrical Six-Phase Induction Motor (SSPIM) Under the Loss of Phase Conditions

Purushottam Kumar, Vishal Rathore, and K. B. Yadav

Abstract The objective of this paper is to analyze the behavior of symmetrical six-phase induction motor (SSPIM) under the loss of phase conditions. Various phase loss configurations are considered to analyze the torque profile and derating factor of the SSPIM and compare them with the healthy machine. To analyze the loss of phase condition, a mathematical model of SSPIM has been developed in MATLAB/Simulink environment; it has been seen that under loss of phase's condition the machine will continue to run with almost similar ripple content but at reduced derating factor (DF) as consequences of which the life of motor gets degrades.

Keywords Loss of phase · Six-phase induction motor · Modeling · Torque ripple · Derating factor

1 Introduction

The various benefits of six-phase machine drives over its three-phase counterparts are high-power handling capability which can be achieved by distributing the required power among the six phases, reduction in the pulsation of torque and is highly reliable. Other benefits of the six-phase system are increased of volt/ampere and torque and

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

P. Kumar · V. Rathore (✉) · K. B. Yadav
Electrical Engineering Department, N.I.T., Jamshedpur, Jamshedpur 831014, India
e-mail: 2018rsee009@nitjsr.ac.in

P. Kumar
e-mail: purushottam.kmr1025@gmail.com

K. B. Yadav
e-mail: kbyadav.ee@nitjsr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_40

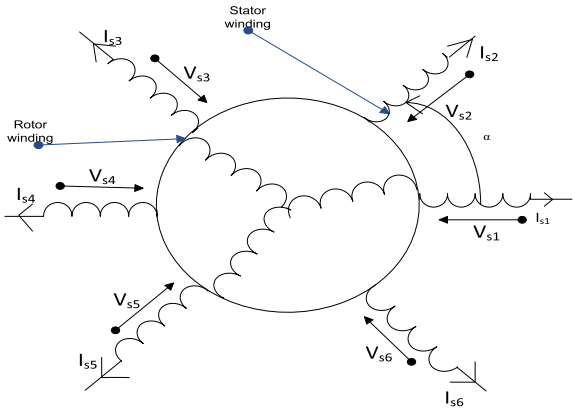


Fig. 1 Winding arrangement of SSPIM

for the same dimension of the motor. Also, rotor current harmonics and stator copper loses get degraded [1, 2].

The application of multiphase machines allows us to gain the advantages of the extra degree of freedom as compared to three-phase machines but the freedom is limited to certain applications such as hybrid electric vehicles, airspace applications, high-power drives, and navel propulsion [3, 4]. If the three-phase motor is working under the open phase required a divided neutral connection and a DC bus. Moreover, we can say, a zero-sequence current is a must to introduce an unbalanced MMF on being of a phase loss condition [5]. On the other hand, the multiphase motor has a better capability of fault tolerance because of the fact that it eliminates the need fora neutral connection required for balancing the rotating MMF [6]. If one phase of an SSPIM is a loss, the phase currents combination to produce an unbalanced rotating forward MMF is not unique. For a healthy operation, it is necessary to generate an optimal value of currents which is required to produce the same MMF [7, 8].

Multiphase induction is capable of self-start and is able to operate under unbalance conditions or when one, two, or three of its stator phases are open. The advantage of symmetrical six-phase induction motor is that the pair of stator three-phase winding is phase displaced 60 °E as shown in Fig. 1 leads to removal of all (6n ± 1)order torque and air gap flux harmonics where $n = 1, 3, 5, 7 \ldots$ [9].

The paper comprises four sections: Mathematical modeling is performed in Sect. 1, and in Sect. 2, loss of phase analysis is presented; Sect. 3 comprises of results and analysis of a study performed, and finally, the conclusion is presented in the last section.

2 Mathematical Modeling of Six-Phase Induction Motor

A vector space decomposition (VSD) model is used to represent a healthy symmetrical six-phase induction motor. Machine variables $[F_k]$ which are based on VSD

model are transformed to a stationary frame of variables $[F_{\alpha\beta xy}]$ via $[D_6]$ [10–13].

$$\left[F_{\alpha\beta xy} = D_6 \cdot F_k\right]$$

where

$$[F_k = F_{s1}\mathrm{F}_{s2}F_{s3}F_{s4}F_{s5}F_{s6}]^{\mathrm{T}}$$

$$\left[F_{\alpha\beta xy} = F_\alpha F_\beta F_x F_y F_{0+} F_{0-}\right]^{\mathrm{T}}$$

$$\boldsymbol{D}_6 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{1}{2} & -1 \\ 1 & -\frac{1}{2} & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & -1 \\ 0 & -\frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & \frac{1}{2} & \frac{1}{2} & -1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

The voltage equations which represent a symmetrical six-phase induction motor in stationary reference frame are as follow,

$$v_{q1} = r_1 i_{q1} + \omega_k \lambda_{d1} + p\lambda_{q1} \tag{1}$$

$$v_{d1} = r_1 i_{d1} - \omega_k \lambda_{q1} + p\lambda_{d1} \tag{2}$$

$$v_{q2} = r_2 i_{q2} + \omega_k \lambda_{d2} + p\lambda_{q2} \tag{3}$$

$$v_{d2} = r_2 i_{d2} - \omega_k \lambda_{q2} + p\lambda_{d2} \tag{4}$$

$$v_{qr} = r_r i_{qr} + \omega_k \lambda_{dr} - \omega_r \lambda_{dr} + p\lambda_{qr} = 0 \tag{5}$$

$$v_{dr} = r_r i_{dr} - \omega_k \lambda_{qr} + \omega_r \lambda_{qr} + p\lambda_{dr} = 0 \tag{6}$$

The equation of flux linkage of six-phase induction motor is given as:

$$\lambda_{q1} = L_{l1} i_{q1} + L_{mq} i_{q1} + L_{mq} i_{q2} + L_{mq} i_{qr} + L_{lm} i_{q1} + L_{lm} i_{q2} + L_{dq} i_{d2} \tag{7}$$

$$\lambda_{d1} = L_{l1} i_{d1} + L_{md} i_{d1} + L_{md} i_{d2} + L_{md} i_{dr} + L_{lm} i_{d1} + L_{lm} i_{d2} + L_{dq} i_{q1} \tag{8}$$

$$\lambda_{q2} = L_{l2} i_{q2} + L_{mq} i_{q1} + L_{mq} i_{q2} + L_{mq} i_{qr} + L_{lm} i_{q1} + L_{lm} i_{q2} + L_{dq} i_{d1} \tag{9}$$

$$\lambda_{d2} = L_{l2} i_{d1} + L_{md} i_{d1} + L_{md} i_{d2} + L_{md} i_{dr} + L_{lm} i_{d1} + L_{lm} i_{d2} + L_{dq} i_{q1} \quad (10)$$

$$\lambda_{qr} = L_{lr} i_{qr} + L_{mq} i_{q1} + L_{mq} i_{q2} + L_{mq} i_{qr} \quad (11)$$

$$\lambda_{dr} = L_{lr} i_{dr} + L_{md} i_{q1} + L_{md} i_{q2} + L_{md} i_{qr} \quad (12)$$

$$\text{Assuming, } L_{dq} = 0 \quad \text{and} \quad L_m = L_{mq} = L_{md} \quad (13)$$

Also,

$$L_1 = L_{l1} + L_{lm} + L_m \quad (14)$$

$$L_2 = L_{l2} + L_{lm} + L_m \quad (15)$$

$$L_3 = L_{lm} + L_m \quad (16)$$

$$L_r = L_{lr} + L_m \quad (17)$$

On substituting the above value in Eq. (7–13), the resultant flux linkage is represented as

$$\lambda_{q1} = L_1 i_{q1} + L_3 i_{q2} + L_m i_{qr} \quad (18)$$

$$\lambda_{d1} = L_1 i_{d1} + L_3 i_{d2} + L_m i_{dr} \quad (19)$$

$$\lambda_{q2} = L_2 i_{q1} + L_3 i_{q2} + L_m i_{qr} \quad (20)$$

$$\lambda_{d2} = L_2 i_{d1} + L_3 i_{d2} + L_m i_{dr} \quad (21)$$

$$\lambda_{qr} = L_r i_{qr} + L_m i_{q1} + L_{md} i_{q2} \quad (22)$$

$$\lambda_{dr} = L_r i_{dr} + L_m i_{d1} + L_m i_{d2} \quad (23)$$

And the resultant electromagnetic torque (T_e) is given by the expression,

$$T_e = \left(\frac{3}{2}\right)\left(\frac{P}{2}\right)\left[\lambda_{md} i_{q1} + \lambda_{md} i_{q2} - \lambda_{mq} i_{d1} + \lambda_{mq} i_{d2}\right] \quad (24)$$

Derating factor (DF): It is the measure of the ratio of modulus of per unit postfault current $I_{\alpha\beta}$ to the rated current provided that the maximum value of $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{post-fault}}$ does not more than the $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{rated}}$. This keeps the machine drive to guard against hot spots and thermal heating at the expense of $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{post-fault}}$ below $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{rated}}$ values [14, 15].

For example, in six-phase induction machines for one phase open $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{post-fault}}$ is 0.762 while the $\left|\mathbf{I}_{\alpha\beta}\right|_{\text{rated}}$ is 0.8186. Therefore, the derating factor (DF) of the machine is 0.9883.

$$DF = \frac{\left|\mathbf{I}_{\alpha\beta}\right|_{\text{post-fault}}}{\left|\mathbf{I}_{\alpha\beta}\right|_{\text{rated}}}$$

3 Loss of Phase Analysis of Six-Phase Induction Machine

When loss of phase occurs in the one phase, two phases and three phases of a six-phase induction machine, under these conditions analysis of dynamic behavior, are performed and compared with a healthy machine, and the ripple content in torque under various conditions for first five cycles is analyzed [16–18]. The current of the stationary frame $I_{\alpha\beta}$ and the derating factor are also calculated. Flowchart of various losses of phase conditions is shown in Figs. 2 and 3 represents voltage phasor of SSPIM.

A. *Under healthy phase condition*

Under healthy phase condition, the percentage of ripple content in torque and $I_{\alpha\beta}$ for the first five cycles is 6.34% and 10.65%, respectively, and the RMS value of $I_{\alpha\beta}$ is 0.8186. Figure 4 shows the waveform of torque, and $I_{\alpha\beta}$ is shown in Fig. 5 (Table 1).

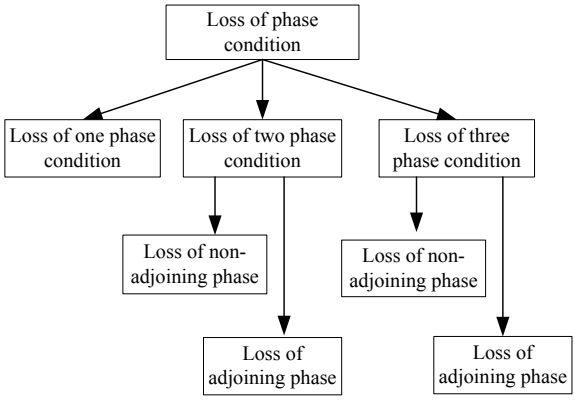


Fig. 2 Flowchart of loss of phase condition

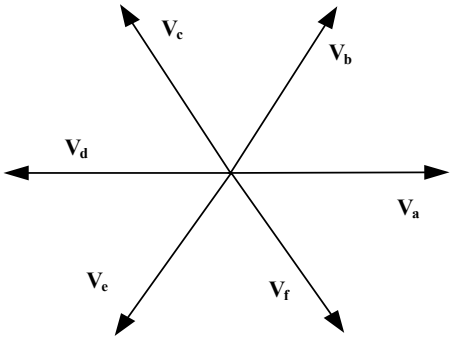


Fig. 3 Voltage phasor of SSPIM

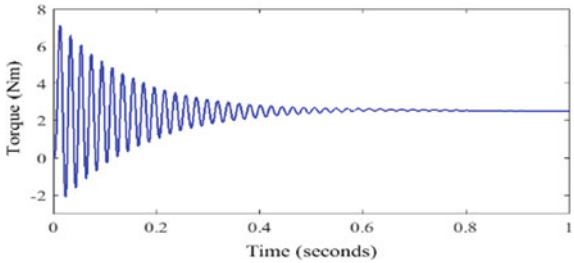


Fig. 4 Torque profile subjected to 6-healthy phase

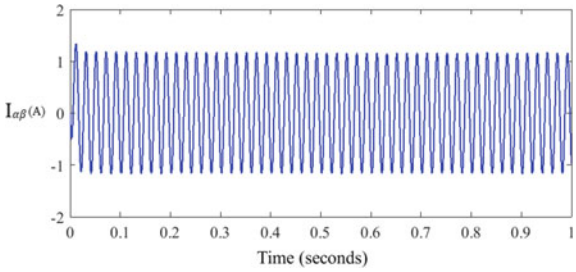


Fig. 5 $I_{\alpha\beta}$ profile subjected to 6-healthy phase

Table 1 Parameters of SSPIM

Parameter	Symbol	Value
Motor rated power	P	1 hp
Leakage inductance of stator	L_s	0.789[H]
Leakage inductance of rotor	L_r	0.789[H]
Magnetizing inductance	L_m	0.728[H]
Stator resistance	R_s	10.5[Ω]
Resistance of the rotor	R_r	9.3[Ω]
Inertia of the rotor	J	0.084[kg*m^2]
Pole pairs	p	2
voltage	V	230(v)

B. *Loss of one phase and five healthy phrases condition*

Under the loss of one phase condition means from healthy phases (*a*, *b*, *c*, *d*, *e*, and *f*), one phase is opened and other phases are closed as shown in Fig. 6. During these conditions, the ripple content in torque and $I_{\alpha\beta}$ is calculated and also derating factor is obtained. Figures 7 and 8 show the waveform of torque and $I_{\alpha\beta}$ for loss of phase (a), respectively, and for other phases are given in Table 2.

C. *Loss of two phases and four healthy phases condition*

Here two possible cases are visualized, first case is the loss of two adjoining phases and the second is loss of two non-adjoining phases. The flowchart of the two cases mentioned above is shown in Fig. 9, and the phasor representation of voltages for 4 healthy phases are shown in Figs. 10 and 11.

During the loss of two nearby phases (*a*, *b*) and non-adjoining phase (*a*, *c*), the waveform of torque and $I_{\alpha\beta}$ of loss of adjoining phase (*a*, *b*) is presented in Figs. 12

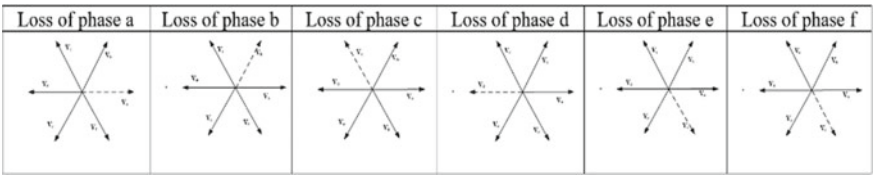


Fig. 6 Voltage phasor of SSPIM with one phase loss condition

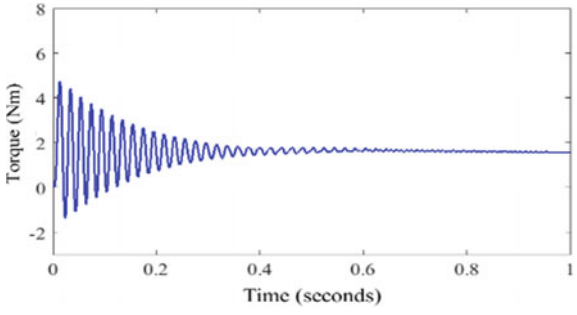


Fig. 7 Torque profile subjected to loss of phase (a)

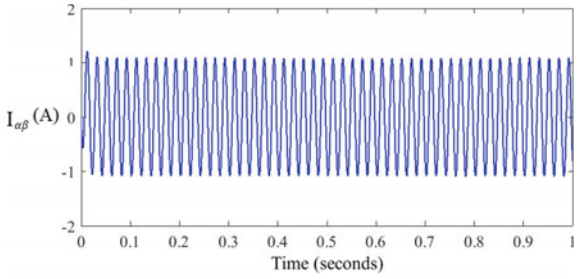


Fig. 8 $I_{\alpha\beta}$ profile subjected to loss of phase (a)

Table 2 Ripple content of torque and $I_{\alpha\beta}$, derating factor for loss of one phase

S.no	Phase condition	%ripple content of torque for first five cycle	$I_{\alpha\beta}$(A)	Derating factor	%ripple content of $I_{\alpha\beta}$
1	a	7.30	0.762	0.9308	12.47
2	b	6.84	0.5473	0.6685	13.25
3	c	7.77	0.7647	0.9341	13.03
4	d	7.30	0.762	0.9308	12.47
5	e	6.84	0.5473	0.6685	13.25
6	f	7.77	0.7647	0.9341	13.03

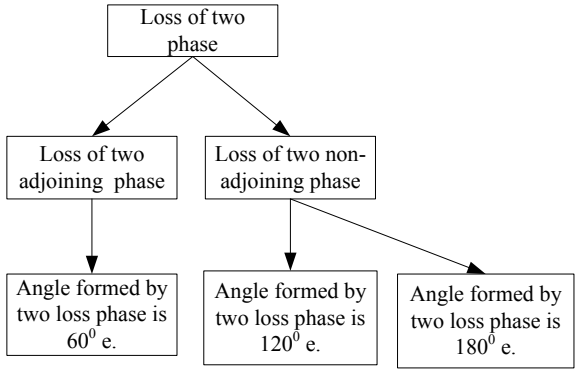


Fig. 9 Flowchart of loss of two-phase condition

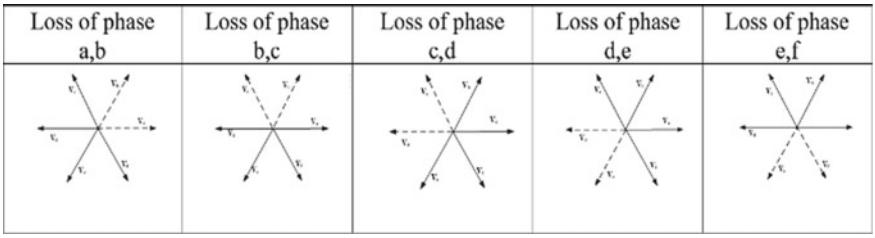


Fig. 10 Voltage phasor of SSPIM with loss of two adjoining phase condition

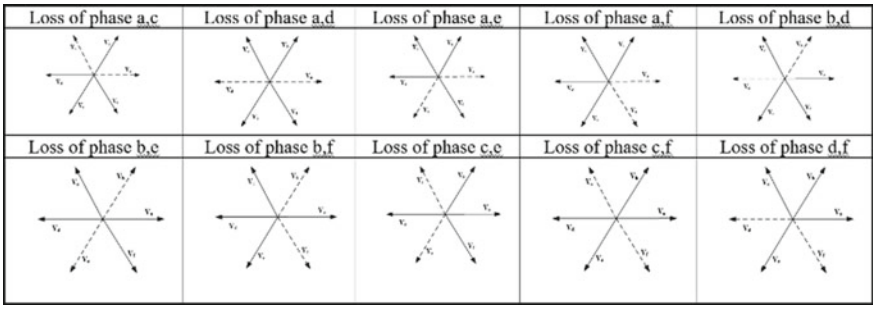


Fig. 11 Voltage phasor of SSPIM with loss of two non-adjoining phase condition

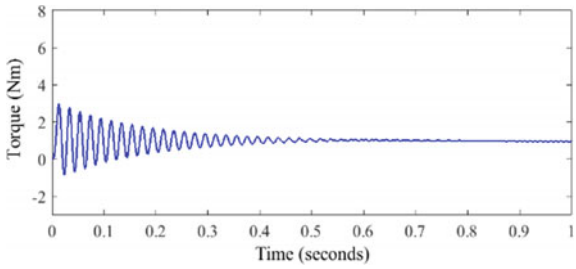


Fig. 12 Torque profile subjected to loss of adjoining phase (a, b)

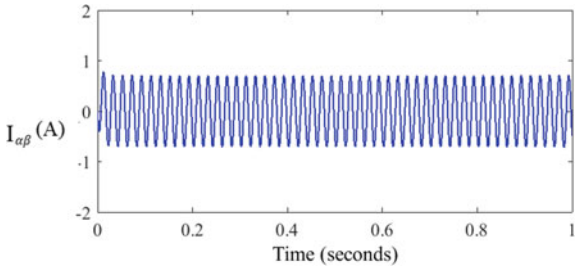


Fig. 13 $I_{\alpha\beta}$ profile subjected to loss of adjoining phase (a, b)

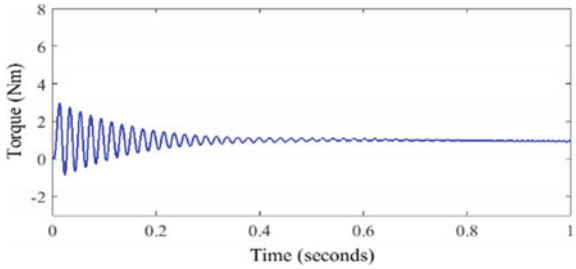


Fig. 14 Torque profile subjected to loss of non-adjoining phase (a, c)

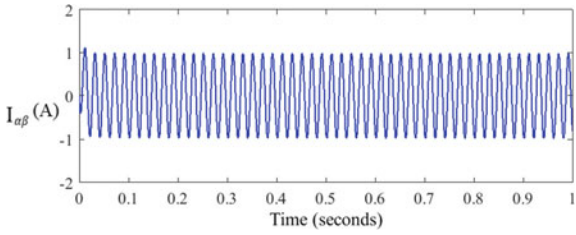


Fig. 15 $I_{\alpha\beta}$ profile subjected to loss of non-adjoining phase (a, c)

and 13 and waveform for loss of non-adjoining phase (a, c) is presented in Figs. 14 and 15, respectively.

Table 3 Ripple content of torque and $I_{\alpha\beta}$, derating factor for loss of two adjoining phases

S.no	Phase condition	%ripple content of torque for first five cycle	$I_{\alpha\beta}$(A)	Derating factor	%ripple content of $I_{\alpha\beta}$
1	ab	6.94	0.4955	0.6053	11.50
2	bc	7.13	0.4962	0.6061	14.10
3	cd	7.25	0.6879	0.8403	11.87
4	de	7.06	0.4955	0.6053	11.96
5	ef	7.13	0.4962	0.6061	14.10

Table 4 Ripple content of torque and $I_{\alpha\beta}$, derating factor for loss of two non-adjoining phases

S.no	Phase condition	%ripple content of torque for first five cycle	$I_{\alpha\beta}$(A)	Derating factor	%ripple content of $I_{\alpha\beta}$
1	ac	7.25	0.6879	0.8403	11.87
2	ad	6.89	0.7289	0.8904	7.74
3	ae	6.94	0.4955	0.6053	11.50
4	af	7.25	0.6879	0.8403	11.87
5	bd	6.94	0.4955	0.6053	11.50
6	be	6.98	0.2749	0.3358	11.24
7	bf	7.13	0.4962	0.6061	14.10
8	ce	7.13	0.4962	0.6061	14.10
9	cf	7.98	0.7293	0.8909	13.52
10	df	7.25	0.6879	0.8403	13.52

The combination of loss of remaining two non-adjoining and adjoining phases condition with their respective value of torque ripple, $I_{\alpha\beta}$ and the derating factor, is given in Table 3 (adjoining phases) and Table 4 (non-adjoining phases), respectively.

D. *Loss of three phases condition and three healthy phases*

All possible combinations of loss of three phases which are separated by 60 °E and loss of each three phases which are separated by 120 °E are shown in Fig. 16 and their phasor voltages in Fig. 17.

Figures 18 and 19 show the waveform of torque and $I_{\alpha\beta}$ under the loss of three adjoining phases (a, b, c) and non-adjoining phase (a, c, e), whereas the waveform of torque and $I_{\alpha\beta}$ for phases (a, b, d) is shown in Figs. 20 and 21, respectively. The ripple content of torque and $I_{\alpha\beta}$ for all combinations mentioned above with their derating factor is given in Table 5.

4 Simulation Result

The objective of this section is to examine the benefits of the proposed method through simulation results. For this, the various topologies of the developed SSPIM model are

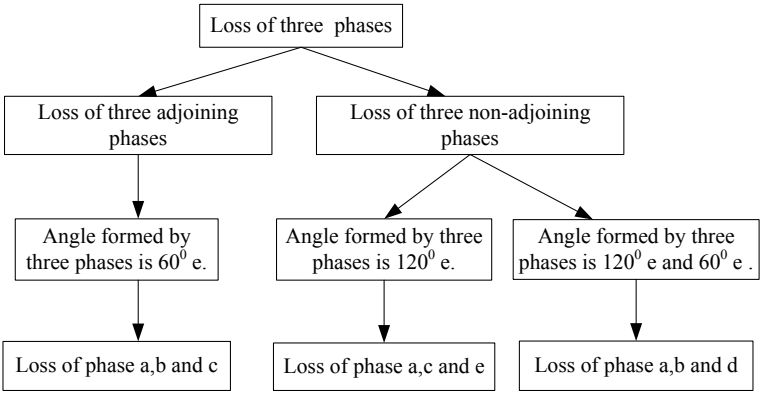


Fig. 16 Flowchart of loss of three-phase condition

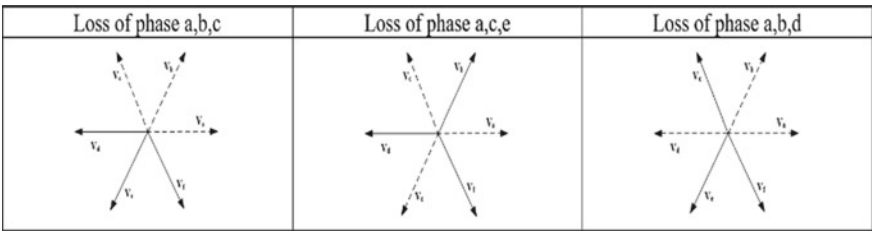


Fig. 17 Voltage phasor of SSPIM with loss of three-phase condition

Fig. 18 Torque subjected to loss of phases (a, b, c) and phases (a, c, e)

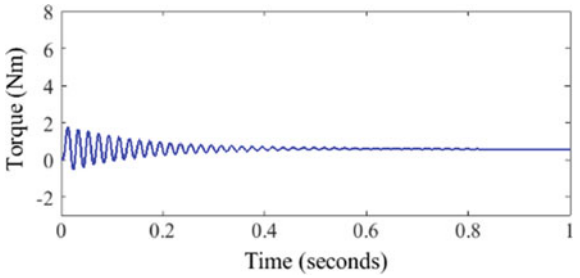
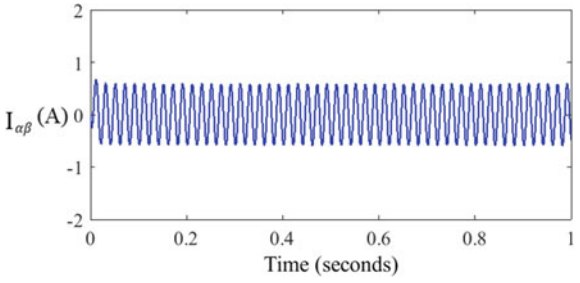


Fig. 19 $I_{\alpha\beta}$ subjected to loss of phases (a, b, c) and phases (a, c, e)



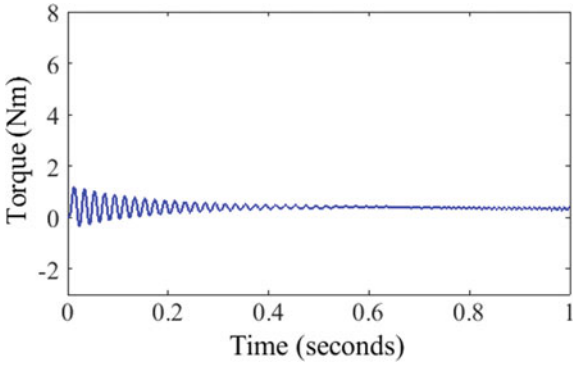


Fig. 20 Torque subjected to loss of phases (a, b, d)

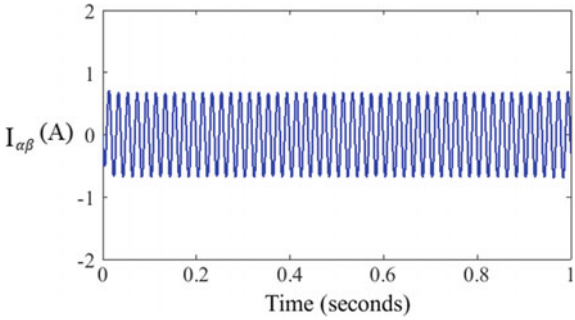


Fig. 21 $I_{\alpha\beta}$ subjected to loss of phases (a, b, d)

Table 5 Ripple content of torque and $I_{\alpha\beta}$, derating factor for loss of three phases

S.no	Phase condition	%ripple content of torque for first five cycle	$I_{\alpha\beta}$(A)	Derating factor	%ripple content of $I_{\alpha\beta}$
1	abc	7.07	0.4132	0.5047	10.38
2	ace	7.07	0.4132	0.5047	10.38
3	abd	6.73	0.4791	0.5852	6.34

simulated in MATLAB/Simulink environment. The proposed model is tested under the different loss of phase conditions, and results obtain from different combinations are tabulated with their derating factor. Parameters of simulated SSPIM are shown in Table 1.

5 Conclusion

A comparative analysis is done under the various loss of phase condition with respect to the healthy phase. Also, ripple content in the torque and current of subspace $I_{\alpha\beta}$

is calculated. The effective value of the derating factor for the different cases is analyzed with respect to a pre-fault condition. It has been concluded that under the loss of phase condition, percentage of ripple content increases due to which losses in the motor are increasing as a result of which motor gets imbalance but continues to run. Also, the derating factor of the machine gets reduces which prone to degrade the life of the motor.

References

1. Levi E, Bojoi R, Profumo F, Toliyat HA, Williamson S (2007) Multiphase induction motor drives—a technology status review. IET Elect Power Appl 1(4):489–516
2. Bose BK (1986) Power electronics and AC drives. Prentice-Hall, Englewood Cliffs, New Jersey
3. Levi E, Barrero F, Duran MJ (2016) Multiphase machines and drives—Revisited. IEEE Trans Ind Electron 63(1):429–432
4. Singh GK (2002) Multi-phase induction machine drive research—a survey. Electric Power Syst Res 61(2):139–147
5. Kumar P, Yadav KB (2013) Torque profiles of asymmetrically wound six-phase induction motor (AWSP-IM) under phase-loss conditions. IOSR J Electric Electron Eng (IOSR-JEEE) 7(2):2278–1676
6. Munim WNWA, Duran MJ, Che HS, Bermúdez M, González-Prieto, I, Abd Rahim N (2013) A unified analysis of the fault tolerance capability in six-phase induction motor drives. IEEE Trans Power Electron 32(10):7824–7836
7. Miranda RS (2013) Current control of six-phase induction motor under open phase fault condition. In: 2013 international electric machines & drives conference
8. Pantea A et al (2015) Six-phase induction machine model for electrical fault simulation using the circuit-oriented method. IEEE Trans Ind Electron 63(1):494–503
9. Kumar P, Yadav KB (2013) Performance prediction of multi-phase induction motor. Int J Current Eng Technol 3(5):1756–1762
10. Miranda RS, Jacobina CB, Lima AMN (2009) Modeling and analysis of six-phase induction machine under fault condition. Power Electron Conf 824–830
11. Zhao Y, Lipo TA (1996) Modeling and control of a multi-phase induction machine with structural unbalance. IEEE Trans Energy Conver 11(3):578–584
12. Rathore V, Dubey M (2012) Speed control of asynchronous motor using space vector PWM technique. Int J Electr Eng Technol (IJEET) 3(3):222–233
13. Klingshirn EA (1983) High phase order induction motors—Part I–Description and theoretical consideration. IEEE Trans Power Apparatus Syst PAS 102, 47–53
14. Abdel-Khalik AS et al (2014) Calculation of derating factors based on steady-state unbalanced multiphase induction machine model under open phase (s) and optimal winding currents. Electric Power Syst Res 106:214–225
15. Gerard A, Humberto H (2009) Experimental Analysis of the d-q-o stator current component spectra of a 42 V fault-tolerant six-phase induction machine drive with opened stator phases. In: Diagnostics for electric machines, power electronics and drives, pp 52–57
16. Martin JP, Meibody-Tabar F (2000) Multiphase permanent magnet synchronous machine supplied by VSIs, working under fault conditions. IEEE Ind Appl Conf 3(8–12):1710–1717
17. Parsa Leila, Toliyat Hamid A (2007) Fault-tolerant interior-permanent-magnet machines for hybrid electric vehicle applications. IEEE Trans Veh Tech 56(4):1546–1552
18. Jacobina CB, Miranda RS, de Corrêa MB, Lima AMN (2004) Disturbance-Free operation of a six-phase AC motor drive system. In: IEEE power electronics specialist conference. Aachen Germany

Verilog Implementation of Biometric-Based Transmission of Fused Images Using Data Encryption Standards Algorithm

Ghanshyam Gaur, Janki Ballabh Sharma, and Lokesh Tharani

Abstract Biometric scan images are used for authentication purposes widely. Because they are unique in nature, the use of biometric scan images is increasing day-by-day in authentication and security. Due to the economic-based systems such as banking and financial sectors are using biometric-based security and authentication, the use of biometric images is increasing every day. In this paper, the fingerprint of the user is taken as key image for DES algorithm and it is used for the encryption of fused image. Hardware implementation of data encryption standards algorithm is designed. For enhancing the target information content in image, using MATLAB software, two multi-focus images are fused. The dual-tree-complex wavelet transform is used for the fusion of images. The comparison is done with other hardware designs in terms of number of LUTs and power and complexity of implementation with other encryption techniques. This paper gives the secure transmission of fused images using DES encryption, with the use of biometric image as a key image, which is strong and unique method for security and authentication in nowadays.

Keywords Image fusion · Wavelet transform · Data encryption standards · Encryption · Decryption · S-box · Verilog hardware description language

1 Introduction

Biometric scan images such as fingerprint, face, iris, thumb, and hand impression are used for authentication and security purposes widely. Because they are distinctive in nature, its use is increasing day-by-day in authentication and security. Due to the

G. Gaur (✉) · J. B. Sharma · L. Tharani
Department of Electronics, University Department RTU Kota (Rajasthan), Rajasthan Technical University (Kota), Kota 324010, India
e-mail: gaurghanshyam08@gmail.com

J. B. Sharma
e-mail: jbsharma@rtu.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_41

economic-based systems such as banking and financial sectors are using biometric-based security and authentication of the use of biometric images is increasing every day. We use dual-tree complex wavelet transform (DT-CWT) here to fuse two multi-focus images.

There are various fusion techniques available today, and their utilization based on requirement is chosen by user. Here in this paper, DT-CWT is used because of its quality as compared to other fusion techniques like averaging method, Pyramid method, curvelet, DWT, and SWT. After the DT-CWT-based fusion operation, DES encryption is applied to this fused image, using the biometric scan image of user such as fingerprint as a key image. Recovery of securely transmitted image is possible by using the same authorized key image [1, 2]. Otherwise, the original fused image is not obtained and this is the main aim of this work done in this paper. Verilog implementation of DES using Xilinx ISE tool is done; design of the hardware and other parameters, like timing, power, and number of LUTs, slices, are obtained. The comparison is done with other hardware designs in terms of number of LUTs and power and complexity of implementation with other encryption techniques like triple DES, AES, and Blowfish [3]. Here the goal of this paper is to provide biometric-based secure transmission of images using encryption technique, as well as to reduce complexity of algorithm used in encryption in terms of key size and block size [4], which are the parameters affecting the area and power distribution and the ease of implementation. Biometric-based secure transmission is strong and unique method for security and authentication in nowadays.

1.1 Related Work

The various fusion techniques have been used by other authors such as averaging pixels, maximum, PCA, and pyramid decompositions and transform-based techniques. Elbirt (et al.) gives the complete details about image fusion and its principles, methods, and applications, and has evaluated FPGA implementation and performance [2].

The details of image fusion techniques based on wavelets are provided by Krista Amolims (et al.). For the segmentation of input image features, a dual-tree complex wavelet transform is used, either jointly or separately to create a map of region [4].

New updated version of the AES-based image encryption algorithm is proposed by Kamali (et al.) [6].

The image encryption algorithm was proposed using the chaotic map and S-DES by Bin and Liu (et al.) [7]. Agarwal and Bedi suggested the use of hybrid image fusion technique to enhance the function of clinical diagnosis [11].

2 Preliminaries

DWT-based approaches have now largely replaced previous fusion methods based on decompositions of pyramid [4]. It is possible to explain the fusion of two test images using the discrete wavelet transform in terms of the 2 recorded input test images S_0 and S_1, the WT is **ω,** and the fusion rule is φ, which is taken as symbol to combine colocated coefficients within the domain of transform. And coefficients of the fused wavelet are inversed to generate the resulting fused image IF using inverse wavelet transform $\boldsymbol{\omega - 1}$ hence [4]

$$\mathbf{IF} = \boldsymbol{\omega} - \mathbf{1}(\varphi(\boldsymbol{\omega}(\mathbf{S}0), \boldsymbol{\omega}(\mathbf{S1}))).$$

2.1 Image Fusion Based on Dual-Tree Complex Wavelet Transform (DT-CWT)

While image fusion based on DWT offers nice results, the associated shift variance has been recognized to produce sub-optimal performance, which is because of the magnitude of a coefficient does not adequately reflect the actual energy due to position of spatial frequency. Hence, it is not a reliable base for comparing perceptive significance inside the domain of transform [3, 4]. Using transforms such as shift-invariant discrete wavelet transform (SI-DWT) has addressed the problem [12]. The transformation is similar to the DWT, but at each point of the decomposition, it eliminates down-sampling. Since down-sampling allows the variance of the shift, the SI-DWT is shift-invariant.

This differentiation (of + and − frequencies) leads to splitting the original three DWT sub-bands into 6 DT-CWT sub-bands based on the orientations of +15/−15, +45/− 45, and +75/− 75, respectively.

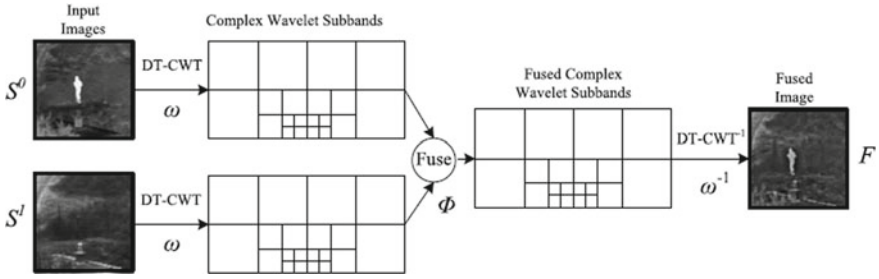


Fig. 1 Image fusion of test images using DT-CWT [15]

The sub-band coefficients resulting are complex, and the magnitude of each coefficient is about the invariant of changes. Figure 1 displays a typical fusion scenario.

Figure 1 shows that it is possible to create and transform coefficients from test images. The fused transformed is achieved. The most popular fusion rule is for the coefficient of wavelets which is max rule of fusion [13].

Input : Multi-focus Images: I_1, I_2

$W^0 \leftarrow$DT-CWT(I_1)

$W^1 \leftarrow$DT-CWT(I_2)

$W^2 \leftarrow$Max-Coefficient Fusion(W^0,W^1)

Initial fused DT-CWT transform

F=DT-CWT^{-1}(W^2)

Output: Fused image I_F

The selection of max rule of fusion or its variants presupposes that a coefficient's perceptual value is directly related to its magnitude [14, 15].

Test images were first processed using DT-CWT for getting transformed coefficients. Theses coefficients are further fused using fusion rule. The inverse DT-CWT is then applied over fused coefficients to get fused image F.

2.2 Data Encryption Standards

This is a block cipher of archetypal-type. This algorithm accepts the input of 64-bit block size and converts it into ciphertext of same length.

DES also uses a key to customize the encryption in such a way that only those who know the particular key used to encrypt are able to decrypt. The secret key consists theoretically of 64 bits; however, the algorithm actually only uses 56 of those bits. Figure 2 displays the entire DES algorithm [1, 3, 5].

A. **S-box**

DES S-boxes differ for all other permutations. [9] The structure of S-box design is shown at the end of Fig. 3.

B. **Verilog implementation for S-box**

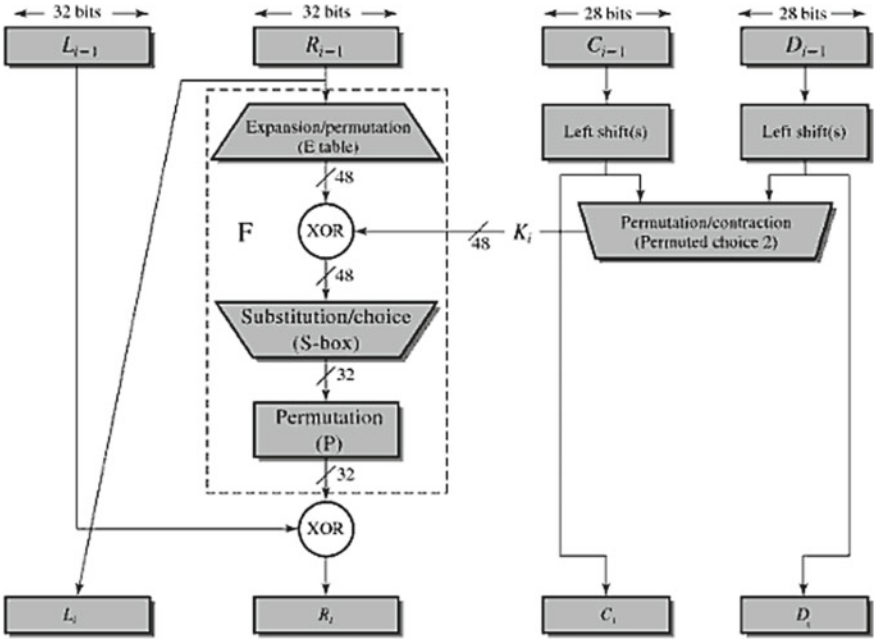


Fig. 2 DES algorithm [1]

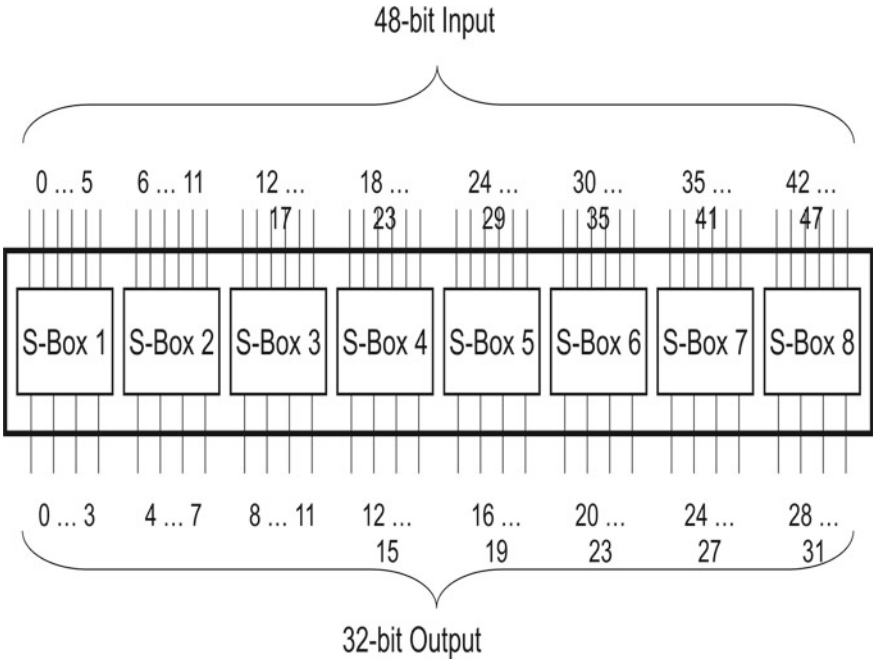


Fig. 3 Eight DES S-boxes [7]

B. Verilog implementation for S-box

```
module Sbox1(in, out);
input [1:6] in;
output [1:4] out;
reg [1:4] out;
always @(in) begin
  case ({in[1], in[6], in[2:5]}) //synopsys full_case
parallel_case
0: out = 14;
1: out = 4;
2: out = 13;
3: out = 1;
4: out = 2;
5: out = 15;
6: out = 11;
7: out = 8;
.
63: 0ut= 13;
  end case
  end
module
```

Similarly we apply it for the other 7 s-boxes.

3 Proposed Architecture

The whole sequence of this work is organized into three parts. The first part is presented using the block diagram shown in Fig. 4. In this part, we performed DT-CWT on two test images, which results in fused image I_F [16, 17]. DT-CWT-based fused image I_F is encrypted using DES algorithm. The unique biometric scan image like fingerprint, iris, or hand impression is used as a key image I_K in order to secure transmission of fused image and also in storage which is shown in Fig. 5. The same key image (fingerprint) is used for DES decryption process, when the user needs his original fused image. After DES decryption process, the original fused image is recovered as shown in Fig. 6. If the key image used is not authorized, then the original fused image is not recovered after decryption. This ensures the secure transmission.

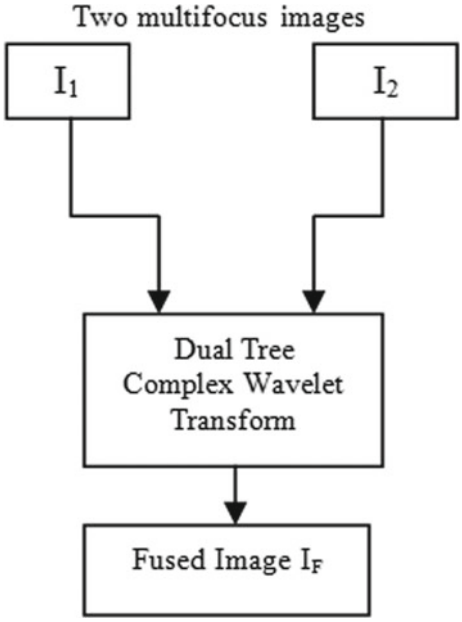


Fig. 4 DT-CWT-based image fusion

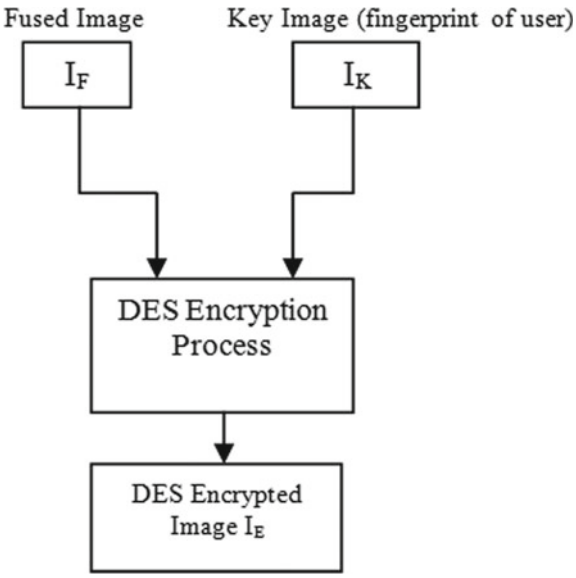


Fig. 5 DES algorithm

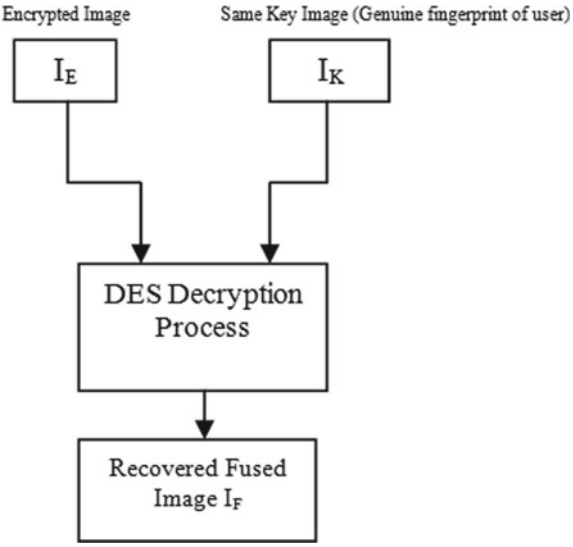


Fig. 6 DES decryption

Biometric images Iris

Fingerprint

Hand impression [11]

4 Assessment Environment

In this paper, we use MATLAB and Xilinx programming. We now use the VLSI format here. Hence, hardware specification language is used to create this design is Verilog HDL. Similarly, requirements and specialties of Verilog make popular for different performance and speed of design implementation. Xilinx ISE 14.2 Design Suit is used as simulation software as well as for synthesis layout [8, 12]. Likewise, software and hardware are used and it is simple to implement on them.

5 Implementation Results

The first result is found using DT-CWT of two multi-focus test images as shown in Fig. 7 using MATLAB, and hence, fused image is found as indicated in Fig. 8. After this, fused image as proposed is encrypted using fingerprint of user with DES method in Xilinx ISE 14.2 Verilog implementation.

These two test images are taken, and they are fused with DT-CWT method in MATLAB of the fused image is stored, and in the proposed scheme, it is encrypted with other key image which is the biometric data of the user and it may be his fingerprint, iris, thumb impression, or hand impression. Both fused image and key image are encrypted using DES method. DES is implemented on Xilinx ISE 14.2. The proposed results are shown in Fig. 9.

After DES, we get encryption result of fusion image and biometric image and then it is used for secure transmission and storage. After decryption using the same unique fingerprint image as a key, we get fused image again. Figure 10 shows the implementation on Xilinx ISE 14.2. The proposed scheme is tested for single block using DES. The input data stream is 64-bit single block, and the key is 56-bit single block as per DES operation. Using this entire programming environment, we get fused image with more information from test images. Further, we get chip-level solution using Verilog HDL implementation for security with DES algorithm.

Fig. 7 Two multi-focus test images [13]

Fig. 8 Fused image I_F

Figure 9 shows the proposed results for Verilog implementation of biometric-based transmission of fused images using DES.

Analysis of Power Distribution in Chip:

The power distribution analysis is essential in the design because the leakage power and dynamic power of the designed chip are also accountable in optimum design that is shown in Table 1. It gives the distribution of power consumed in clocks, logics, signals, and inputs/outputs (IOs).

a) Fused Image I_F b) Fingerprint c) DES Encrypted image d) Decrypted image

Fig. 9 Proposed scheme

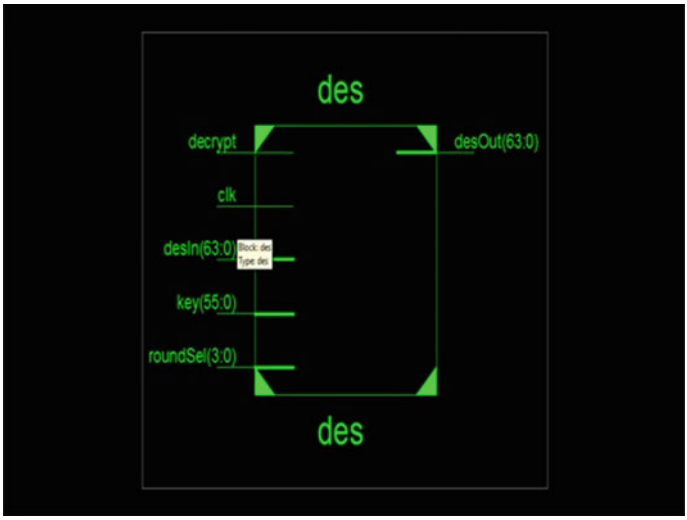


Fig. 10 Xilinx ISE 14.2 implementation

Table 1 Power distribution in chip

On chip	Power(W)	Used
Clocks	0.002	1
Logic	0.003	437
Signals	0.008	673
IOs	0.131	190
Leakage	0.041	–
Total	0.184	–

Table 2 Comparison table for power distribution

Design	Power (W)
Sumanth Kumar Reddy S et al.	0.655
Ours	0.184

Table 3 Encryption algorithms

Algorithm	Key Size (Bits)	Block Size (Bits)
DES	64	64
3DES	192	64
AES (Rijndael)	256	128
Blowfish	448	64

The power distribution window is generated by clicking appropriate option of *analyze power distribution* in the Xilinx ISE 14.2

6 Comparisons

Here the comparison is made with the other author's results such as Sumanth Kumar Reddy et al. [10]. Table 2 shows the power distribution comparison.

Further Table 3 shows the generally used other algorithm such as triple-Data Encryption Standards (DES), Advanced Encryption Standards (AES) (Rijndael), and Blowfish. Also we see in this DES algorithm with respect to number of bits used, the key size and block size are minimum as used in other algorithms.

Therefore, it is obvious from the above comparison that this design comparison in proposed scheme in terms of power distribution gives better results as compared to other work done. Also, it utilizes a fix 64-bit key and allows 64-bit block size, and therefore, making it easier to implement on Verilog HDL in Xilinx ISE14.2 tools.

7 Conclusions

Each person has a unique identity, and their biometric image is stored with their records in a database. The DES key which is a biometric data is not accessible to unauthorized user. So it becomes more secure. Same key is used for decryption of transmitted fused images. Only authenticated users can access their data after giving their fingerprint or iris scan genuine information, etc. Hence, we conclude that image fusion is very useful in increasing the information content of target image and DES algorithm performed secure transmission of fused image.

References

1. National Bureau of Standard-Data Encryption Standard. FIPS Publication 46 (1977)
2. Elbirt.: An FPGA Implementation and Performance Evaluation of The Cast-256 Block Cipher. Cryptography and Information Security Group, ECE Department, Worcester Polytechnic institute, Worcester, Ma, Technical Report (1999)
3. William S.: Network Security Essentials (Application and Standards). Pearson Education (2004)
4. Krista, A., Zhang, Y., Dare, P.: Wavelet Based Image Fusion Techniques- An Introduction, Review and Comparison. Published by Elsevier (2007). http://dx.doi.org/10.1016/j.isprsjprs.2007.05.009
5. Monicalibertori, Fernando, O., Bonadero, J.C., Castineira, J.: AES-128 Cipher High Speed, Low Cost FPGA Implementation, pp. 195–198. IEEE (2007)
6. Kamali, S.H., Shakerian, R., Hedayati, M., Rahmani, M.: A New Modified Version of Advance Encryption standard Based Algorithm for Image Encryption, Volume-I, VI-141 to VI-145. ICEIE (2010)
7. Bin, L., Liu, L., Jan, Z.: "Image Encryption Algorithm Based on Chaotic Map and S-DES, pp. 41–44. IEEE (2010)
8. Deergh Rao, K., Gangadhar, C.:VLSI Realization of a Secure Cryptosystem for Image Encryption and Decryption, pp. 543–547. IEEE (2011)
9. Zhang, Y-H.: Image encryption using extended chaotic sequences. In: Fourth International Conference on Intelligent Computation Technology & Automation, pp. 143–146 (2011)
10. Sumanth Kumar Reddy, S. Sakthivel, R., Praneeth, P.: Int. J. Adv. Eng. Sci. Technol. (IJAEST) **6**(1), 022–026 (2011)
11. Agarwal and Bedi (2015) Implementation of hybrid image fusion technique for feature enhancement in medical diagnosis. Human-centric Comput. Inf.n Sci. 5:3. https://doi.org/10.1186/s13673-014-0020-z Springer
12. Moonon, A.-U., Hu, J., Li, S.: Remote sensing image fusion method based on nonsubsampled shearlet transform and sparse representation. Sens. Image. Int. J. **16**(1), 1–18 (2015)
13. Amita, N., Hamurabi, G.R.: Enhanced image fusion using directional contrast rules in fuzzy transform domain. Springer Plus **5**, 1846 (2016). https://doi.org/10.1186/s40064-016-3511-8
14. Paul, S., Sevcenco, I.S., Agathollis, P.: Multi-exposure and multi-focus image fusion in gradient domain. J Circu. Syst. Comput. **25**(10). https://doi.org/10.1142/S0218126616501231
15. Hill P, Al-Mualla ME, Bull D (2017) Perceptual image fusion using wavelets. IEEE Trans Image Process 26(3):1076–1088
16. Hamza O.I., Ibrahim O.S., Gorkem S., Nizamettin A.: The effect of nonlinear wavelet transform based de-noising in sperm abnormality classification. In: Computer Science and Engineering (UBMK) 3rd International Conference, pp. 658–661 (2018)
17. Asha, C.S., Shyam, L., Varadraj, P.G., Prakash Saxena, P.U.: Multi-modal medical image fusion with adaptive weighted combination of NSST bands using chaotic grey wolf optimization. Access IEEE **7**, 40782–40796 (2019)

Study and Analysis of Test Case Prioritization Technique

Dharmveer Kumar Yadav, Sandip Dutta, and Chandrashekhar Azad

Abstract Context: The purpose of regression testing is to validate the altered program and identify the adverse affect in program due to modification. The technique of test case prioritization is used to plan test instances to maximize certain objectives. Regression testing is mainly a maintenance activity. The primary goal behind this study and analysis is to provide a basis for conducting research in the area of prioritization of regression testing. Objective: This paper examines each technique of the prioritization process and addresses the open problem and possible directions for future study. A number of academic papers from 2000 to 2018 were evaluated in this study. We explored the database of Google Scholar, IEEE explores, Springer, ACM, Science Direct, and Inspec.

Keywords Test case prioritization · Regression testing · Software testing

1 Introduction

In the regression testing, test cases are re-executed. In that order, run the test cases in order to identify the faults in the program as soon as possible within a period of time. During the maintenance stage, this is a significant activity as it builds confidence that the modified code is correct. This article provides a study on prioritization of the regression testing. Regression testing is done when the software is modified. The

D. K. Yadav (✉) · S. Dutta
Computer Science Department, Birla Institute of Technology Mesra, Ranchi, India
e-mail: kumar.dharmveer@gmail.com

S. Dutta
e-mail: sandipdutta@bitmesra.ac.in

C. Azad
Department of Computer Applications, National Institute of Technology Jamshedpur, Jamshedpur, India
e-mail: csazad.ca@nitjsr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_42

purpose of regression test is to make sure that any introduced modifications do not interfere with the behaviors of the original unchanged software program.

The prioritization of test case methods enhance testing cost-effectiveness and make a number of benefits, like early identification of faults and early feedback system that helps to testers. Various research provides proof that regression testing may benefit from the prioritization method [1–3]. Many methods are code-based coverage, and there is very few techniques for design-based methods [4]. It is possible to distinguish test case prioritization methods using the following information [5]:

The code coverage: Prioritization is based on reference to the number of statements and fundamental blocks. Total method coverage: This method schedules the test cases if the test case covers the methods of the program [5]. Use feedback: The use of code modifications information. Other information: Many researchers presented prioritization using other information like estimating the test costs, fault severity, data on test history [5].

Although the literature contains many prioritization techniques, very few systematic studies on the present research prioritization technique are presented. This area's research is timely owing to the increasing knowledge base on TCP issues. Academics and practitioners can use the findings of this paper. Researchers can identify techniques to prioritize and implement test cases in their tools. The purpose of this study is to examine the prioritization method of the test case. We conducted a systematic survey for this purpose and summarized related to a prioritization technique.

Given the fact that the literature contains various TCP methods, there are no recent progressive literature reviews that illustrate the current significance of prioritization in research software testing. This study, therefore, tries to conduct a systematic review of the literature on the recent prioritization technique [6].

The literature review is an expert, uncompromising research study [7]. The purpose of a review paper is not to merely summarize all present evidence based on study issues, and it is also anticipated to enhance the researcher's advice on evidence-based studies. This review paper of the literature is organized as follows. A systematic study of the prioritization is proposed from 1997 to 2011 in the regression test case prioritization research [8]. From 65 research, 49 have been recognized in their job to initiate a distinct strategy, two to increase previous research, and 14 to re-analyze previously tested results.

Around eight different prioritization methods were also evaluated and recognized by the review paper. These consist: genetic-based, requirement, coverage history, fault, composite, and other techniques.

The author focuses on the area of regression testing that involves test suite minimization, selection of test case, and prioritization technique [9].

The present state-of-the-art methods were categorized into several classifications for prioritization. These methods were classified based on requirements, model, code coverage, historical information, probabilistic, and other approaches. The authors survey 159 studies from 1977 to 2007 period.

Section 2 describes related work of prioritization technique. Section 3 presents the research method which consists of research questions, motivation, source of information, the process of study selection, and data extraction. Next, Sect. 4 is

about the prioritization technique. Section 5 describes the result and analysis section based on various research questions (RQ). Section 6 describes the research findings. In Sect. 7, presents the conclusion part of the survey paper.

1.1 Test Case Prioritization

The prioritization of test cases performs running the test case in order to achieve some goals. Prioritization of test cases addresses a severals goals, including:

1. Testers may want to improve the faults detection rate, which is the probability of exposing faults in regression testing.
2. Testers may improve the capability to detect the high-risk faults.
3. Testers may increase the chance of discovering errors in the regression testing method related to specific code changes.
4. Testers may improve their coverage of easily covered code at a faster pace in the scheme being tested.
5. Testers may improve their confidence at a faster speed in the reliability of the system being tested.

2 Related Work

Researchers presented a survey on minimizing, selecting, and prioritizing regression testing [10]. They explored many papers from year 1999 to 2009. Prioritization is performed using the following categories:

Coverage based: This approach enhances structural coverage. This method helps to increase the detection of faults. This method allows to execute maximum number of statements in the program, blocks, and method. Hence, maximum fault can be detected by this approach.

Human based: This type of approach uses a human tester. Prioritization is based on human tester's comparisons.

Requirements based: Requirements are used in this method like a customer who may assign priorities for the requirements, and developer requirement complexity can be used [11].

Model based: Different models are used instead of code blocks, such as UML sequence or activity diagrams. The UML is the standard modeling software.

Catal proposed systematic survey paper for prioritization [12]. They searched paper fron online repositories papers: IEEE Explorer, Science Direct, ACM Digital Library, Springer, and Wiley. They first searched for string extracted 202 research, but 120 papers were identified for prioritizing test cases after further evaluation of titles and abstracts of papers. The literature consists several techniques, with coverage-based method using statements and basic blocks. Model-based method is also used

by many researchers. The finding of [12] is recommendations for researchers that are provided:

(1) Give preference instead of proprietary datasets to public datasets.
(2) Introduce many model-based techniques for prioritization
(3) Further research on trying to compare prioritization methods
(4) Assess the performance of proposed approach with metrics and compare their results.
(5) Use industrial project datasets.

Author [13] proposed both a survey and analysis of selection, minimization, and prioritization technique.

In paper [14] proposed a systematic review of prioritization of test case research. Sixty-five papers and fifteen case studies have been found. Approx 106 different approach has been analyzed. The proposed method was also analyzed and compared with different measures such as type of study, technique, input, and metrics. Their research finding shows that fifty percent of prioritization techniques are not dependent on language implementation. Future research should focus on filling the gap found in the existing technique. A survey on test case prioritization is presented [15]. The finding of this paper is that a total of 90 scholarly articles in this paper from the year 2001 to 2018 have been analyzed. They explored various databases. They found each method of prioritization with findings and subject programs. Their finding research was based on three types of research questions.

Researchers used repositories to extract various proposed techniques [16]. They searched according to criteria defined and classified in the categories of journals, conference papers, symposiums, and workshops. The review strategy nominated 69 primary studies. They used 40 journals, 21 conference papers, that is used for the result. Each prioritization strategy has prospective values, benefits, and limitations indicated. They explored the systematic literature review on the test case prioritization method.

3 Research Method

A structured method for performing this review paper with a particular end objective. The study was conducted to examine TCP-related research [6, 7, 17] influenced the systematic and organized technique. There are four main methods involved in this review paper, which consists of research question, research motivation, source of information, and process of study selection. The main study issues were presented in the first phase to answer to the primary objective of this paper review. Next, the choice was made of the corresponding repositories. This is reduced by using a search approach that includes specifying search strings and search methods that were scheduled on the basis of the study issues articulated. The search phase output was then transferred to the research studies.

3.1 Research Questions

We will discuss answer of the following research questions.

RQ1-The most researched methods for prioritization: RQ2-The most frequently used methods for research:

RQ3-The evaluation metrics most frequently used for TCP: RQ4- What are the programs frequently used in prioritization research?:

RQ5-What are the processes involved in TCP?:

(1) *Research Motivation*: Identify TCP prioritization techniques developments and opportunities.

Identify widely used study subjects in TCP articles. Identify new deficiencies in the research methods used in TCP articles.

Identify the most considerable journal for the priority test case.

Identify the most used evaluation metrics for prioritization.

3.2 Source of Information

As Kitchenham proposed in [18], searching provides more searching space. In accordance with the guidelines, to cover the maximum possible information, the following databases were explored for journals and conference proceedings.

- Google Scholar
- Springer LNCS
- IEEE eXplore
- ACM digital library
- Science Direct
- Inspec.

3.3 Process of Study Selection

Initially, all sources researched possibly relevant articles. Because of the use of string-like regression and testing, the elementary search produced an enormous quantity of literature.

When the systematic review was selection techniques for regression testing, in the next phase, software or regression testing title-based exclusions were made for articles that were meaningless. The author suggested consideration of papers irrespective of their language, but in any language other than English, we excluded papers [19]. Then, rejections involve based on the abstract of the research paper. software testing, selection technique, reduction technique, and hybrid method were

not selected for the study. Only prioritization technique related papers in included for study purposes.

3.4 Data Extraction

Data were extracted to answer the research questions from the primary research studies. The data was collected and stored in MS Excel to find related information for study and analysis purposes.The following information was collected from the paper.

- Research Paper Title
- Publication year
- Testing method
- Research questions
- Different prioritization method
- Limitations in this field
- Various metrics used for evaluation
- Result obtained
- Answers to research question
- Future Work
- Findings
- Conclusions.

4 Prioritization Technique

Prioritization of test case includes running the test cases in some order to meet the particular objective, such as, fault detection rate, code coverage, etc. The key goal of this approach is priority setting is to minimize the cost of regression testing. Use the prioritization technique for test case, the efficiency of testing procedure can be improved.

The current discussion described in the sections below also provides a answer to research question by specifying the comparative methods.

4.1 Background: Test Case Prioritization

The test case prioritization approach schedules the test case so that greatest advantage can be achieved by the tester using prioritization. The methodology was first presented by [20]. proposed the methodology in a progressively broad area in prioritization [21], which was experimentally evaluated by the author [22]. Consider, an

Table 1 Test case detects faults

Test case	Fault detected by test case		
	$T1$	$T2$	$T3$
A	*		
B	*		
C	*	*	*
D		*	*
E	*		

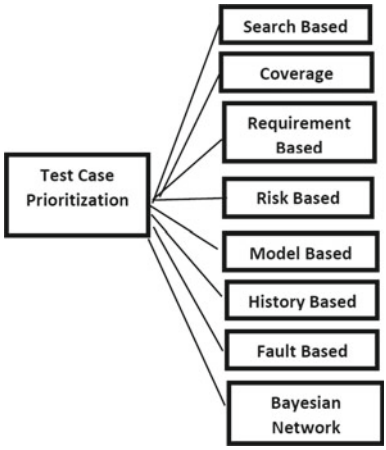


Fig. 1 Test case prioritization technique

example illustrated in Table 1. In this case, fault detection is known before. Prioritization is aimed at maximizing the early detection of faults. The test case ordering A-B-C-D-E shows fault detected by the test cases. The effectiveness of this algorithm is calculated through metrics. Many researchers used the APFD metric (Fig. 1).

(A) *Positioning Figures and Tables*

4.2 Prioritization Technique

- Search based: Prioritization was proposed using a search-based approach like GA-based [23–30], greedy algorithm [31, 32], ant colony [33–35], and string distance [36]. The genetic algorithm (GA) technique [37] does not produce good result than a greedy algorithm. The search-based approach is therefore dependent on several such as test suite and fitness function. Using the GA approach, several researchers obtained the best result. The GA method functions slowly [26]. Swarm-based prioritization method is proposed [38].
- Requirement based: A framework is constructed dependent on its necessities, in this manner, by utilizing the prerequisites data it might conceivably help to arrange

a few essential experiments something beyond by utilizing data of code. The work in [39] accepts that prioritization of necessities for the test is a procedure of setting up the estimation of the significance of a product prerequisite. Assessment of port depends on some basic components like customer need, mistake, and execution challenges. Late examination announced that a mix of at least two variables may deliver a superior outcome than a solitary consider prioritization terms of testing viability. Prioritization calculation is based on traceability, changes in requirements, client need. The methodology outcome was in terms of fault detection rate.

- A Risk-based
- Prioritization of the risk-based approach was primarily employed with the software to be built in a project usually associated with risk-related values [40].
- Requirements risk information is used for prioritization and to identify the test cases that were expected to determine the system's risk-related errors [41, 42]. Proposed prioritization of test cases through requirement risk values [43]. Risk value is determined first, then prioritization starts for the test case. During the early stage, it has ability to detects more faults.
- Fault based: Initial method for prioritization using fault is suggested by [44, 45]. According to authors, the fault would be revealed by a test case, depends not only on whether some particular statement is executed but also on the chance that the statement fault cause failure. They suggested two methods total fault exposing potential and additional fault observed by the test case. The author proposed fault-based prioritization approach which detects more fault [46, 47].
- Elbaum et al. [24] proposed six different types of function-level methods for prioritization using faults. Some of the methods are based on function-level fault prioritization, the other two methods used the concept of fault index for the function, and the remaining two methods use a combination of both fault index and FEP. Prioritization using the functional-level method was also proposed [27]. The syntactic differences technique was used between two versions of the code in this approach. They used Unix diiff command to measure the degree of change for the function available in both the version.
- History based: History-based prioritization is performed using historical data. The author [48] claimed that their proposed method helps to increase fault detection capability using historical data. Their approach was an extension of Kim and Porter [49] existing history-based prioritization method. They used eight C programs in their experiments. The new method is proposed in which they observed improvements in fault detection using a modified cost cognizant technique for prioritization [49]. They used Unix utilities sed and flex in their experiments.
- Bayesian Network (BN)
  Bayesian network (BN) method for prioritization of test cases has been proposed [50]. BN model discovers a certain error based on the data available. Their target strategy collects different information from the source code, integrating all data for a single BN model using probabilistic approach. BN model performs better result compared to random method. Test case prioritization proposed using Bayesian

networks [50]. They demonstrated using a combination of three methods called modification, coverage, and fault. The proposed result is compared with the other ten prioritized techniques which consist original, random, and optimal (three techniques) and six other methods using coverage-based technique. From the result, it is observed that the proposed method performs better result in the other technique.

- Cost-aware based
  Prioritization of the test case method is proposed based on the cost of the test case and fault present in the system [51]. The proposed approach produces better results in some of the cases. A cost-aware approach is suggested by Huang [52]. They proposed modified cost-cognizant prioritization technique to find units of faults discovered by the test cases. They used flex and sed of UNIX in their experiments. They compare their results with optimal, random historical fault. Reduce testing costs in their proposed approach.
- Model-based TCP: UML-based technique is presented for prioritization. They used activity diagram(AD), sequence diagram(SD) with rule mining. Test case ere generated through activity sequence graph constructed from AD and SD diagram. They trace affected nodes by forward slicing technique. This work detects faults at an early stage which improves the prioritization method [53]. Static- and dynamic-based techniques are proposed for Java programs. They perform experiments on method- and class-level tests. They used static and dynamic approaches for class- and method-level tests with the APFD metric to calculate the effectiveness of an algorithm [54]. Hybrid technique is proposed for object-oriented program [55]. The proposed work consist source code and UML models to represent the modified classes. Dependency graphs of the program is used in their work [56].

5 Result and Analysis

This section tries to answer the following research questions.

- RQ1-The most researched approach for prioritization: Methods of prioritization based on coverage dominate. Most papers introduced techniques based on coverage. A comparison is performed for the proposed approach with methods based on coverage.
- RQ2-The most frequently used methods for research: The papers were categorized as per the method of applied research, which includes prioritization method survey, experiment, review, and development. Most papers presented a new technique of prioritization. This shows that researchers try to focus primarily on developing new methods of prioritization.
- RQ3-The evaluation metrics most frequently used for TCP:

Many researchers presented new evaluation metrics, called average percentage of detected faults (APFD), relative position (RP), coverage effectiveness (CE), APFD per cost, normalized APFD used to measure the performance of the proposed technique.

- RQ4-What are the programs frequently used in prioritization research?:
- This question is about which programs are used for experimentation with prioritization techniques. Many researchers used C/C+ +/JAVA programs, and Web-based applications have been found that are used for TCP studies.
- RQ5-Processes involved in prioritization: Software engineering is very concerned about how the processes of engineering are applied systematically to software development. It is, therefore, necessary to investigate these questions. Each experiment should have its own process. The prioritization process begins with targeted data.

6 Research Findings

This is our research finding in this study for the test case prioritization technique. Prioritizing test cases in software development is an essential part in the testing process [57, 58]. Developed software must meet the user requirements without any faults. The prioritization technique helps to reduce costs, time, and software faults. The various evaluation metrics are used in these primary studies. APFD metric is used in all approaches for evaluation purposes. Many new metrics were proposed in various proposed works. The percentage of a different approach is shown in Fig. 2.

Coverage based: This technique arrange test based on code coverage.

Various prioritization techniques are classified in this paper.

Human based: This is based on the human tester.

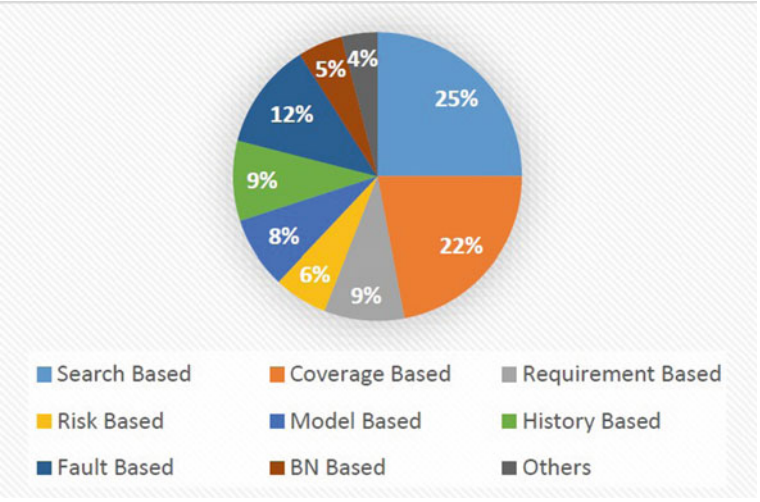


Fig. 2 Studies technique percentage

Probabilistic approach: In this technique, the probabilistic theory method is proposed.
History based: In this method, fault or history information is used for prioritization techniques.
Requirements based: Requirements are considered for prioritization technique.
Model based: Model-based prioritization method is also proposed using the UML diagram.

7 Conclusion

Prioritization of test cases is one of the important activities in regression testing. Numerous researchers have recommended various TCP techniques. This study has been working on analyzing various techniques based on research questions. Study and analysis is done from selected studies, classifying, analyzing, and deducting them. The main reason for this study is to find out which technique can be enhanced in future work. By exposing the latest trends, this study provides a basis for future work for testing community with a deep insight. Researchers may also benefit from this study because it illustrates the limitations and gaps in the current research.

References

1. Do H, Mirarab S, Tahvildari L, Rothermel G (2010) The effects of time constraints on test case prioritization: a series of controlled experiments. IEEE Trans Softw Eng 36(5):593617
2. Elbaum SG, Malishevsky AG, Rothermel G (2000) Prioritizing test cases for regression testing. In: Proceedings of the 2000 ACM SIGSOFT international symposium on software testing and analysis (ISSTA 2000). ACM, New York, NY, pp 102112
3. Elbaum S, Gable D, Rothermel G (2001) Understanding and measuring the sources of variation in the prioritization of regression test suites. In: Proceedings of the 7th international symposium on software metrics (METRICS 01). IEEE Computer Society, Washington, DC, USA
4. Korel B, Koutsogiannakis G (2009) Experimental comparison of code-based and model-based test prioritization. In: Proceedings of the IEEE international conference on software testing, verification, and validation workshops (ICSTW 09), pp 7784
5. Do H, Rothermel G, Kinneer A (2006) Prioritizing JUnit test cases: an empirical assessment and cost-benefits analysis. Empirical Softw Eng 11(1):3370
6. Kitchenham B (2004) Procedures for performing systematic reviews, vol 33, no. TR/SE-0401. Keele University, Keele, UK, p 28
7. Kitchenham B, Brereton OP, Budgen D et al (2009) Systematic literature reviews in soft- ware engineering a systematic literature review. Inf Softw Technol 51(1):715
8. Singh Y (2012) Systematic literature review on regression test prioritization tech- niques difference between literature review and systematic literature. Informatica 36:379408
9. Yoo S, Harman M (2007) Regression testing minimisation, selection and prioritisa- tion : a survey. Test Verification Reliab: 17
10. Yoo S, Harman M (2012) Regression testing minimization, selection and prioritiza- tion: a survey. Softw Test Verification Reliab 22(2):67120
11. Srikanth H, Hettiarachchi C, Do H (2016) Requirements based test prioritization using risk factors: an industrial study. Inf Softw Technol 69:7183
12. Catal C, Mishra D (2012) Test case prioritization: a systematic mapping study. Softw Qual J 21 (3): 445478; Kumar A, Singh K A literature survey

13. Yoo S, Harman M (2007) Regression testing minimisation, selection and prioritisation : a survey. Softw Test Verif Reliab 17. https://doi.org/10.1002/000
14. Singh Y, Kaur A, Suri B, Singhal S (2012) Systematic literature review on regression test prioritization techniques. Informatica 36:379408
15. Mukherjee R, Patnaik KS A survey on different approaches for software test case prioritization. J King Saud Univ Comput Inf Sci
16. Khatibsyarbini M, Isa MA, Jawawi DNA, Tumeng R (2018) Test case prioritization approaches in regression testing: a systematic literature review. Inf Softw Technol 93: 7493
17. Achimugu P, Selamat A, Ibrahim R, Naz M (2014) A systematic literature review of software requirements prioritization research. Inf Softw 56:568585
18. Kitchenham BA (2007) Guidelines for performing systematic literature re- views in software engineering version 2.3. Technical Report S.o.C.S.a.M. Software Engineering Group, Keele University and Department of Com- puter Science University of Durham
19. Dyba T, Dingso yr T, Hanssen GK (2007) Applying systematic reviews to diverse study types: an experience report. In: First international symposium on empirical software engineering and measurement, ESEM 2007, pp 225–234
20. Eric Wong W, Horgan JR, London S, Mathur AP (1998) Eect of test set minimization on fault detection eectiveness. Softw Pract Exper 28(4): 347369
21. Harrold MJ (1999) Testing evolving software. J Syst Softw 47(23):173181
22. Rothermel G, Untch RH, Chu C, Harrold MJ (1999) Test case prioritization: an empirical study. In: Proceedings of international conference on software maintenance, ICSM 1999. IEEE Computer Society Press, , p 179188, Aug 1999
23. Maheswari R, Mala D (2015) Combined genetic and simulated annealing approach for test case prioritization. Ind J Sci Technol
24. Lou Y, Hao D, Zhang L (2016) Mutation-based test-case prioritization in software evolution. In: IEEE 26th symposium on software reliability engineering, pp 4657
25. Yuan F, Bian Y, Li Z, Zhao R (2015) Epistatic genetic algorithm for test case prioritization. In: International symposium on search based software engineering. Springer, pp 109–124
26. Catal C (2012) On the application of genetic algorithms for test case prioritization: a systematic literature review. In: Proceedings of the 2nd international workshop
27. Kaur A, Goyal S (2011) A genetic algorithm for fault-based regression test case prioritization. Int J Comput Appl 32(8):975–8887
28. Jun W, Yan Z, Chen J (2011) Test case prioritization technique based on genetic algorithm. In: International conference on internet computing and information services, pp 173–175
29. Yadav DK, Dutta S (2017) Regression test case prioritization technique using genetic algorithm. In: Advances in computational intelligence. Springer, Singapore, pp 133–140
30. Deb K, Pratab A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm. IEEE Trans Evol Comput 6(2):182197
31. Li Z, Harman M, Hierons RM (2007) Search algorithms for regression test case prioritization. IEEE Trans Softw Eng 33(4):225237
32. Yadav DK, Dutta S (2016) Test case prioritization technique based on early fault detection using fuzzy logic. In: 3rd international conference on computing. for sustainable global development (INDIACom), pp 1033–1036
33. Solanki K, Singh Y, Dalal S, Srivastava PR (2016) Test case prioritization: an approach based on modified ant colony optimization. In: International conference on computer, communication and control. IEEE, pp 1–6
34. Gao D, Guo X, Zhao L (2015) Test case prioritization for regression testing based on ant colony optimization. In: Software engineering and service
35. Noguchi T, Washizaki H, Fukazawa Y (2015) History-based test case prioritization for black box testing using ant colony optimization. In: Proceedings of the 8th IEEE international conference on software testing, verification, and validation, testing in practice track, pp 13–17
36. Ledru Y, Petrenko A, Boroday S, Mandran N (2012) Prioritizing test cases with string distances. Autom Softw Eng 19(1):65–95

37. Jiang B, Chan WK (2015) Input-based adaptive randomized test case prioritization. J Syst Softw: 91–106
38. Yadav DK, Dutta S (2018) A new cluster-based test case prioritization using cat swarm optimization technique. In: Proceedings of the third international conference on microelectronics, computing and communication systems MCCS 2018
39. Srikanth H, Williams L, Osborne J (2005) System test case prioritization of new and regression test cases. In: International symposium on empirical software engineering: 6271
40. Srikanth H, Hettiarachchi C, Do H (2016) Requirements based test prioritization using risk factors. Inf Softw Technol 69(C): 7183
41. Yoon M (2012) A test case prioritization through correlation of requirement and risk. J Softw Eng Appl 5(10):823836
42. Yoon H, Choi B (2011) A test case prioritization based on degree of risk exposure and its empirical study. Int J Softw Eng Knowl Eng 21(2):191209
43. Hettiarachchi C, Do H, Choi B (2014) Effective regression testing using require- ments and risks. In: Proceedings 8th international conference on software security and reliability, SERE 2014, p 157166
44. Rothermel G, Untch R, Chu C, Harrold MJ (1999) Test case prioritization: an empirical study. In: Proceedings of international conference software maintenance, pp 179–188, Aug 1999
45. Elbaum S, Malishevsky A, Rothermel G (2000) Prioritizing test cases for regression testing. In: Proceedings of the international symposium on software testing and analysis, pp 102–112, Aug 2000
46. Yadav DK, Dutta S (2019) Test case prioritization based on early fault detection technique. Recent Patents Comput Sci 12(4). ISSN: 1874-4796 (Online), ISSN: 2213-2759 (Print). https://doi.org/10.2174/2213275912666190404152603
47. Yadav DK, Dutta S (2016) Test case prioritization technique based on early fault detection using fuzzy logic. In: 2016 3rd international conference on computing for sustainable global development (INDIACom). IEEE, pp 1033–1036
48. Kim JMKJM, Porter A (2002) A history-based test prioritization technique for regression testing in resource constrained environments. In: Proceedings of the 24th international conference on software engineering ICSE 2002, pp 119129
49. Huang YC, Peng KL, Huang CY (2012) A history-based cost-cognizant test case prioritization technique in regression testing. J Syst Softw 85(3):626637
50. Mirarab S, Tahvildari L (2007) A prioritization approach for software test cases based on Bayesian networks. In: International conference on fundamental approaches to software engineering. Springer, Berlin, Heidelberg, p 276290
51. Malishevsky AG, Ruthruff J, Rothermel G, Elbaum S (2006) Cost-cognizant test case prioritization. In: Technical Report TR-UNL-CSE-2006- 0004, University of Nebraska-Lincoln
52. Simons C, Paraiso EC (2010) Regression test cases prioritization using failure pursuit sampling. In: Proceedings of tenth international conference on intelligent systems design and applications, pp 923–928
53. Mahali P, Mohapatra DP (2018) Model based test case prioritization using UML behavioural diagrams and association rule mining. Int J Syst Assur Eng Manage: 1063–1079
54. Luo Q, Moran K, Zhang L, Poshyvanyk D (2018) How do static and dynamic test case prioritization techniques perform on modern software systems? An extensive study on GitHub projects. IEEE Trans Softw Eng
55. Panigrahi CR, Mall R (2012) A hybrid regression test selection technique for object-oriented programs. Int J Softw Eng Appl 6(4):1734
56. Panigrahi CR, Mall R (2010) Model-based regression test case prioritization. ACM SIGSOFT Software Eng Notes 35
57. Rothermel G, Untch RH, Chu CC, Harrold MJ (1999) Test case prioritization: an empirical study. In: Software maintenance, ICSM
58. Schwartz A, Do H (2016) Cost-effective regression testing through adaptive test pri- oritization strategies. J Syst Softw 115:6181

Effect of Ultrasonication and Centrifugation on the Pore Size of PVDF/CNT Nanocomposites

Sudha R. Karbari, Raghav Singhal, M. Uttara Kumari, and G. Shireesha

Abstract The paper conducts an initial study of the effect of ultrasonication and centrifugation along with the annealing temperature on the crystalline phase and porosity of PVDF/CNT films. The precursor solution was made using PVDF pellets and CNT with *N*, *N*-dimethylformamide acting as a solvent. The dispersion of CNT into the precursor solution of PVDF was aided by ultrasonication and centrifugation. The films were fabricated using a spin coating on a glass substrate. The films were annealed at a relatively high temperature and the resultant crystalline phases, and pore structure was studied using SEM and XRD analysis. The process described confirms the formation of β phase without the need for poling and stretching.

Keywords PVDF/CNT nanocomposites · Crystalline phases · Annealing · Centrifuge · Ultrasonicate probe

1 Introduction

The crystalline properties along with morphology of polyvinylidene fluoride for ultrafiltration membranes synthesized using the phase inversion method were investigated under ultrasound-assisted conditions in a bath. An increase in the intensity of ultrasonication increases the flux of membranes. The process of phase separation

S. R. Karbari · R. Singhal (✉) · M. Uttara Kumari
Department of Electronics and Communication Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India
e-mail: raghav97blr@gmail.com

S. R. Karbari
e-mail: karbari.rsudha@gmail.com

M. Uttara Kumari
e-mail: uttarakumari@rvce.edu.in

G. Shireesha
Department of Physics, Rashtreeya Vidyalaya College of Engineering, Bengaluru, India
e-mail: shireeshag@rvce.edu.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_43

was accelerated and macrovoids were developed by increasing the intensity resulting in an improvement in membrane porosity. The β crystalline phase in the ultrasound-assisted process using phase inversion method results in an elongation that is greater than the pristine membrane [1].

The effects of temperature and ultrasonication on the morphology and crystal structure of PVDF membranes were studied. As the temperature increased, the crystalline-induced gelatin line approached toward binodal, which is a dominant initial stage of precipitation. The top surface of the membrane observed using SEM and AFM was found to be composed of large stacks of crystalline lamella with the interlamellar distance of 13–20 nm [2].

Monir Noroozi et al. studied Nanofluids containing aluminum oxide nanoparticles of 11 nm dispersed in a distilled water at low temperature and various concentrations using two different ultrasonic devices, a probe and bath sonicator for dispersion. The overall process effect of ultrasonic system on the stability and the thermal diffusivity was investigated using the photopyroelectric technique. The observed thermal diffusivity of nanofluids resulting from sonication improved the stability and dispersion. Greater influence of sonication type, NP size and concentration on the thermal diffusivity enhancement of nanofluids was analyzed using PPE technique. The thermal diffusivity is completely dependent on the power of the sonication device. The thermal diffusivity improved with the smaller size of nanoparticles. There is a prominent improvement in the thermal diffusivity using probe sonicator compared to the bath sonicator at higher particle concentration with smaller particle size [3].

Graphene, as a reinforcement material, possess some problems while dispersion as it takes a long time that makes it difficult to produce applications at a larger scale. Mechanisms to improve the dispersion are developed and reported with various advantages and disadvantages compared and analyzed. Three methods of dispersion include physical, non-covalent bonding, and covalent bonding. The physical method damages the graphene structure which is irreversible and complicated to repair. Henceforth, these methods are used as additive methods and assisting in the synthesis process. Covalent bonding includes organic and inorganic bond formation. But this still affects the combining of nano-dopants with the matrix material. Non-covalent bonding classified as π–π bonding, ionic bonding, and hydrogen bond is simple to use for synthesis but need surfactants to disperse graphene [4].

Various studies investigate Solutions prepared by dissolving PVDF into DMF and a mixture of acetone and DMF. The casting of solutions in the film forms by using a spin coater and annealed. The parameters for processing include spin speed, baking temperature, solution viscosity, poling voltage, and the ratio of solutions is optimized to achieve the required characteristics. To improve the surface morphology for surface charge density of the films with mechanical characteristics, different solvents are used along with acetone. PVDF films prepared by spin coating require a spin speed of 2000 rpm and baked at 50–90 °C [5–8]. As the content of the solvent increases, the thickness of the film also increases. The acetone content increased the formation of α phase at and above 60 °C is decreased, whereas the rate of decrement in the β phase is decreased. It is aiding the increased evaporation rate of the solvent facilitating

the curing of molecular chains in stretching states resulting in the formation of piezoelectric β phase [9–11].

PVDF with mixed matrix membranes that are applicable for hydrophilic applications show improvement in the durability and properties with the use of hydrophilic dopants. This is facilitated by using two concepts of dope sonication and the introduction of PVP layer by layer. The sonication of the dopants is carried out for a longer time than compared to the older techniques, and this resulted in the pore formation and also functioned pore connector with particle dispersion. This henceforth improved the particle dispersion capability of nano-dopants into the matrix for surface wettability. The addition of PVP in a quantified manner resulted in the thermal stability of the nanocomposite and the mass fraction of the beta phase content [12–14].

Two different approaches of PVDF with CNT as nano-dopants are prepared by sonication of the solution and mechanical mixing in an appropriate ratio. The piezophase exists with the β-phase in the composite of CNT prepared by sonication, whereas no β-phase is observed in composite prepared by the mechanical mixture. With reference to the density functional theory, large amounts of energies required for transforming trans-gauche-trans-gauche0 (TGTG0) into a configuration of trans–trans and the molecular chain of TT with the absorption of CNT surface tight. The beta phase formation is independent of zigzag carbon atoms on the CNT surface. Because of sonication, dope diffusivity is improved and some TT chains are adsorbed on the CNT surface as nucleating agents for strong confirmation of crystalline morphological CNT nanocomposites with PVDF as the matrix [15]. Dope sonication as seen from the previous reports by the researchers resulted in efficient particle dispersion, with a very few improvements of the membrane surface wettability [14].

PVDF as a membrane is an important material for water treatment. The morphological changes are achieved under various conditions with different practical usage in distillation and ultrapure cleaning of water and protein attachment. The numerous effects of conditions of the dissolution of solvents on the performance of PVDF are observed. Four different PVDF ultrafiltration membranes are synthesized from dopants by dissolving and stirring at 24, 90, 120 °C and sonicated. The dope sonication resulted in improved mechanical and thermal stability, with high flux recovery of 26% at 24 °C. This linked to the slightly lower porosity, narrow distribution of small pore sizes and smooth surface morphology [17].

2 Materials and Methods

2.1 Materials Processing

PVDF pellets is a product of sigma Aldrich, the multiwalled CNTs having a diameter from 10 to 50 nm and length varying from 10 to 50 nm were bought. PVDF dissolved

in DMF using magnetic stirring. Figure 1 shows the solution prepared using the standard process.

One of the parameters for magnetic stirring is the temperature for the set up that leads to the conversion of pellets to a viscous liquid. Samples were all synthesized using 1% by weight DMF solution of PVDF copolymer.

Spin coating is a process by which thin uniform films of a polymer are fabricated. A prepared and clean substrate is placed on the chunk of a spin coater with the parameters set as shown in Fig. 2. The substrate is cleaned using piranha cleaning standard process. Spin coater parameter for the operation depends on viscosity, drying rate, surface tension, and solvent used. Samples were made on glass slides. The parameters for spin coating include spin speed, acceleration, and the duration with

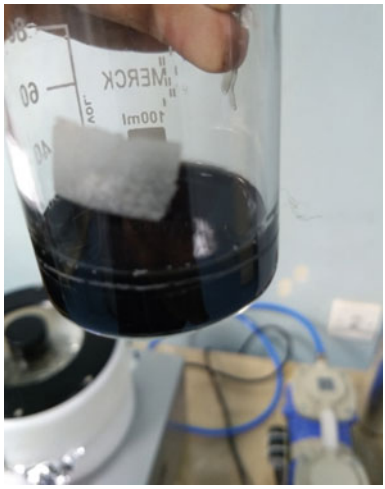


Fig. 1 A solution of PVDF/CNT nanocomposites after magnetic stirring

Fig. 2 Parameters setup for spin coater (fabrication of nanocomposites)

values of 200 rpm, 25 rpm/s, and 80 s, respectively, obtaining a uniform thickness of the films of 15um.

2.2 *Process of Ultrasonication and Centrifuge*

Procedure 25 ml of DMF was taken in a beaker 0.025 g of MWCNT was measured out and placed on a watch glass. The nano-dopant (CNT) added to the DMF and stirred for 4 continuous hours on a magnetic stirrer. The beaker containing the sample was then placed in an ultrasound water bath for an hour. The spin-coated sample prepared from this method leads to clump formation and uneven distribution across the matrix. In the second approach, a probe-type ultrasonicator was used and the CNT was dried (as these were the suspected causes of uneven mixing). The CNT was then placed in a vacuum drier and heated to remove any present moisture. The precaution was taken to wear a mask so as to not accidentally inhale any CNT while handling the sample. The mixed sample was kept under a probe-type ultrasonicator as shown in Fig. 3. Care was taken to keep the beaker in cold water/ice bath as the process generates a lot of heat and can lead to the evaporation of DMF. This improved the results as compared to the first approach as the supernatant fluid was visibly darker indicating a more even and dense dispersion.

There were still a small number of clumps of CNT observed. To separate them from the supernatant, decantation was ineffective, so the sample was separated in a

Fig. 3 Ultrasonicator Probe station for sample preparation

centrifuge. The sample was placed in 2 ml centrifuge tubes and spun at 5000 rpm for 10 min. The supernatant was then collected from the tubes, and the obtained dispersion was added to a recently mixed solution of PVDF pellets in DMF. The resultant mixture was spun using a magnetic stirrer. This sample was then used to create films using a spin coater. Films from the second procedure were of a darker appearance indicating a good dispersion of CNT. While drying, it was noticed that the films of the second sample started separating from the glass substrate before drying and minor shrinkage occurred. This was not observed in pristine PVDF samples. Double-coated samples and thick samples (slow spin rate) were observed to have higher opacity and porosity as compared to pristine PVDF samples which are prescribed.

The samples are annealed at a higher temperature to increase the porosity and the transition of nonpolar to polar phases. Figure 4 shows the opacity and transparency of the films prepared after spin coating and left for annealing.

Fig. 4 Samples of spin-coated PVDF/CNT nanocomposites

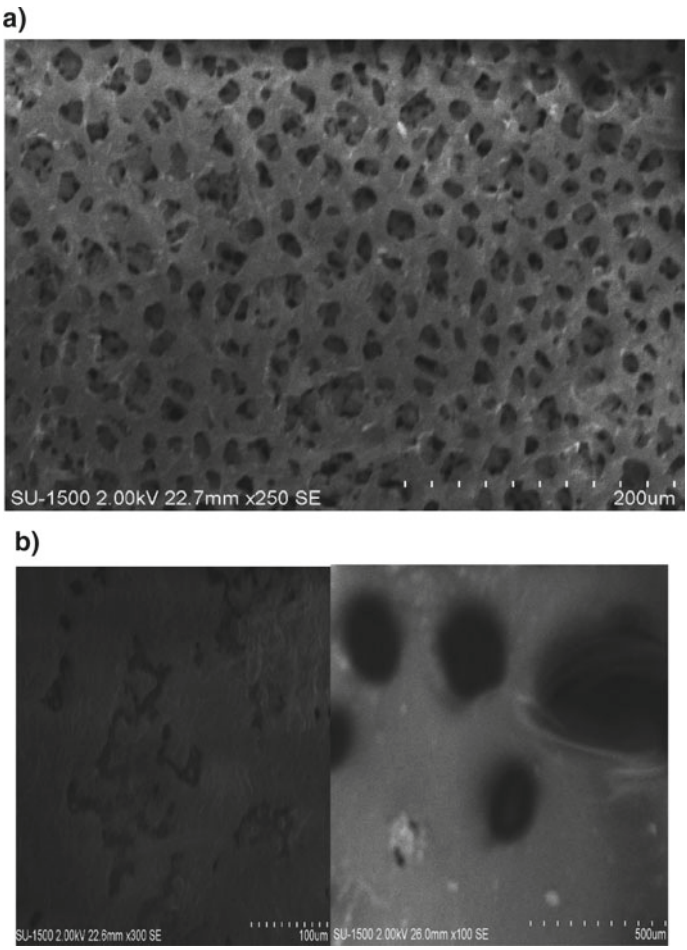


Fig. 5 **a** SEM images showing pore structure (at 200 μm magnification) **b** SEM images showing pore structure (at 100 μm and 500 μm magnification)

3 Results and Discussion

3.1 SEM Analysis

The SEM images show the existence of different components such as nanoparticles, amorphous and crystalline phases in a nanocomposite PVDF/CNT. SEM micrographs of pure PVDF dispersed with CNT nanoparticles are shown in Fig. 5. The image shows all the cases of sizes of pores for concentrations.

The size of pores increased due to the process followed. There is uniformity in the pores of the matrix with CNT as nano-dopants henceforth confirming and dispersion.

It is also found that fine spherulites were arranged in a well-defined manner in all cases. One of the most important tests that help to verify the particle dispersion. Small amounts of CNT are dispersed, and the particle agglomeration can be seen from the above image. Surface analysis of deposited nanoparticle thin films by spin coating is investigated by scanning electron microscopy. This indicates that the long-term sonication with centrifuge proved useful in mechanical unbundling and uniform distribution of pores at the nanolevel.

3.2 Crystal Phase Characterization

The samples were scanned in the 2θ range of 10–45 with a step of 0.0167. The crystalline phase with the porosity increase due to ultrasonication using a probe method prepared by solution casting without experiencing any mechanical stretching related to the crystalline rate of the solution as shown in Fig. 6 [18–20].

The X-ray diffraction (XRD) measurement is performed to investigate the crystalline nature of the samples of PVDF/CNT nanocomposites. The XRD pattern of the PVDF CNT nanocomposite obtained by the procedure as mentioned above shows the following characteristics at various peak points. The peaks of 2θ that correspond to 19.30 (110), 22.40 (100), 38.30 (111), 45.50 (200) are indicating the various phases of the PVDF polymers processed [21–23]. The intensity of (200) peak is significantly larger compared to the other peaks. The peaks at $2\theta = 19.30$ (110) and 22.40 (100) are related to the crystalline peaks of β and γ-phase of crystalline PVDF. The peaks

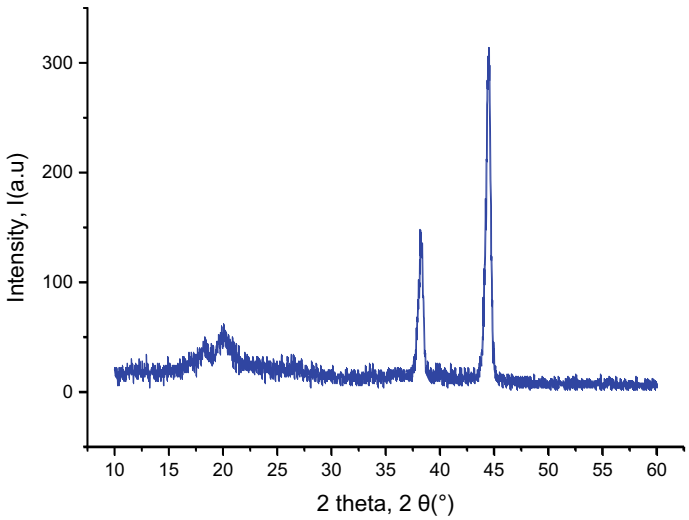


Fig. 6 XRD plots of PVDF/CNT nanocomposites

at 38.80 and 45.50 corresponding to (200) and (111) of the monoclinic α-phase, respectively.

4 Conclusion

In the study, the process parameter effect of ultrasonication and centrifugation on the crystalline phase of PVDF is observed. XRD analysis confirms the formation of β phase for PVDF /CNT nanocomposite along with α and γ phases. The process of annealing and effective incorporation of CNT dopant changed the phase of the material from α to β. The conventional method of CNT addition leads to the formation of clumps, indicating agglomeration. In an effort to overcome this, experiments were carried out leading to the development of a novel process which includes the usage of ultrasonication probe station with a centrifuge, and the annealing temperature variation. This results in an increase in the pore size with uniformity throughout the matrix of PVDF nanocomposite film.

Acknowledgements The experimentation was carried out at Interdisciplinary Research Center (IDRC) RV College of Engineering.

References

1. Qu T, Pan K, Li Li, Liang B, Wang L, Cao B (2014) Influence of ultrasonication conditions on the structure and performance of poly(vinylidene fluoride) membranes prepared by the phase inversion method. Ind Eng Chem Res 53:8228–8234. https://doi.org/10.1021/ie5012905
2. Cheng L-P (1999) Effect of temperature on the formation of microporous PVDF membranes by precipitation from 1-Octanol/DMF/PVDF and water/DMF/PVDF systems. Macromolecules 32:6668–6674. https://doi.org/10.1021/ma990418l
3. Noroozi M, Radiman S, Zakaria A (2014) Influence of sonication on the stability and thermal properties of Al_2O_3 nanofluids. J Nanomater https://doi.org/10.1155/2014/612417
4. Liang A, Jiang X, Hong X, Jiang Y, Shao Z, Zhu D (2018) Recent developments concerning the dispersion methods and mechanisms of graphene. Coatings 8:33. https://doi.org/10.3390/coatings8010033
5. Zeng C, Hossieny N, Zhang C, Wang B (2010) Synthesis and processing of PMMA carbon nanotube nanocomposite foams. Polymer 51:655–664. https://doi.org/10.1016/j.polymer.2009.12.032
6. Rezvantalab H, Ghazi N, Ambrusch M, Infante J, Shojaei-Zadeh S (2017) An aqueous-based approach for fabrication of PVDF/MWCNT porous composites. Sci Rep 7. https://doi.org/10.1038/s41598-017-01770-9
7. Dhand V, Hong S, Li L, Kim JM, Kim S, Rhee KY, Lee H (2019) Fabrication of robust, ultrathin and light weight, hydrophilic, PVDF-CNT membrane composite for salt rejection. Compos B Eng 160:632–643. https://doi.org/10.1016/j.compositesb.2018.12.106
8. Mishra S, Kumaran KT, Sivakumaran R, Pandian S, Kundu S (2016) Synthesis of PVDF/CNT and their functionalized composites for studying their electrical properties to analyze their applicability in actuation and sensing. Coll Surf A Physicochem Eng Aspects 509. https://doi.org/10.1016/j.colsurfa.2016.09.007

9. Martins P, Lopes A, Lanceros-Méndez S (2014) Electroactive phases of poly(vinylidene fluoride): determination, processing and applications. Prog Polym Sci 39:683–706. https://doi.org/10.1016/j.progpolymsci.2013.07.006
10. Cardoso V, Minas G, Costa C, Tavares C, Lanceros-Méndez S (2011) Micro and nanofilms of poly(vinylidene fluoride) with controlled thickness, morphology and electroactive crystalline phase for sensor and actuator applications. Smart Mater Struct 20:087002. https://doi.org/10.1088/0964-1726/20/8/087002
11. Jee T, Lee H, Mika B, Liang H (2007) Effects of microstructures of PVDF on surface adhesive forces. Tribol Lett 26:125–130. https://doi.org/10.1007/s11249-006-9163-z
12. Gregorio Jr R (2006) Determination of the α, β, and γ crystalline phases of poly(vinylidene fluoride) films prepared at different conditions. J Appl Polym Sci 100:3272–3279. https://doi.org/10.1002/app.23137
13. Gregorio Jr R, Borges D (2008) Effect of crystallization rate on the formation of the polymorphs of solution cast poly(vinylidene fluoride). Polymer 49:4009–4016. https://doi.org/10.1016/j.polymer.2008.07.010
14. Ike I, Ludovic D, Groth A, Orbell J, Duke M (2017) Effects of dope sonication and hydrophilic polymer addition on the properties of low pressure PVDF mixed matrix membranes. J Membrane Sci 540. https://doi.org/10.1016/j.memsci.2017.06.056
15. Yu S, Zheng W, Yu W, Zhang Y, Jiang Q, Zhao Z (2009) Formation Mechanism of beta-Phase in PVDF/CNT composite prepared by the sonication method. Macromolecules 42. https://doi.org/10.1021/ma901765j
16. Anjalin M (2014) Synthesis and characterization of MWCNTs/PVDF nanocomposite and its electrical studies. Der Pharma Chemica 6:354–359
17. Ike I, Zhang J, Groth A, Orbell J, Duke M (2017) Effects of dissolution conditions on the properties of PVDF ultrafiltration membranes. Ultrasonics Sonochem 39. https://doi.org/10.1016/j.ultsonch.2017.05.041
18. Khan W, Sharma R, Saini P (2016) Carbon nanotube-based polymer composites: synthesis, properties and applications. https://doi.org/10.5772/62497
19. Hudaib B, Gomes V, Shi J, Liu Z (2017) Poly (vinylidene fluoride)/ Polyaniline/MWCNT nanocomposite ultrafiltration membrane for natural organic matter removal. Sep Purif Technol 190. https://doi.org/10.1016/j.seppur.2017.08.026
20. Xia W, Zhang Z (2018) PVDF-based dielectric polymers and their applications in electronic materials. IET Nanodielectrics 1. https://doi.org/10.1049/iet-nde.2018.0001
21. Tsonos C, Pandis C, Soin N, Sakellari D, Myrovali E, Kripotou S, Kanapitsas A, Siores E (2015) Multifunctional nanocomposites of poly(vinylidene fluoride) reinforced by carbon nanotubes and magnetite nanoparticles. eXPRESS Polymer Lett 9: 1104–1118. https://doi.org/10.3144/expresspolymlett.2015.99
22. Mago G, Kalyon D, Fisher F (2008) Membranes of polyvinylidene fluoride and PVDF Nanocomposites with carbon nanotubes via immersion precipitation. J Nanomater https://doi.org/10.1155/2008/759825
23. Aqeel S, Huang Z, Walton J, Baker C, Falkner D, Liu Z, Wang Z (2017) Polyvinylidene fluoride (PVDF)/polyacrylonitrile (PAN)/carbon nanotube nanocomposites for energy storage and conversion. Adv Compos Hybrid Mater 1. https://doi.org/10.1007/s42114-017-0002-5

Controlling of Twin Rotor MIMO System (TRMS) based on Multivariable Model Predictive Control

Ujjwal Manikya Nath, Chanchal Dey, and Rajani K. Mudi

Abstract Objective of this study is to design a model predictive control (MPC) for twin rotor multi-input multi-output (MIMO) system (TRMS). Major challenges for controlling of TRMS are that it involves bi-directional motions in relation with yaw and pitch movements. Moreover, there is a strong interaction in between them resulting nonlinear behaviour. Therefore, conventional proportional integral derivative (PID) controller fails to provide satisfactory performance for such nonlinear multi-input multi-output (MIMO) system. However, multivariable MPC may be considered to be a useful control methodology in such cases. Efficacy of the designed multivariable MPC is estimated by computing integral absolute error (IAE) in control action during closed-loop operation realized with the help of MATLAB environment.

Keywords Model predictive control · Multivariable MPC · Twin rotor MIMO system · TRMS process

1 Introduction

For controlling linear systems, substantial amount of developments have already been reported in literature to ascertain online optimization, stability, and closed-loop

U. M. Nath (✉)
Department of Instrumentation Engineering, Jorhat Engineering College, Garmur, Jorhat 785007, Assam, India
e-mail: um.nath@yahoo.com

C. Dey
Instrumentation Engineering Section, Department of Applied Physics, University of Calcutta, Kolkata 700009, West Bengal, India
e-mail: cdaphy@caluniv.ac.in

R. K. Mudi
Department of Instrumentation and Electronics Engineering, Jadavpur University, Salt Lake Campus, Kolkata 700098, West Bengal, India
e-mail: rajanikanta.mudi@jadavpuruniversity.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_44

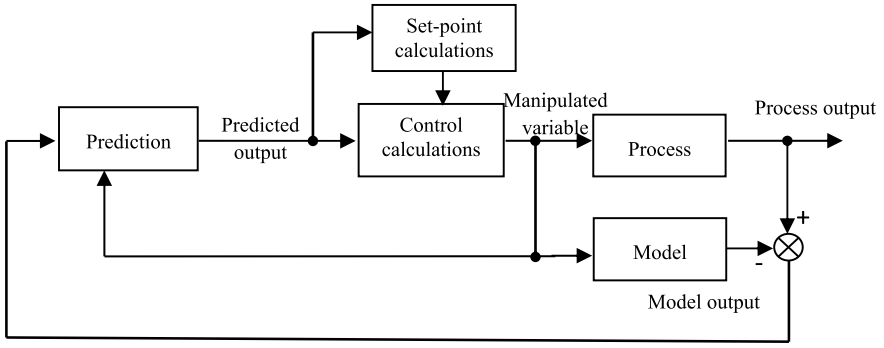


Fig. 1 Block diagram of model predictive control technique

performance enhancement [1]. But, in comparison with linear systems, relatively less satisfactory progress has been observed towards controlling nonlinear and/or multivariable processes. For industrial processes, many practical closed-loop control applications are still managed by manual control based on prior acquired knowledge. Till date, hardly, any mathematical relation has been found for online computation towards deciding appropriate control methodology for such systems [2].

In general, controlling task becomes substantially difficult for processes with more than one interacting loop, i.e. for MIMO systems. Stronger the interaction between different channels, selection of suitable control policy for such cases becomes more difficult. Extensive interaction among the input and output variables [3] leads to nonlinear process characteristics. Hence, conventional PID controller is not capable enough to provide satisfactory control performance for such MIMO processes. Twin rotor MIMO system (TRMS) [4–7] is well-known example for such multivariable process with significant nonlinearity due to considerable interaction between yaw and pitch motions.

Model predictive control (MPC) [8, 9] appeared in industry [10, 11] as an effective solution to deal with multivariable constrained control problems. MPC contains some unique features [12–16] over the conventional PID controllers – employs an explicit model of the concerned process, utilizes moving horizon control policy and exploits online optimization technique to calculate future control moves. Here, in the reported work, a multivariable MPC is designed for TRMS process and its quantitative assessment is made through computation of performance index—IAE (integral absolute error) and total variation (TV) in control action during closed-loop set point tracking and load recovery responses.

2 Model Predictive Control

Basic block diagram of MPC [8]-based control technique is shown in Fig. 1. Designing of MPC can be formulated by the following steps:

Step I	Control horizon M and prediction horizon P are chosen such that $M < P$.
Step II	Present time instant is considered as $t = k$.
Step III	Output for all time samples $y(k+1) \ldots y(k+P)$ are searched to select the closest output to the set point R (optimization problem).
Step IV	Corresponding objective function is framed

$$\mathrm{O} = \sum_{i=1}^{P} [y(k+i) - R]^2 + \sum_{i=1}^{M} [u(k+i-1)]^2$$

where $u(k)$ is manipulated variable.

Step V	Minimization of the objective function (Step IV) is done by manipulating $u(k) \ldots u(k+M-1)$with$u(k+m) = u(k+m-1)$ where m is $M, M+1 \ldots P$.
Step VI	Optimized control moves $u^*(k) \ldots u^*(k+M-1)$ at time k are calculated.
Step VII	Implementation is made related to control move $u^*(k)$.
Step VIII	For time $k+1$, same steps are repeated from Step III.

Figure 2 shows the timing diagram of MPC, where future control moves (M) are calculated based on past and present control position. In case of MIMO process, outputs of the process are considered to be y_1 and y_2 and they can be represented by the following relations

$$y_1(k) = \sum_{i=1}^{y_{11}} u_1(k+j-i) + \sum_{i=1}^{y_{12}} u_2(k+j-i),$$

$$y_2(k) = \sum_{i=1}^{y_{21}} u_1(k+j-i) + \sum_{i=1}^{y_{22}} u_2(k+j-i).$$

Figure 3 shows the block diagram of multivariable MPC where the set points are R_1 and R_2, corresponding control actions are u_1, u_2 and outputs are y_1, y_2 respectively.

3 Twin Rotor MIMO System

Typical schematic of TRMS [4–7] is shown in Fig. 4 where the yaw and pitch motors are driven by two DC motors connected on a common shaft. TRMS structure resembles the dynamics of a helicopter with two propellers which are perpendicular

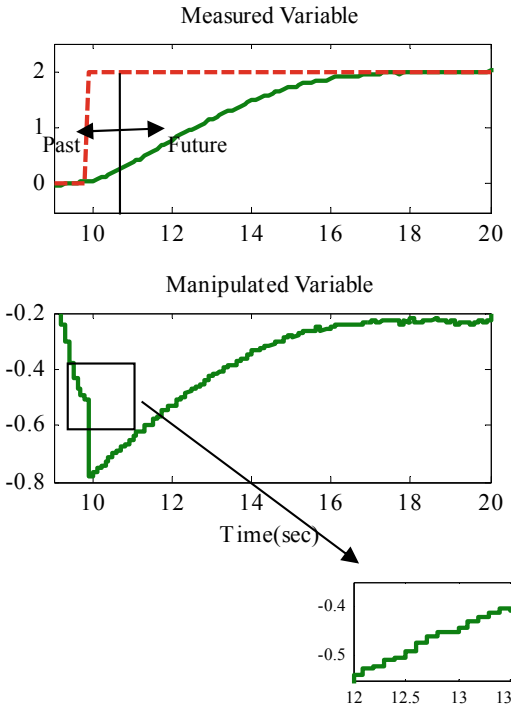


Fig. 2 Timing diagram of MPC

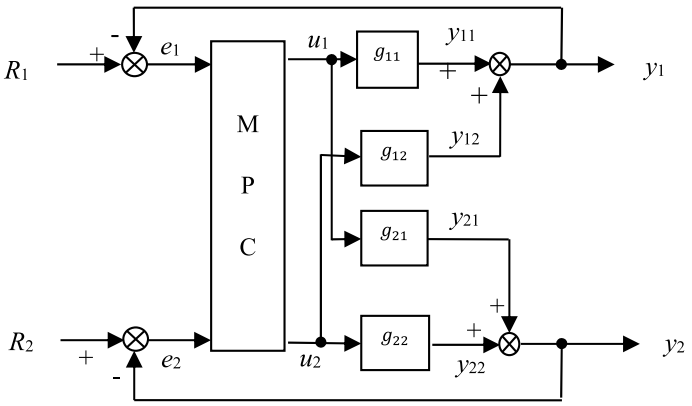


Fig. 3 Multivariable MPC for controlling of TITO process

to each other. Connection between the two rotors is built by a beam pivoted on its base that can move freely in both horizontal and vertical planes. Moreover, there is a counterweight mounted by another beam fixed perpendicular to the main one. The beam can move in vertical as well as in horizontal planes by changing the supply voltages of the motors to affect the rotational speed and directions of the rotors . Due

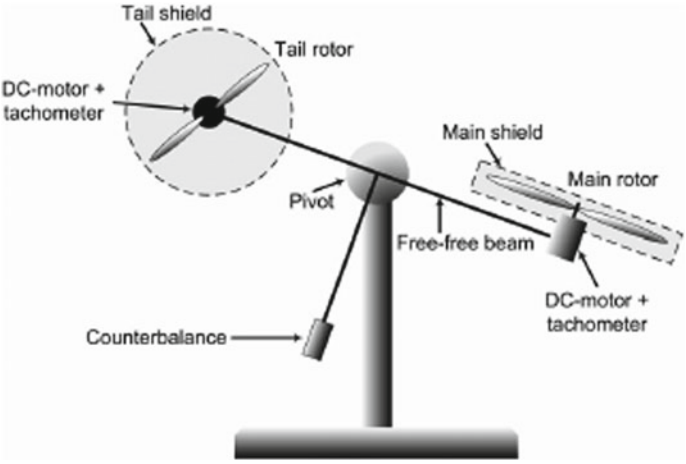


Fig. 4 Typical diagram of TRMS

to simultaneous motion of the yaw and pitch rotors, there is a strong cross-coupling between the main rotor (yaw) and the tail rotor (pitch).

4 Simulation Study

For simulation study TRMS process model is chosen from literature [4] as given by the following state space model:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -4.738 & -0.032 & 5.660 & 0 & 0 & 0 \\ 0 & 0 & -0.752 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -0.123 & -0.186 & 1 \\ 0 & 0 & 0 & 0 & 0 & -2.326 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 1 \\ 0 & 0.239 \\ 0.752 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 2.326 \end{bmatrix},$$

$$C = \begin{bmatrix} 1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0\,0 \end{bmatrix}.$$

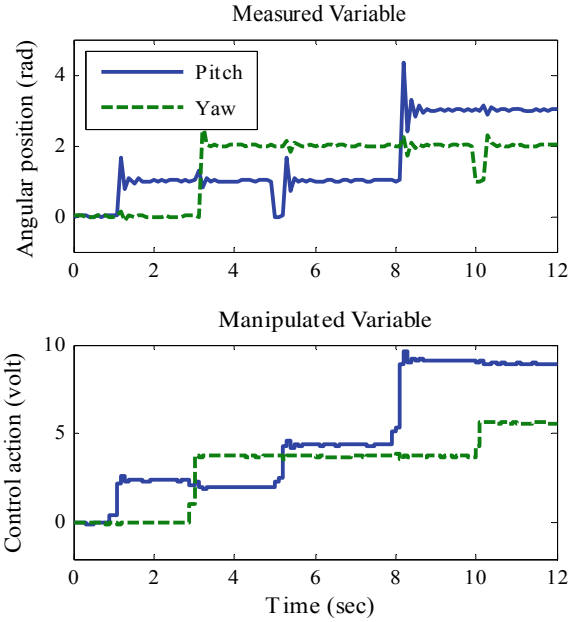


Fig. 5 Yaw and pitch response of TRMS

Table 1 Performance indices

	IAE	TV
Pitch	18.059	5.64
Yaw	17.283	3.69

Here, A is the system matrix, B is the input matrix and C is the output matrix of the chosen TRMS model. During closed-loop simulation, two-step input changes are provided to pitch at 1 s and 8 s, respectively, and load disturbance is introduced at 5 s. In addition, step input change and load disturbance are provided to yaw at 3 s and 10 s, respectively. Corresponding yaw and pitch response of the TRMS are shown in Fig. 5. For quantitative estimation of closed-loop responses performance indices, integral absolute error (IAE) and total variation (TV) in control action are calculated for both yaw and pitch motions as listed in Table 1.

5 Conclusion

In this work, performance evolution of multivariable MPC is made on TRMS. The main constraint of such multivariable process is that it has more than one input and output variables. In future scope, the objective function of MPC may be enhanced by incorporating change in control action (Δu) and in such case, the resulting objective

function will be $O = \sum_{i=1}^{P} [y(k+i) - R]^2 + \sum_{i=1}^{M} [u(k+i-1)]^2 + \Delta u$ and performance comparison will be made with existing MPC algorithm.

References

1. Bequette BW (2004) Process control modeling, design, and simulation, 2ed edn. Prentice Hall, India
2. Diehl M, Bock HG, Schloder JP, Findeisen R, Nagy Z, Allgower F (2002) Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. J Process Control 12: 577–585
3. Seborg DE, Edgar TF, Mellichamp DA (2004) Process dynamics and control. Wiley, Singapore
4. Gabriel C Modelling, simulation and control of a twin rotor MIMO-system. In: Project thesis. Department of Systems Engineering and Control, University Politecnica de Valencia
5. Rashad R, Badaway AE, Aboudonia A (2017) Sliding mode disturbance observer-based control of a twin rotor MIMO system. ISA Trans 69:166–174
6. Tastemirov A, Lecchini-Visintini A, Morales-Viviescas RM (2017) Complete dynamic model of the twin rotor MIMO system (TRMS) with experimental validation Control engineering practice 66: 89–98
7. Azar AT, Sayed AS, Shahin AS, Elkholy HA, AmmarHH (2019) PID controller for 2-DOFs Twin Rotor MIMO system tuned with particle swarm optimization. In: Proceedings of the international conference on advanced intelligent systems and informatics, pp 229–242
8. Wang L (2009) Model predictive control system design and implementatioin using MATLAB. Springer-Verlag London Limited, Spain
9. Grune L, Pannek J (2016) Nonlinear model predictive control. Springer
10. Vazquez S, Rodriguez J, Rivera M, Franquelo LG, Norambuena M (2017) Model predictive control for power converters and drives: advances and trends. IEEE Trans Ind Electron 64(2):935–947
11. Li H, Yan W, Shi Y (2017) Continuous-time model predictive control of under-actuated spacecraft with bounded control torques. Automatica 75:144–153
12. Morari M, Lee JH (1999) Model predictive contol: past, past and future. Comput Chem Eng 23:667–682
13. Scattolini R (2009) Architectures for distributed and hierarchical model predictive control—a review. J Process Control 19:723–731
14. Darby ML, Nikolaou M (2012) MPC: Current practice and challenges. Control Eng Pract 20: 328–342
15. Mayne DQ (2014) Model predictive control: recent developments and future promise. Automatica 50:2967–2986
16. Mesbah A (2016) Stochastic model predictice control: an overview and perspective for future research. IEEE Control Syst Mag 36(6):30–44

Implementation of a Multi-level Inverter (MLI) with a New Structure with Different Values of DC Sources

Bidyut Mahato, Saikat Majumdar, Sambit Vatsyayan, and K. C. Jana

Abstract Recent advances in the domain of medium-voltage, high-power, and multilevel inverter have risen to such a level due to their amazing facts such as acceptable harmonic spectrum, ease in control, reduced EMI, filter-free circuit, reduced stress on switches, and reduced common-mode voltage. This paper describes a new single-phase structure of multilevel inverter, generalized in nature employing a smaller number of overall power components, especially the power switches. 9-level structure is thoroughly explained and discussed with necessary details. In addition, 13-level inverter is also designed and tested with the RL-Load. Extensive simulations are carried out in MATLAB and the corresponding experimental results verifies the performance of the proposed MLI. Nonetheless, the MLI topology is compared with some of the recent reported literatures and in every way considered better.

Keywords Power components · DC/AC power conversion · 7-level inverter · 13-level inverter · SPWM

1 Introduction

Abundant inadequacies such as larger voltage stress on power switches and the higher harmonics in the output voltage are observed in the classical inverters (two-level) [1]. Nonetheless, the introduction of multilevel inverter (MLI) provides numerous advantages, some of which are lower harmonic content [2, 3], reduced voltage stress as well as reduced switching losses. In addition, some of the important features such as higher efficiency, electromagnetic compatibility, high voltage, and high power applications [4] make MLI suitable for the different industries. The first topology to obtain a higher voltage levels was, the "neutral point clamped", [5]. The other

B. Mahato (✉) · S. Majumdar · S. Vatsyayan · K. C. Jana
Department of Electrical Engineering, IIT (ISM), Dhanbad, India
e-mail: bidyut1990@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_45

renowned topology called "Flying Capacitor" structure being implemented to overcome the problems caused by the increasing amount of clamping diodes (in NPC-MLI). Nonetheless, the main weakness lies with the FC-MLI is the voltage balancing techniques for the large amount of available capacitors [6]. Nevertheless, the shortcomings of the two classical MLIs mentioned above are overshadowed by the advent of H-bridge configuration due to the fact that it does not need any surplus components in the circuit such as diodes and capacitors. The design of the H-bridge [7] can be cascaded with the same or different magnitudes of DC sources and generally referred to as symmetric CHB-MLI and asymmetric CHB-MLI, respectively, with using same number of power electronic switches. It, therefore, involves the development of such MLIs that can operate with the significant amount of component reduction.

The researchers working in the field of MLI, targets to introduce a new design and establish various different reduced MLI switch topologies which should have fewer power switches, input DC voltages, gate drivers, and losses. In this work, a 9-level inverter as well as 13-level inverter are built, and the experimental results verifies the simulation outcomes. MLI topologies recently proposed in [8–13] are being investigated and compared with the topology suggested in this paper.

This paper is distributed in various sections where Section-II illustrates the operating principle of the suggested MLI topology, Section-III deals with the control algorithm, Section-IV shows the comparison with the newly proposed MLI topologies, Section-V includes extensive simulation and experimental results, and Section-VI concludes the work.

2 Proposed Topology

Unlike other MLI structures, the proposed topology only uses two power switches, S_x and S_x' instead of H-bridge and the generalized structure of the proposed MLI is shown in Fig. 1a. The magnitude of DC source, V_x is the summation of the rest of the DC sources in the MLI topology. Each unit comprises of two power switches along with a DC source. Arrangement of the MLI topology proposed with two units is depicted in Fig. 1b which can be easily extended to nth unit. The employed power devices (switches) in each of the units are two in number. First unit consists of switches, S_{11} and S_{12}, second unit consists of switches, S_{21} and S_{22}, and similarly, the nth unit comprises of switches, S_{n1}, S_{n2}, respectively. The voltage steps specification is determined by the magnitude of the DC sources. The MLI topology can be performed for both the symmetric and asymmetric values of DC sources. To avail the maximum number of voltage levels, values of DC sources are to be taken as 1:2:2 for three DC sources (V_{cd}, V_1 and V_{ab}) or one unit i.e. $V_{cd} = V_{dc}$, $V_1 = 2V_{dc}$, $V_{ab} = 2V_{dc}$, $V_x = 5V_{dc}$, 1:2:2:2 for four DC sources (V_{cd}, V_1, V_2 and V_{ab}) or two units i.e. $V_{cd} = V_{dc}$, $V_1 = 2V_{dc}$, $V_2 = 2V_{dc}$, $V_{ab} = 2V_{dc}$, $V_x = 7V_{dc}$, 1:2:2:2:2 for five DC sources (V_{cd}, V_1, V_2, V_3 and V_{ab}) or three units $V_{cd} = V_{dc}$, $V_1 = 2V_{dc}$, $V_2 = 2V_{dc}$, $V_2 = 2V_{dc}$, $V_{ab} = 2V_{dc}$, $V_x = 9V_{dc}$ and so on. Uni-directional switches are all the only control switches used in this topology. Four power switches are fixed in this topology, S_a,

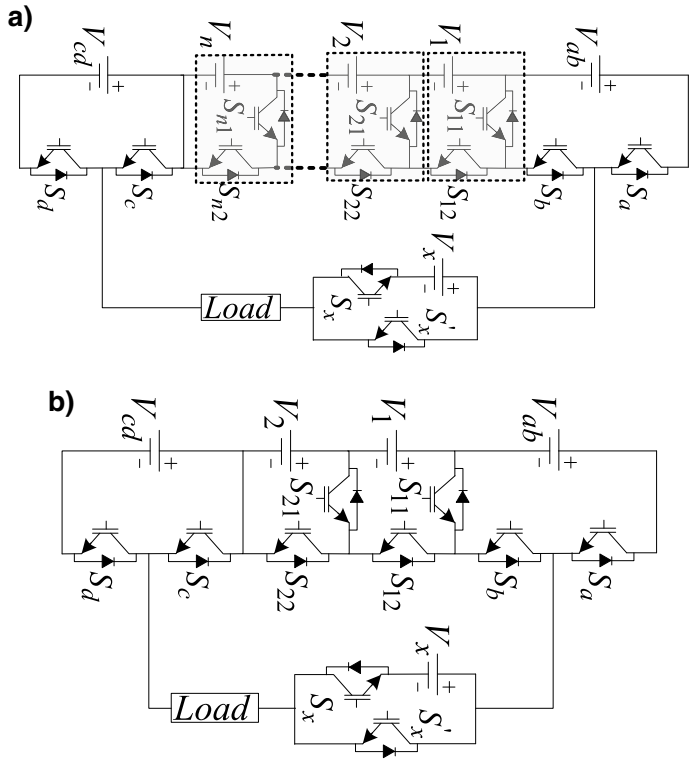


Fig. 1 **a** Proposed topology structure (generalized), **b** proposed topology with two units

S_b, S_c, S_d and two DC sources V_{ab}, V_{cd} which does not depend on the increasing number of units.

The proposed topology has a basic structure that produces 9-level voltage output for the equal values of input DC supplies as well as 13-level voltage output for all different DC magnitudes values.

3 Modes of Operation

The controller method suggested in this paper for the MLI is the technique of pulse width modulation (PWM). In this simple and well-defined technique, triangular carriers are chosen and a reference sine signal is compared to generate the corresponding pulse patterns.Level-shifted PWM (LS-PWM) is chosen among some of the well-known control strategies, i.e. selective harmonic elimination (SHE) [14], nearest-level control (NLC) [15, 16], phase-shifted PWM [17], sine-based carrier PWM [18, 19], to cultivate the pulses in MATLAB because it is simple and easy to

control. The control switches are provided with the switching frequency to be 3 kHz. DSPACE-1103 controller is used in the prototype model to generate experimental pulses. For this topology, eight triangular carriers are kept upon each other in the fashion of different upper and lower levels. The switching patterns obtained from the simulation and the controller are shown in Figs. 2a–c and 3a–c for one reference signal loop, respectively. Figure 4 illustrates the various modes that generate the different steps of voltage have been well presented, the dark path shows the current path as well as the corresponding voltage generation of the proposed 9-level inverter (symmetrical) topology. The voltage steps are well discussed in the theoretical and pictorial explanations, and the switching pattern of the ON switches are explained too in Table 1.

The generation of the various voltage steps for the 7-level inverter in different modes is discussed beneath.

Mode-1 ($+4V_{dc}$): Together with the only one DC source $V_x = +114$ V, power switches S_b, S_c, S_{12}, S_{22} and S_x are turned ON to generate the peak voltage of magnitude +112 V.

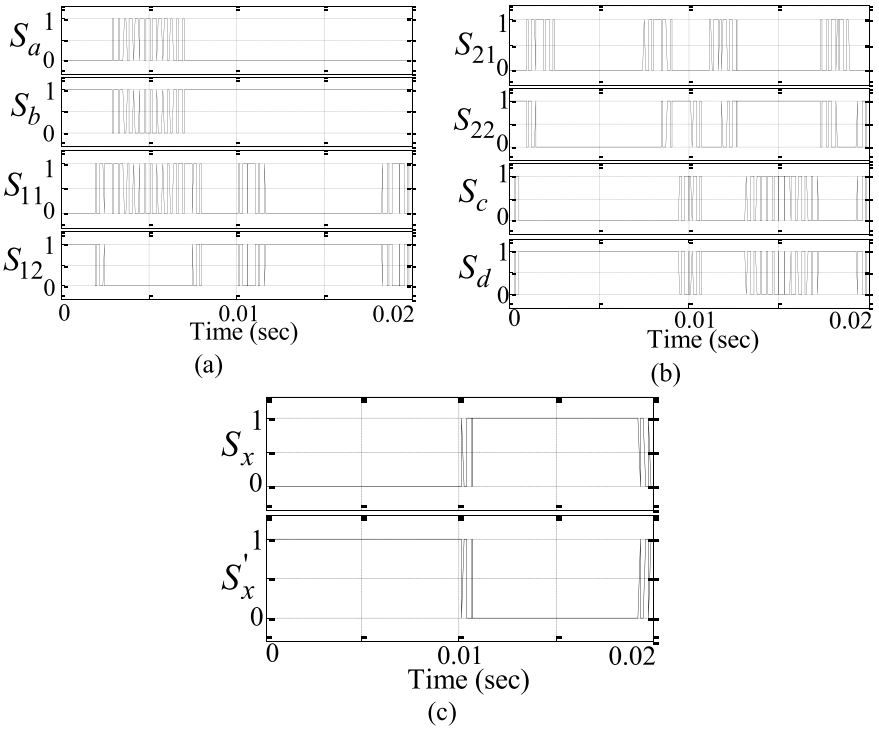


Fig. 2 Simulation results of gate pulses of the power switches. **a** S_a, S_b, S_{11}, S_{12}, **b** S_{21}, S_{22} S_c, S_d, **c** S_x and S'_x

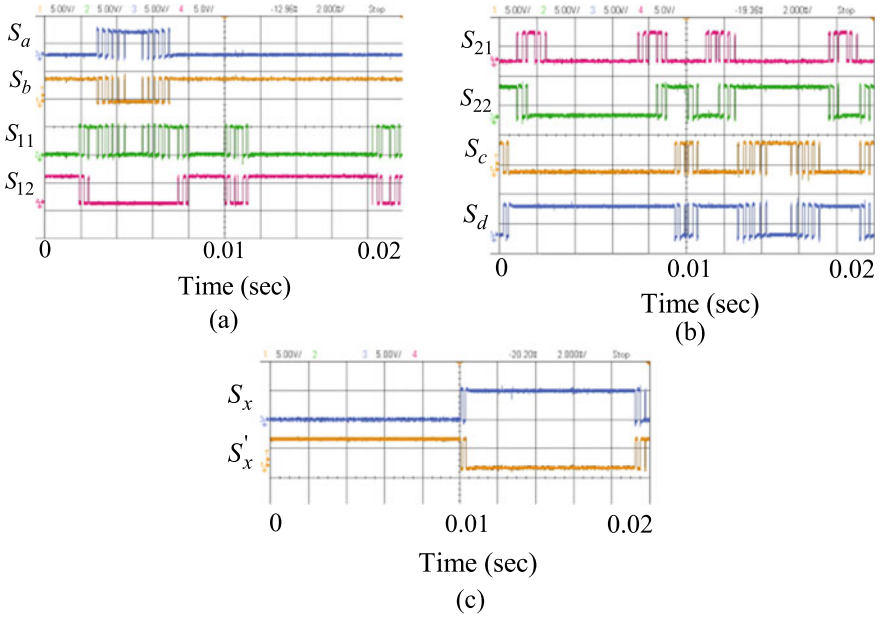


Fig. 3 Experimental results of gate pulses of the power switches. **a** S_a, S_b, S_{11} and S_{12}, **b** S_{21}, S_{22} S_c and S_d, **c** S_x and $S_x^{'}$

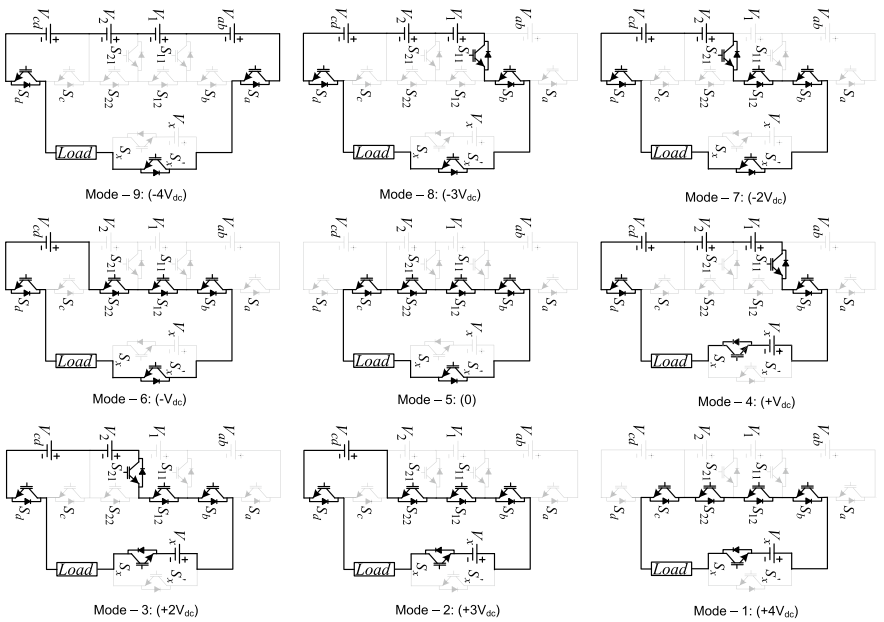


Fig. 4 Modes of operation

Table 1 Switching table for proposed 9-Level MLI ($V_1 = V_2 = V_{ab} = V_{cd} = V_{dc}$) symmetrical.

On = 1; Off = 0										Voltage levels
S_a	S_b	S_c	S_d	S_{11}	S_{12}	S_{21}	S_{22}	S_x	$S_x{}'$	
0	1	1	0	0	1	0	1	1	0	$+4V_{dc}$
0	1	0	1	0	1	0	1	1	0	$+3V_{dc}$
0	1	0	1	0	1	1	0	1	0	$+2V_{dc}$
0	1	0	1	1	0	0	0	1	0	$+V_{dc}$
0	1	1	0	0	1	0	1	0	1	0
0	1	0	1	0	1	0	1	0	1	$-V_{dc}$
0	1	0	1	0	1	1	0	0	1	$-2V_{dc}$
0	1	0	1	1	0	0	0	0	1	$-3V_{dc}$
1	0	0	1	0	0	0	0	0	1	$-4V_{dc}$

Mode-2 ($+V_{dc}$): The power switches S_b, S_d, S_{12}, S_{22} and S_x are turned ON to generate the peak voltage of magnitude +84 V.
Mode-3 ($+2V_{dc}$): The power switches S_b, S_d, S_{12}, S_{21} and S_x are turned ON to generate the peak voltage of magnitude +56 V.
Mode-4: generates $+V_{dc}$, when the power switches S_b, S_d, S_{11} and S_x are kept ON, rest of the power switches being kept OFF and thus +28 V is generated across the output.
Mode 5 (0): Power switches S_b, S_c, S_{12}, S_{22} and S'_x are switched ON to generate zero voltage.
Mode-6 ($-V_{dc}$): The power switches S_b, S_d, S_{12}, S_{22} and S'_x are turned ON to generate the peak voltage of magnitude −28 V.
To achieve $-2V_{dc}$ (Mode-7), the power switches S_b, S_d, S_{12}, S_{21} and S'_x are kept ON, rest of the power switches being kept OFF. and thus −56 V is generated across the output.
To get $-3V_{dc}$ (Mode-8), the power switches S_b, S_d, S_{11} and S'_x are kept ON, rest of the power switches being kept OFF are switched OFF. and thus −84 V is generated across the output.
To get $-4V_{dc}$ (Mode-9), the power switches S_a, S_d and S'_x are kept ON, rest of the power switches being kept OFF are switched OFF, and thus −112 V is generated across the output.

4 Comparison with New Topologies

The power components required for the proposed inverter are compared with few of recent well-known MLIs reported in [8–13]. The generalized formulae of the different power components (total employed input DC supplies, N_{dc}, control switches, N_{sw}, uni-directional switches, N_U, bi-directional switches, N_B) of MLIs reported in [8–13] are illustrated in Table 2. In addition to the three conventional MLI topologies,

Table 2 Comparison of generalized formulae for the various power components

Types of MLI	Components count			
	(N_L)	(N_B)	(N_U)	(N_{DC})
[8]	($N_{sw}-1$)	NA	N_{sw}	($N_{sw}-2$)/2
[9]	($2N_{sw}-9$)	NA	N_{sw}	($N_{sw}-4$)/2
[10]	($N_{sw}-1$)	NA	N_{sw}	($N_{sw}-2$)/2
[11]	($2N_{sw}+3$)/3	NA	N_{sw}	(N_{sw}/4)
[12]	($N_{sw}-1$)	NA	N_{sw}	($N_{sw}-2$)/2
[13]	($8N_{sw}$/7) + 1	(N_{sw}/7)	($6N_{sw}$/7)	($3N_{sw}$/7)
Proposed topology	($2N_{sw}-5$)	NA	N_{sw}	(N_{sw}/2)

MLIs developed in recent years are drawing attention for their important features such as simple configuration of the suggested circuit, ease in maintainenece, and the economic cost. Few newly inverter topologies as reported in [8–13] are compared with the proposed topology.

In addition, the graph is incorporated between the voltage levels (N_L) versus power switches (N_{sw}) and voltage levels (N_L) versus driver circuits (N_{dr}) as depicted in Fig. 5a–b, respectively.

It can be found that power switch reduction reduces the number of driver circuits and makes the circuit less complicated, easier to implement and thereby making the overall inverter economical lower cost.

5 Simulation and Experimental Verification

It is chosen to design and test a 9-level inverter with the same values of the DC sources ($V_{ab} = V_{cd} = V_1 = V_2 = V_{dc} = 28$ V, $V_x = 112$ V) in MATLAB and being further implemented in laboratory prototype where the peak voltage of the output (maximum and minimum) is +112 V and −112 V, respectively. The simulations are performed employing the RL-Load considering $R = 90\ \Omega$, $L = 100$ mH. The corresponding results of the load voltage and the load current are presented for $t = 100$ ms and $t = 70$ ms as represented in Fig. 6a, b, respectively.

An experimental set-up for a 9-level symmetrical inverter is being developed in the laboratory. The simulation together with the experiment is carried out for the same load parameters as well as the same input DC supply magnitude for the performance verification of the proposed MLI as shown in Fig. 7a, b respectively. It can be noted that the findings of the experiment confirm the results of the simulation.

Moreover, a 13-level proposed inverter is also made having the magnitude of the input DC supplies in the fashion of 1:2:2, i.e. $V_2 = V_{cd} = V_{dc} = 21$ V, $V_1 = V_{cd} = V_{dc} = 42$ V $V_x = 126$ V, generating the peak voltage (maximum) of +126 V, and −126 V, the peak voltage (minimum) for the load value of $R = 95\ \Omega$, $L = 100$ mH. Figure 8a,

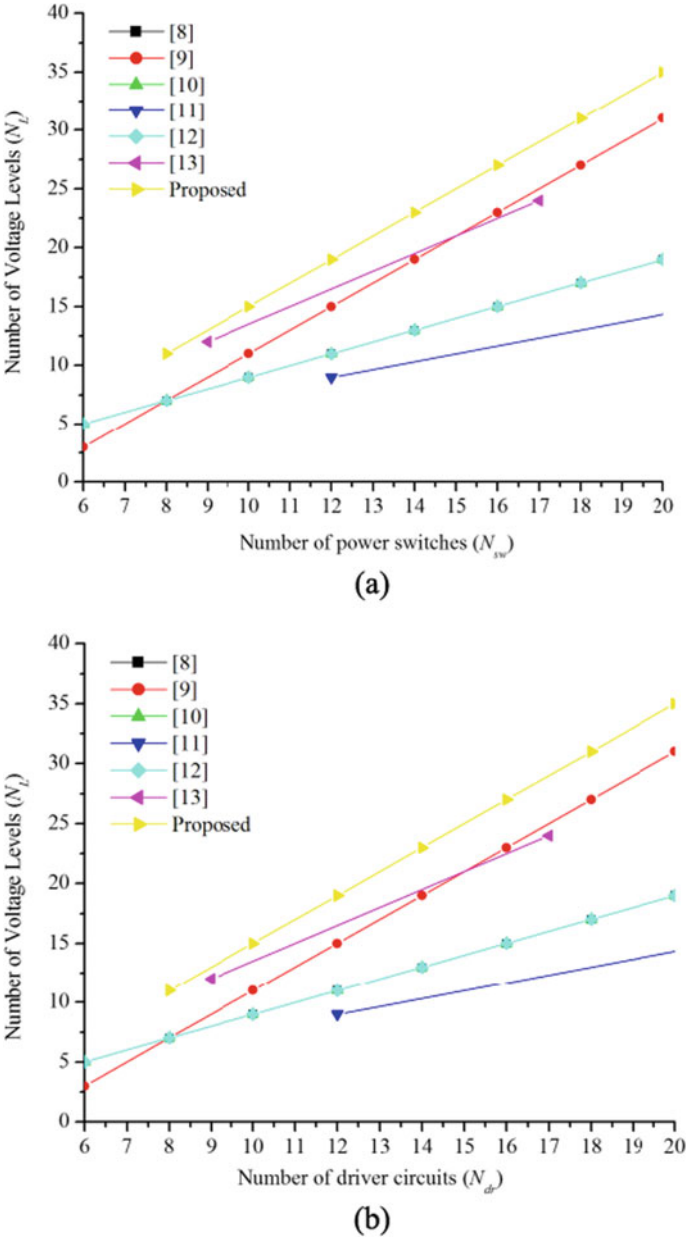


Fig. 5 Comparison graph plotted between proposed inverter and few suggested recent MLIs for, **a** N_{L} and N_{dr}, **b** N_{L} and N_{sw}

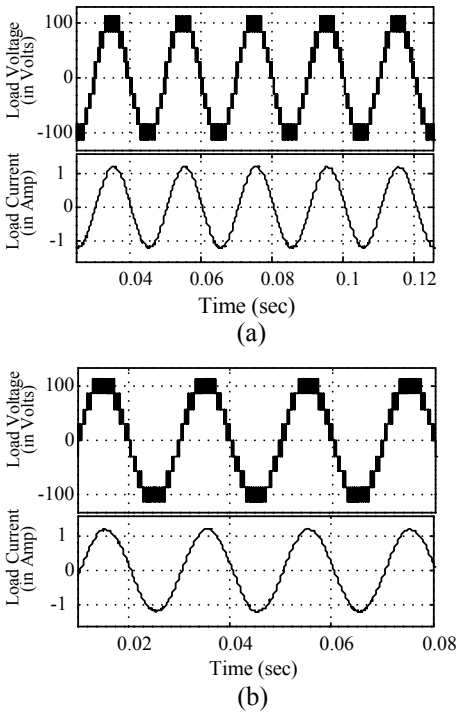


Fig. 6 Simulated results (load voltage and load current) for the proposed 9-level inverter for, **a** $t =$ 100 ms, **b** $t = 70$ ms

b indicates the output voltage with the load current for $t = 100$ ms and $t = 50$ ms, respectively. The experimental results (for both the symmetrical and asymmetrical) are incorporated to proof the effectiveness of the proposed MLI topology.

6 Conclusion

The topology suggested in this paper presents a new and generalized structure which is easily expanded by addition of the aforesaid explained units. The 9-level (symmetrical) and 13-level (asymmetrical) inverter based on proposed MLI structure is replicated in the simulation and the matching experimental outcomes for the same easily verifies the effective working of the inverter. This MLI structure shows better outcomes in terms of overall components required. In addition, the generalized equations for each required components are calculated for the proposed inverter as well as for the few recent inverter topologies as reported in [8–13]. Therefore, the comprehensive studies shows that the proposed inverter with reduced power electronic devices is comparatively better than few newly suggested similar type of multilevel inverters.

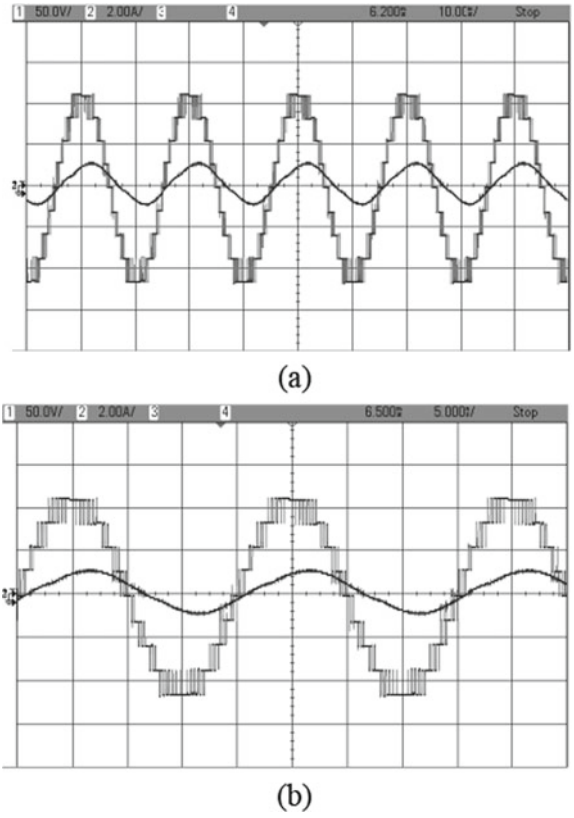


Fig. 7 Hardware results (load voltage and load current) for the proposed 9-level inverter for, **a** $t =$ 100 ms, **b** $t = 50$ ms

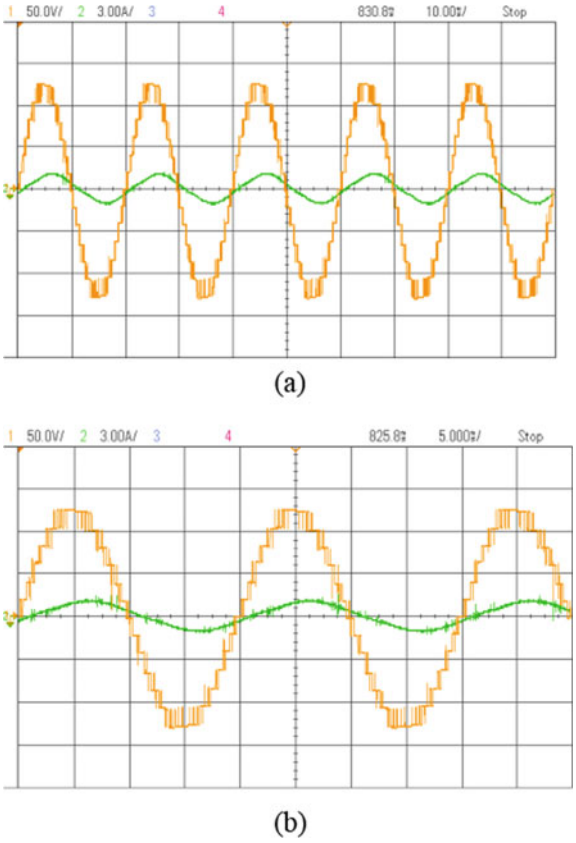


Fig. 8 Hardware results (load voltage and load current) for the proposed 13-level inverter for, **a** $t =$ 100 ms, **b** $t = 50$ ms

References

1. Williamson SS, Rathore AK, Musavi F (2015) Industrial electronics for electric transportation: current state-of-the-art and future challenges. IEEE Trans Ind Electron 62(5):3021–3032
2. Mahato B, Thakura PR, Jana KC (2014) Hardware design and implementation of unity power factor rectifiers using microcontrollers. In: 2014 IEEE 6th India international conference on power electronics (IICPE), pp 1–5
3. Jha KK, Mahato B, Prakash P (2016) Active power factor correction for rectifier using micro-controller. In: 2016 3rd international conference on recent advances in information technology (RAIT), pp 331–336
4. Rodríguez J, Lai JS, Peng FZ (2002) Multilevel inverters: a survey of topologies, controls, and applications. IEEE Trans Ind Electron 49(4):724–738
5. Rodriguez J, Bernet S, Steimer PK, Lizama IE (2010) A survey on neutral-point-clamped inverters. IEEE Trans Ind Electron 57(7):2219–2230
6. Sadigh AK, Hosseini SH, Sabahi M, Gharehpetian GB (2010) Double flying capacitor multicell converter based on modified phase-shifted pulsewidth modulation. IEEE Trans Power Electron 25(6):1517–1526
7. Malinowski M, Gopakumar K, Rodriguez J, Pérez MA (2010) A Survey on cascaded multilevel inverters. In: IEEE Trans Ind Electron 57(7): 2197–2206

8. Kangralu MF, Babaei E (2013) Cross-switched multilevel inverter: an innovative. IET Power Electron 6(4):642–651
9. Babaei E, Hosseini SH (2009) New cascaded MLI topology with minimum number of switches. J Energy Convers 50(11):2761–2767
10. Gupta KK, Jain S (2014) A novel multilevel inverter based on switched DC sources. IEEE Trans Ind Electron 61(7):3269–3278
11. Lee SS (2018) Single-stage switched-capacitor module (S^3CM) topology for cascaded multilevel inverter. IEEE Trans Power Electron: 1–1 (Early Acess). https://doi.org/10.1109/TPEL.2018.2805685
12. Gupta KK, Jain S (2013) Multilevel Inverter topology based on series connected switched sources. IET Power Electron 6(1):164–174
13. Gautam SP, Kumar L, Gupta S (2018) Single-phase multilevel inverter topologies with self-voltage balancing capabilities. IET Power Electron 11(5):844–855
14. Dahidah MSA, Konstantinou G, Agelidus VG (2015) Review of multilevel selective harmonic elimination PWM: formulations, solving algorithms, implementation and applications. IEEE Trans Power Electron 30(8):4091–4106
15. Mahato B, Raushan R, Jana KC (2016) comparative study of asymmetrical configuration of multilevel inverter for different levels. In: 2016 3rd international conference on recent advances in information technology (RAIT), pp 300–303
16. Mahato B, Jana KC, Thakura PR (2018) Constant V/f control and frequency control of isolated winding induction motor using nine-level three-phase inverter. Iranian J Sci Technol Trans Electr Eng 6(1): 1–13
17. Jana KC, Biswas SK, Chowdhury SK (2016) Dual reference phase shifted pulse width modulation technique for a N-level inverter based grid connected solar photovoltaic system. IET Renew Power Gener 10(7):928–935
18. Kumar C, Mahato B et al (2016) Comprehensive study of various configurations of three-phase Multilevel inverter for different levels. In: 2016 3rd international conference on recent advances in information technology (RAIT), pp 310–315
19. Mahato B, Majumdar S, Jana KC (2018) Carrier-based PWM techniques for multi-level inverters : a comprehensive performance study. J Sci Part A Eng Innov 5(3):101–111

Bidyut Mahato (Senior Research Fellow) pursuing Ph.D. from IIT (ISM), Dhanbad received B.Tech and M.E. degrees from WBUT, Kolkata and BIT, Mesra, Ranchi, in the year 2011 and 2014, respectively. Mr. Mahato has been declared as Chartered Engineer by The Institution of Engineers (INDIA). He is recipient of Quarterly Franklin Membership Award given by Editorial Board of London Journals Press UK (2020) for distinctive research contribution in Electrical Engineering. Currently, he acts as a Lead Guest Editor for the Special Issue, "Power Electronics & Drive System Applications" in the Journal International Journal of Electrical Components and Energy Conversion. Mr. Mahato Serving as a Joint Secretary for MCCS and NCCS organized by IETE & ISVE Ranchi Centre till date since 2018. He is Life-Member of ISTE, SSI and ISCA, Member of IAENG, Associate Member of IEI (India), the IRED and IETE. He is an active reviewer of IEEE Trans. on Industrial Electronics, IEEE Journal of Emerging and Selected Topics in Power Electronics, IET Power Electronics, Electric Power Components and Systems, Electric Power systems Research etc., and has reviewed more than 100 manuscripts till date. His

research interest includes Power electronic converters, multi-level inverters, Power factor correction and hybrid electric vehicles.

Saikat Majumdar (Junior Research Fellow) received his B.Tech and M.Tech from AOT and IIT (ISM), Dhanbad, in the year 2014 and 2016, respectively. He is working towards his Ph.D. from IIT (ISM), Dhanbad, since 2016 in the area of multi-level inverter, renewable energy sources, and drives.

Sambit Vatsyayan (M.Tech) pursuing Masters from IIT (ISM), Dhanbad, since 2017 in the domain of Power Electronics and Electrical Drives. He completed his B.Tech degree from IIT (ISM), Dhanbad, in 2017. His project interest includes multi-level inverters, AC/DC power conversion systems, and traction drives.

Kartick Chandra Jana (Asst. Prof.) received his M.Tech and Ph.D. degree from NIT Durgapur and Jadavpur University in 2003 and 2013, respectively. He is currently serving as assistant professor in IIT (ISM), Dhanbad, from June 2012. He has been working in Birla Institute of Technology, Mesra, from July 2003 till May 2012. His research is in area of modelling and design of multilevel inverter, real-time control of power electronics devices, and design and implementation of efficient power converters.

Design of Superlens Using 2D Triangular Photonic Crystal Under Both TE and TM Mode of Propagations

Moumita Banerjee, Pampa Debnath, and Arpan Deyasi

Abstract Dispersion characteristics have been analytically computed using plane wave expansion method for two-dimensional triangular photonic crystal for superlens design. Both types of polarizations have been taken into account for realistic calculation, and corresponding wavelengths for visible range are evaluated. Different combinations semiconductor materials are considered for simulation, for which already superlattice structures are formed; and it is seen that sub-wavelength resolution can be achieved for a wider spectrum in visible and UV ranges. More precisely, the entire superlens is designed using positive index materials, where Si–air combination outperforms GaAs-air array; and thus it can be revealed that metamaterial is not the priority for superlens construction.

Keywords Superlens · Dispersion characteristics · TE mode · TM mode · Plane wave method · Triangular lattice

1 Introduction

Lens is the basic tool used in the field of optics which basically works on the principle of far-field radiation. But conventional optical material, which is nothing but glass, cannot work with the near field radiation, though it is unable to propagate and stays

M. Banerjee
Department of Applied Physics, University of Calcutta, Kolkata, India
e-mail: banarjee.moumita@gmail.com

P. Debnath · A. Deyasi (✉)
Department of Electronics and Communication Engineering, RCC Institute of Information Technology, Kolkata, India
e-mail: deyasi_arpan@yahoo.co.in

P. Debnath
e-mail: poonam.4feb@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_46

very near to the conventional lens. Several works are carried to capture the near-field radiation [1–3] in the last decade for extracting information from the near-field radiation, and different materials are considered for that purpose. Though it is well-established, a few years back that metamaterials or LHM are quite capable to work with the near-field [4, 5], but people are also trying for the other materials which are compatible with the existing fabrication techniques. This is possible as the metamaterials can work with the all possible Fourier components of the image, whereas sharpness of conventional lens is limited by half-wavelength of light. This basic property drives the researchers to make design of novel superlens.

Photonic crystal is the novel realization of combination of materials in the form of 1D, 2D or 3D stacks [6, 7] which now be considered as the building block of photonic components, where photons are only responsible for carrier transport, a quite contrary to the optoelectronic devices. This novel apprehension can now be used as design of optical transmitter [8], receiver [9], sensor [10], communication channel [11], filter [12], information processing [13], etc. Recently, people are interested to work with the visible range using photonic crystal, for which analyzing dispersion properties becomes essential. This is the unique profile which speaks about its candidature for the lens design that can be applicable at sub-wavelength. The benefit of superlens design is that the objects, smaller than the wavelength of emitted/reflected light, can be visualized, and provided the spectra fall within the category of visible and UV ranges. In the present paper, this unique property is achieved by using semiconductor photonic crystal, where triangular lattice structure is chosen. Both TE and TM mode of propagations are taken into account for simulation, and for each case, several wavelengths are identified for the visualization purpose and also in ultra-violet range. Plane wave expansion method is utilized for computation, and results suggest the suitability of using proper combination of semiconductor materials for the superlens design purpose.

2 Mathematical Formulation

Plane wave equation for any electromagnetic wave can be written in the following form:

$$\vec{\nabla} \times \frac{1}{\varepsilon(\vec{r})} \vec{\nabla} \times \vec{H}(\vec{r}) = \frac{\omega^2}{c^2} \vec{H}(\vec{r}) \tag{1}$$

where magnetic intensity is defined as.

$$\vec{H}(\vec{r}) = \sum_{k,\lambda} \hat{p}(k, \lambda) h(k, \lambda) \exp\left[i \vec{k}.\vec{r} \right] \tag{2}$$

In Eq. (2), $\hat{p}(k, \lambda)$ denotes the unit polarization vectors. For transverse field, we can write.

$$\vec{k}.\hat{p}(k,\lambda) = 0 \tag{3}$$

This is the normalization condition. TM and TE waves under the normalization condition can be formulated as.

$$\Theta^{k}_{\mathrm{TM}}(V, V') = |k+V|\left|k+V'\right| f\left(V-V'\right) \tag{4a}$$

$$\Theta^{k}_{\mathrm{TE}}(V, V') = |k+V|.\left|k+V'\right| f\left(V-V'\right) \tag{4b}$$

where Θ_k matrix is defined in the following manner.

$$\begin{vmatrix} \Theta^k(\lambda_1,V_1,\lambda_1,V_1) & \Theta^k(\lambda_1,V_1,\lambda_2,V_1) & \Theta^k(\lambda_1,V_1,\lambda_1,V_2) & \Theta^k(\lambda_1,V_1,\lambda_2,V_2) & \ldots & \Theta^k(\lambda_1,V_1,\lambda_2,V_N) \\ \Theta^k(\lambda_2,V_1,\lambda_1,V_1) & \Theta^k(\lambda_2,V_1,\lambda_2,V_1) & \Theta^k(\lambda_2,V_1,\lambda_1,V_2) & \Theta^k(\lambda_2,V_1,\lambda_2,V_2) & \ldots & \Theta^k(\lambda_2,V_1,\lambda_2,V_N) \\ \Theta^k(\lambda_1,V_2,\lambda_1,V_1) & \Theta^k(\lambda_1,V_2,\lambda_2,V_1) & \ldots & \ldots & \ldots & \ldots \\ \Theta^k(\lambda_2,V_1,\lambda_1,V_1) & \Theta^k(\lambda_2,V_2,\lambda_2,V_1) & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \Theta^k(\lambda_2,V_N,\lambda_1,V_1) & \Theta^k(\lambda_2,V_N,\lambda_2,V_1) & \ldots & \ldots & \ldots & \Theta^k(\lambda_2,V_N,\lambda_2,V_N) \end{vmatrix} \tag{5}$$

In the reciprocal space representation, considering the packing fraction, we can formulate the space function as

$$f_{\mathrm{triangular}}(V) = \begin{cases} \frac{1}{\varepsilon_b} + \left(\frac{1}{\varepsilon_a} - \frac{1}{\varepsilon_b}\right) P_f & \text{for } V = 0 \\ \left(\frac{1}{\varepsilon_a} - \frac{1}{\varepsilon_b}\right) P_f \frac{2J_1(\mathrm{VR})}{(\mathrm{VR})} & \text{for } V \neq 0 \end{cases} \tag{6}$$

3 Results and Discussions

Using Eq. (6), dispersion characteristics is computed and plotted for investigation fo equal frequency contour.We have used MATLAB® software for simualtion. At first, GaAs–air combination is considerd for basic simualtion, and result is plotted in Fig. 1.

In the simulation, we have started to calculate with 0.836 packing fraction and 0.48 as fill factor. Once the material property is modified, it is customary to choose new packing fraction and fill factor to get the wavelengths in visible spectrum. Similar works are carried out for other material combinations, and results are displayed in the following figures. Here, the wavelengths are measured as 361.27 nm (2nd band), 328.51 nm (3rd band), 271.73 nm (4th band), and 211.77 nm (5th band). The first two wavelengths are near visible ranges while later two are under UV range. Similarly, plots are made under TE modes (Fig. 1b), and result shows similar behaviour.

But unfortunately, this configuration gives only 133.97 nm (2D band) and 91.97 nm (3rd band) wavelengths, which can safely be disregarded due to very higher frequency, and far away from desired spectrum.

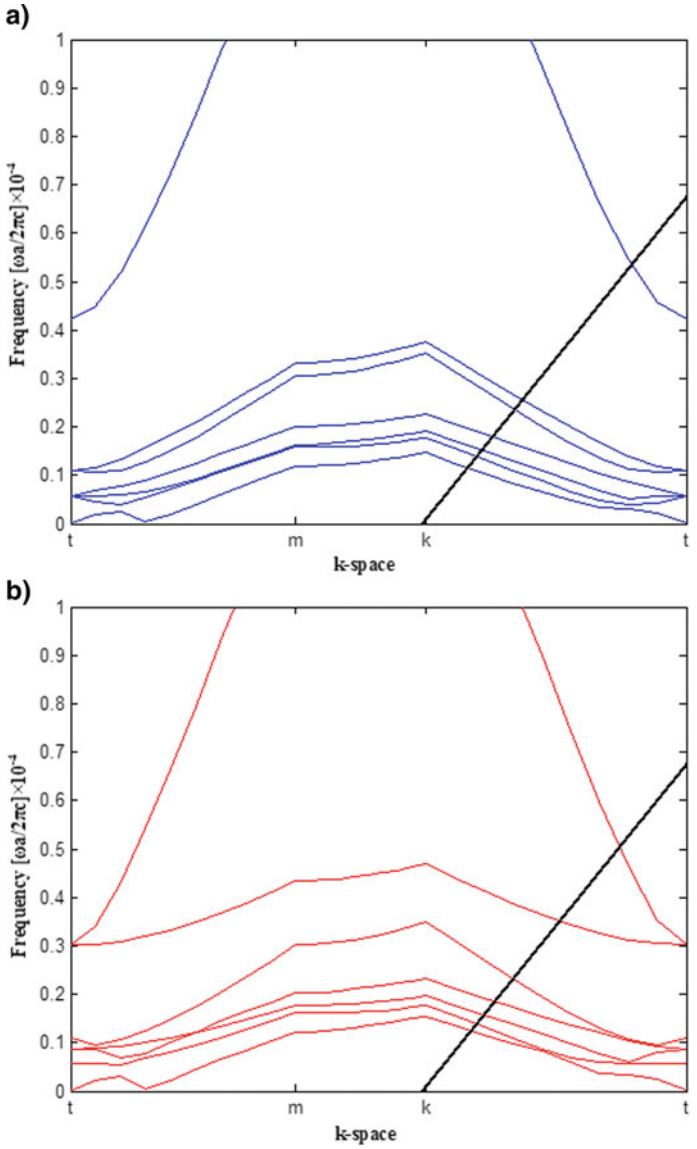


Fig. 1 **a** Dispersion curve with GaAs–air combination under TM mode, **b** dispersion curve with GaAs-air combination under TE mode

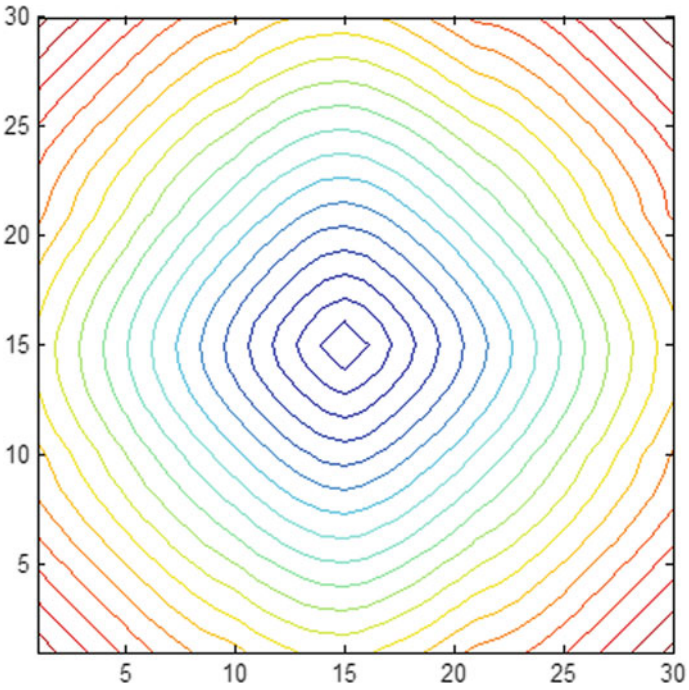


Fig. 2 Contour plot for GaAs–air composition for TE mode

The equal frequency contour plot is shown in Fig. 2 for TE mode, which justifies its working as superlens.

Figure 3a shows the data for TE mode propagation inside Si–air combination, with 0.34 packing fraction and 0.5 fill factor. From the characteristic profile, we get visible wavelength at 366.83 nm (2nd band), 318.26 nm (3rd band), 266.66 nm (4th band), and 195.84 nm (5th band). All the wavelengths are under UV ranges. For the same material combination, if TM mode is considered, then corresponding frequencies are 374.81 nm (2nd band), 321.33 nm (3rd band), 288.85 nm (4th band), and 260.68 nm (5th band), respectively. Figure 3b shows the profile.

The corresponding contour plot for TM mode is shown in Fig. 4. The inner lines are full circles which depicts the region where superlens can be formed.

From the dispersion curve, we find the intersection point of light-line, in which the EFC is circular and in this region, the negative R.I. property are seen, and at frequency range, the photonic crystal can be act as a super lens (Table 1).

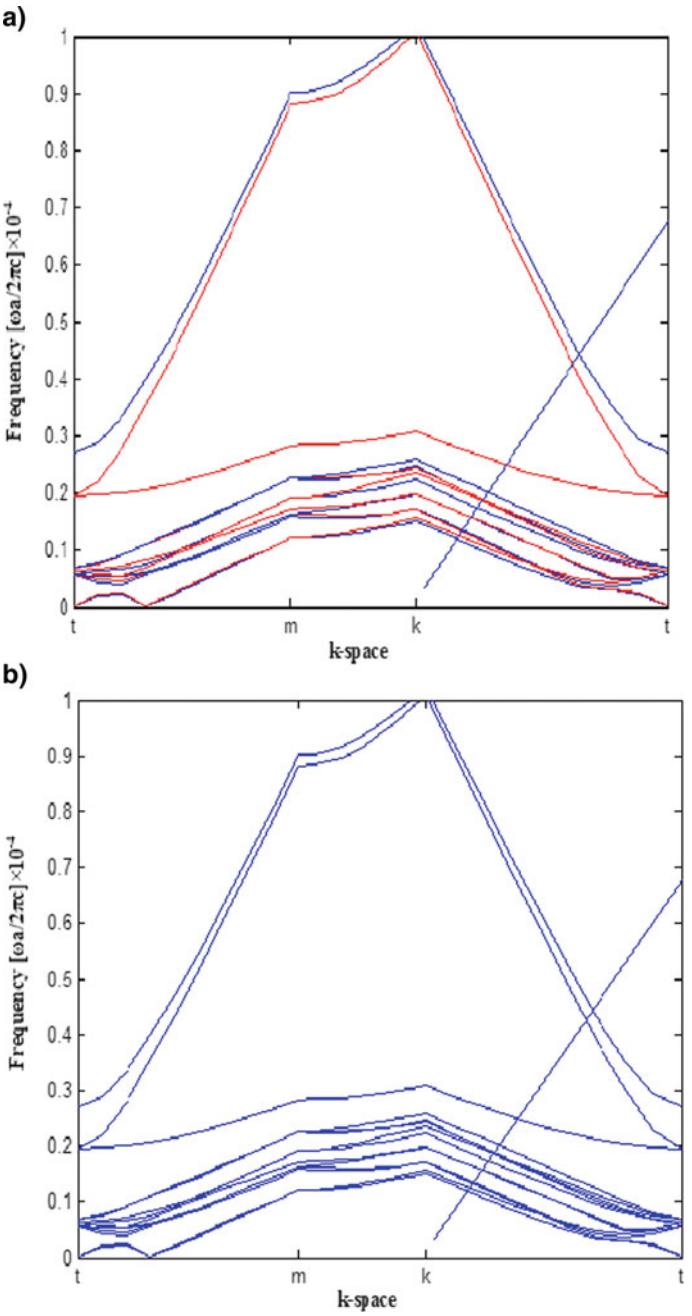


Fig. 3 **a** Dispersion curve with Si–air combination under TM mode, **b** dispersion curve with Si–air combination under TE mode

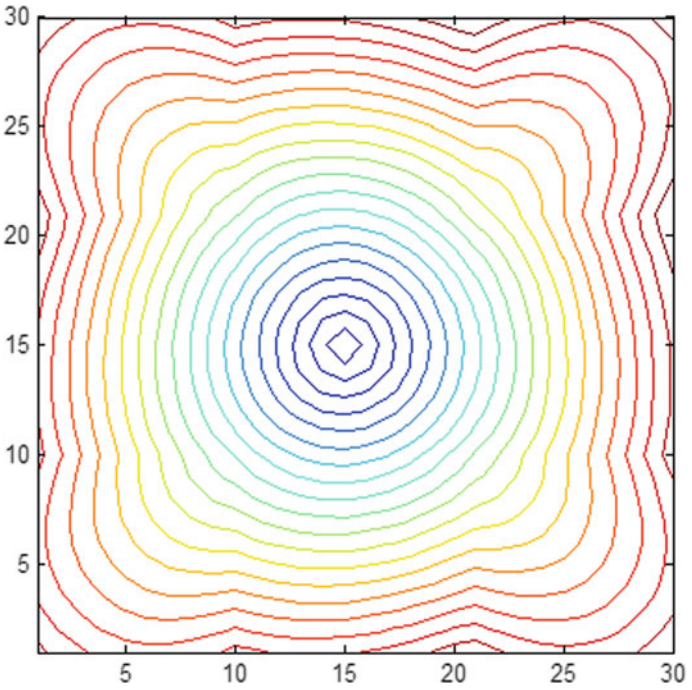


Fig. 4 Contour plot for Si–air composition for TM mode

Table 1 Wavelengths for different material combinations for superlends design

Band	Wavelength in nm for GaAs–air		Wavelength in nm for Si–air	
	TE mode	TM mode	TE mode	TM mode
2nd band	133.97	361.27	366.83	374.81
3rd band	91.97	328.51	318.26	321.33
4th band	…	271.73	266.66	288.85
5th band	…	211.77	195.84	260.68

4 Conclusion

Both GaAs–air and Si–air combinations have been tried for design of superlens at visible and UV ranges. For simulation purpose, both TE and TM modes of propagations are considered which clearly reveals that Si–air combination provides better output than the GaAs–air combination, and wavelengths are obtained at both the desirable ranges. This result also helps the designers to construct superlens using conventional materials which are compatible with the existing fabrication techniques.

References

1. Schoenlineer B, Wu X, Ebling JP, Eleftheriades GV, Rebeiz GM (2002) Wide-scan spherical-lens antennas for automotive radars. IEEE Trans Microw Theory Tech 50(9):2166–2157
2. Hua CZ, Wu X, Yang N, Wu W (2013) Air-filled parallel-plate cylindrical modified Luneberg Lens antenna for multiple-beam scanning at millimeter-wave frequencies. IEEE Trans Microw Theory Tech 64:436–443
3. Karttunen A, Saily J, Lamminen AEI, Ala-Laurinaho J, Sauleau R, Raisanen AV (2012) Using optimized eccentricity rexolite lens for electrical beam steering with integrated aperture coupled patch array. Prog Electromagnet Res B 44:345–365
4. Haxha S, AbdelMalek F, Ouerghi F, Charlton MDB, Aggoun A, Fang X (2018) Metamaterial superlenses operating at visible wavelength for imaging applications. Sci Rep 8:16119
5. Aydin K, Ozbay E (2007) Left-handed metamaterial based superlens for subwavelength imaging of electromagnetic waves. Appl Phys A 87(2):137–141
6. Limpert J, Schreiber T, Nolte S, Zellmer H, Tunnermann T, Iliew R, Lederer F, Broeng J, Vienne G, Petersson A, Jakobsen C (2003) High power air-clad large-mode-area photonic crystal fiber laser. Optic Express 11:818–823
7. Zhang Y, Kan Q, Wang GP (2014) One-way optical transmission in silicon grating-photonic crystal structures. Opt Lett 39:4934–4937
8. Xu Y, Lin J, Dubé-Demers R, LaRochelle S, Rusch L, Shi W (2018) A single-laser flexible-grid WDM silicon photonic transmitter using microring modulators. In: Optical fiber communication conference, OSA technical digest, paper W1I.3
9. Nozaki K, Matsuo S, Fujii T, Takeda K, Kuramochi E, Shinya A, Notomi M (2017) Forward-biased photonic crystal photodetector towards amplifier-free bias-free receiver. In: Conference on lasers and electro-optics, OSA technical digest, paper STh4N.1
10. Lu TW, Wu CC, Lee PT (2018) 1D photonic crystal strain sensors. ACS Photonics 5(7):2767–2772
11. Belhadj W, AbdelMalek F, Bouchriha H (2006) Characterization and study of photonic crystal fibres with bends. Mater Sci Eng C 26:578–579
12. Banerjee M, Dey R, Deyasi A, Dey S, Das A (2016) Butterworth filter design using ternary photonic crystal structure under polarized incidence of E.M Wave. In: springer proceedings in physics: advances in optical science and engineering, chap 24, pp 205–210
13. Azuma H (2008) Quantum computation with kerr-nonlinear photonic crystals. J Phys D Appl Phys 41:025102

Smart Intelligent System Design for Accident Prevention and Theft Protection of Vehicle

Shalini Sil, Subhomoy Daw, and Arpan Deyasi

Abstract In this paper, an automatic system is designed to prevent car accident in moving vehicle along with simultaneous protection of the theft of automobile. It is basically an IoT-based embedded system with built-in application, which works as assistance of driver to make the journey safer by taking crucial decisions at important time. The blynk app interface is set-up and registered with user's number, and it is real-time connected with the motor; so that any unidentified activity with the motor can be stopped using the mobile. DHT, smoke, alcohol, pulse, and speed sensors are connected once the motor is started which together helped to take various decisions at crucial juncture, and also to communicate the message to the registered android-enabled device. Therefore, this intelligent embedded system can provide the real guideline for a protected journey.

Keywords Intelligent system · Accident protection · Internet of things · Blynk app · Theft protection

1 Introduction

With an exponential increase in number of vehicles in last twenty years [1], while area for driving increases in a very slow linear manner, probability of car accident increases enormously in India in that passage of time [2]. Planning commission of Govt. of India already submitted a report of road safety and awareness programme along with rules and regulations for driving ten years ago [3], but that is not enough

S. Sil · S. Daw · A. Deyasi (✉)
Department of Electronics and Communication Engineering, RCC Institute of Information Technology, Kolkata, India
e-mail: deyasi_arpan@yahoo.co.in

S. Sil
e-mail: shalinisil81@gmail.com

S. Daw
e-mail: subhomoydaw16@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_47

with the latest reports published on road accident. Moreover, theft of vehicles is an increasing crime, and several reputed media have published reports on that [4, 5] in last few years. Henceforth, it is one of the challenging jobs for the technocrats and engineers to find suitable remedy for the drivers and car owners; in order to prevent road accidents as far as practicable, as well as to protect the car owners from theft. The present paper throws some light in this ever-increasing problem.

In this paper, a novel Internet of things-based embedded system is made which is a low-cost solution to the above-mentioned ever-increasing problem. The idea of theft-proof car is not a new one, and several research articles and innovative designs are already proposed for that purpose. Blynk application is added with a series of sensors in order to prevent accidents and thefts. DHT sensor is used to avert collisions under rainy conditions, alcohol sensor checks the maximum permissible consumption range, smoke sensor checks the environmental conditions under smog or sandstorms, pulse sensor checks the health condition which is allowable for driving, and finally, speed sensor checks for over-speeding. Though a few reports are published earlier [6–9] for design of smart app which will help the driver, but real-time implementations of this application are published less, and also the system is not taking care of environmental conditions. Moreover, individual systems are designed to prevent either accidents [10, 11] or thefts [12, 13]; but a combination of them is yet to be reported, as far the knowledge of the authors. In this paper, the combined system is represented, along with various data under working conditions for prevention of road accident as well as providing security to the owners. The next section of the paper deals with the working of various sensors, whereas results with real-circuit are illustrated in Sect. 3. Paper is summarized in Sect. 4.

2 Working of Various Sensors in Present System

The blynk app interface is setup and registered on a user's number. Multiple people can have access to it. Whenever the motor is stated, a notification is sent on the blynk app, so that the user knows it has been started. On identification of any unidentified activity, the user has the ability to stop the car at the press of a button in the app.

2.1 DHT Sensor

The moment the module is started, the sensors, namely DHT sensor, smoke sensor, alcohol sensor, pulse sensor, and speed sensor start taking values.

The DHT sensor shows the temperature and humidity values on Blynk as the car starts, these are very important parameters in taking better driving decisions. Rains are among the top five causes of accidents in vehicles.

2.2 Alcohol Sensor

The alcohol sensor takes the alcohol value in driver's breath and checks if the value is above prescribed legal values for driving. If not,

- A notification is sent on blynk, that the driver is drunk, provided multiple access even his/her family/friends get to know of this.
- An alert mail is sent to a registered mail id about the situation
- The motor is stopped with immediate effect.
- The module backlight blinks as a norm which considers car stopping at the middle of the road should have its backlight blinking.

2.3 Smoke Sensor

The smoke sensor takes the parts per million values in driver's breath and checks if the value is above safe environmental level. Since phenomena like photochemical smog, sandstorms, and "Dhund (India)" are all pollutant based and reduce visibility to a great extent, causing frequent accidents. If unsafe values get scanned:

- A notification is sent on blynk, that the visibility is low,
- The buzzer blares alerting the driver.
- If the above condition is scanned along with proximity from any large-sized obstacle less than safe distances,
- A notification is sent on blynk, that the "visibility is low, driver at risk".
- An alert mail is sent to a registered mail id about the situation.
- The motor is stopped with immediate effect.
- The module backlight blinks as a norm which considers car stopping at the middle of the road should have its backlight blinking.

2.4 Pulse Sensor

The pulse sensor fitted to the steering keeps checking the pulse rate of driver. In case the pulse rate is below the prescribed rate of 50 bpm, a physical condition called bradycardia prevails. It severely hinders blood flow to brain, hence hampering decision taking. This can happen due to sleepiness or sickness. In such a case:

- A notification is sent on blynk, that the driver's "pulse is dropping".
- An alert mail is sent to a registered mail id about the situation.
- The buzzer and a bright-coloured led tries to bring the driver back to alertness.

2.5 Speed Sensor

The speed sensor checks for over speeding. On detecting over speeding:

- A notification is sent on blynk, that "the driver's over speeding".
- Buzzer blares alerting him/her about the same.

As the vehicle is started, the location of the same is also shown on blynk and can be shared with people if the driver wants.

All these features provide for a compact, IoT app-based car accident remedy, and car safety module with an advanced intelligent driver assistant, which has the long-time goal of reducing vehicular accidents as much as possible alongside helping drivers be more aware of ambient conditions, his own mistakes while driving, and hence take better decisions to overall create a safe driving environment.

3 Architecture and Workflow

The architecture for the design is explained using the block diagram as shown in Fig. 1. Once the motor starts, all the sensors are in active mode, and blynk app is set. In case, if someone unwanted person wants to start the car, the system will send an

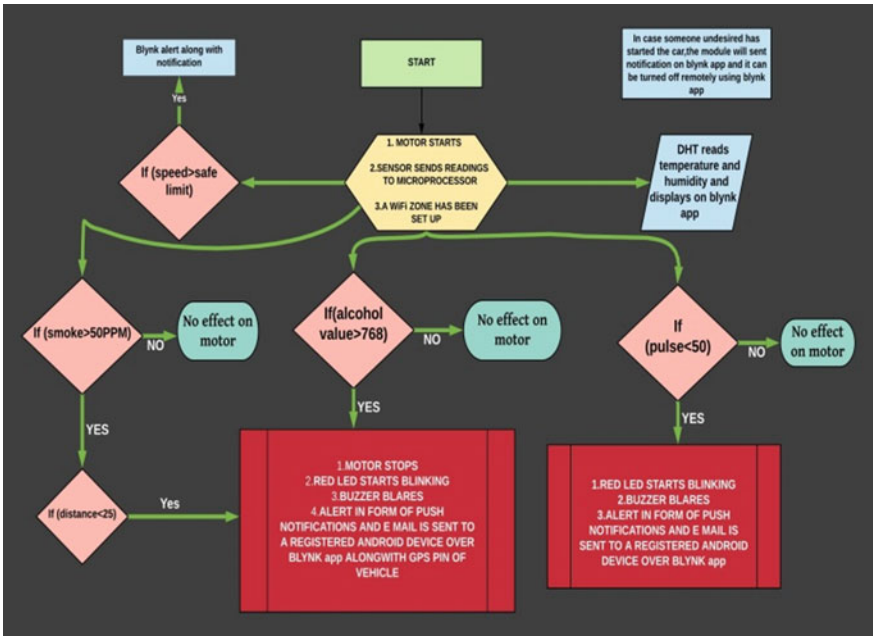


Fig. 1 Block diagram of the design

signal/notification to the blynk application, and motor will be turned off as soon as possible.

If the authorized person starts the motor, the DHT sensor reads the temperature and humidity data. If spped is greater than pre-defined limit, then the app will show the data to the driver. If smoke is greater than 50 ppm (limit is defined for this system, it can be changed), or if alcohol value exceeds 768, or if pulse rate goes below 50, then for each case, the motor stops and red LED starts blinking. For all the cases, alert message will go to the registered mobile number along with GPS location of the car.

Corresponding circuit is shown in Fig. 2.

In this context, it is customary to mention about the threshold levels set in the circuit.

According to the law, individuals found with 30 mg per 100 ml of alcohol in their blood are considered incapable of driving [14]. According to datasheet, 30 mg/100 ml corresponds to an analog value of 768.

During sleep, a slow heartbeat with rates around 40–50 b.p.m is common and is considered normal [15].

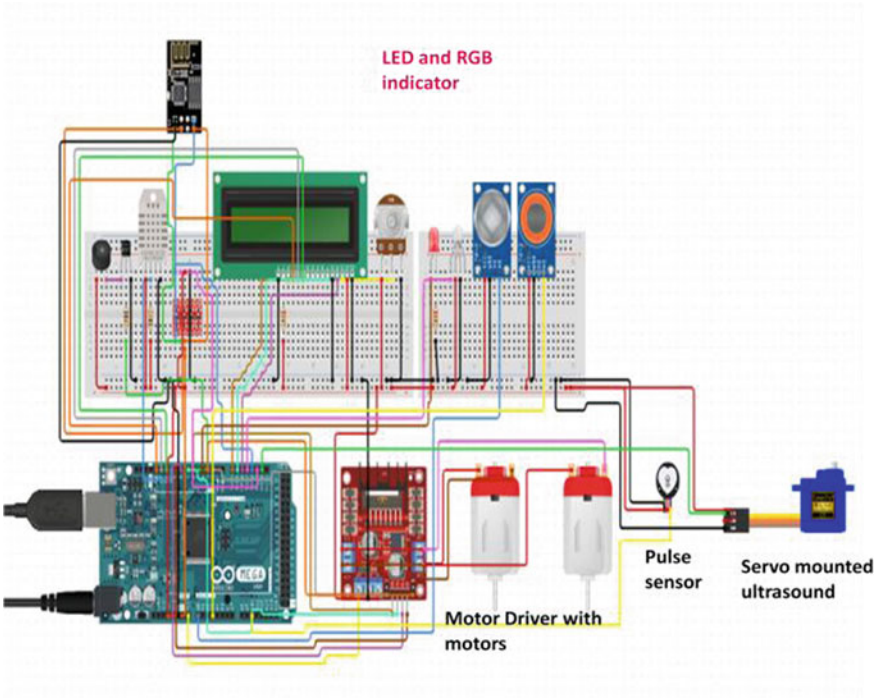


Fig. 2 Real-time ciruit representation of the proposed architecture

According to datasheet, safe ppm value of smoke is 50; hence, we have taken a value threshold of 60 to consider unsafe value of smoke leading to visibility loss [16, 17].

4 Results

Based on the above circuit, a few runtime snapshots are captured, and displayed in Fig. 3 for differnet adverse conditions. Corresponding log report is also shown.

Therefore, all the adverse conditions are taken care as far as practicable in this prototype, and successfully implemented.

5 Conclusion

This prototype can therefore be used as car safety modules to resist against a few most major accident causing phenomena. The present system can effectively prevent and alert against accident caused by excess smoke, alcoholism drowsiness, etc. It has the potential to revolutionize the vehicle safety market, as the already existing reports and literatures never consider both theft and accident prevention conditions simultaneously.

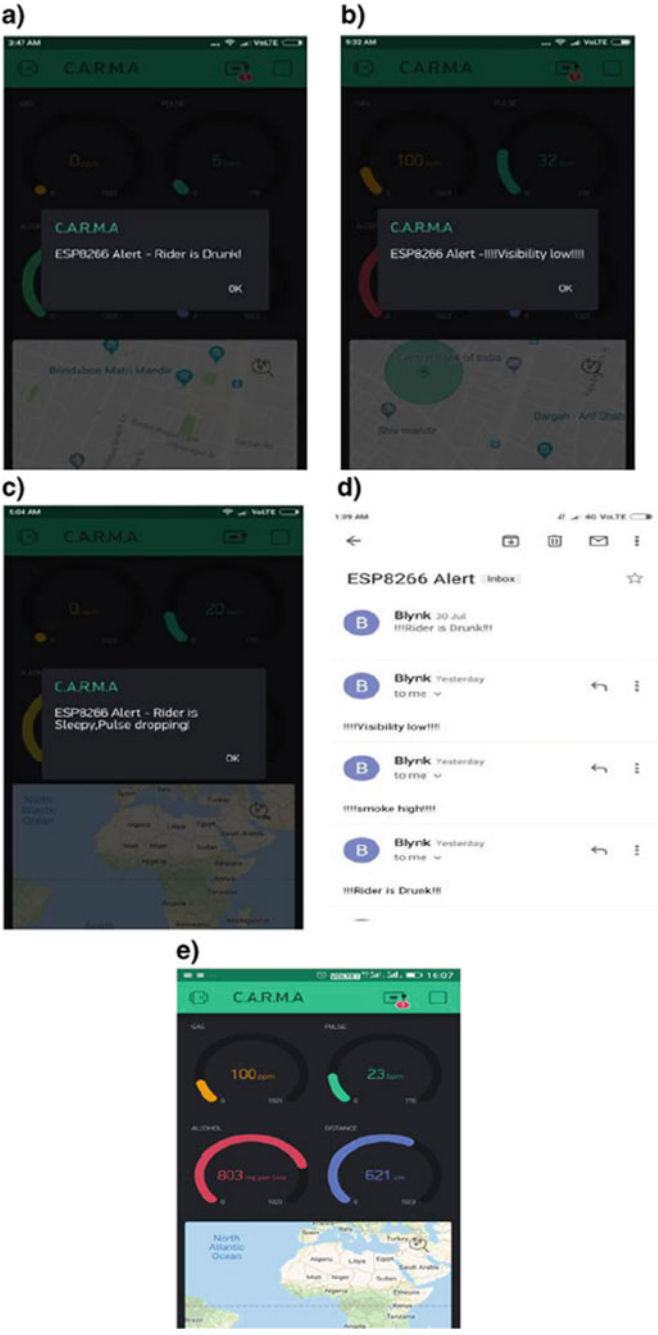


Fig. 3 **a** Alert message under drunk condition, **b** alert message under low visibility condition, **c** alert message under low pulse rate, **d** log of messages under different conditions, **e** display panel of safety parameters under run-time conditions

Acknowledgements The prototype, already displayed in the Dream Big Championship at Internshala held at New Delhi 2019, and achieved 3rd rank. This project is partially funded by RCC Institute of Information Technology.

References

1. https://www.siamindia.com/statistics.aspx?mpgid=8&pgidtrail=9
2. https://www.indiaenvironmentportal.org.in/files/file/road%20accidents%20in%20India%202017.pdf
3. Report of the Committee on Road Safety and Traffic Management, The Secretariat for the Committee on Infrastructure, Planning Commission, Government of India, Feb 2007
4. https://www.newindianexpress.com/business/2019/jul/10/how-big-of-a-problem-is-vehicle-theft-in-india-2002048.html
5. https://www.thehindu.com/news/cities/Delhi/police-data-show-motor-vehicle-theft-the-least-solved-crime/article25954331.ece
6. Vatti NR, Vatti PL, Vatti R, Garde C (2018) Smart road accident detection and communication system. In: IEEE international conference on current trends towards converging technologies
7. Nanda S, Joshi H, Khairnar S (2018) An IOT based smart system for accident prevention and detection. In: IEEE 4th international conference on computing communication control and automation
8. Selvathi D, Pavithra P, Preethi T (2017) Intelligent transportation system for accident prevention and detection. In: IEEE international conference on intelligent computing and control systems
9. John A, Nishanth PR (2017) Real time embedded system for accident prevention. In: IEEE international conference of electronics, communication and aerospace technology
10. Massoudian N, Eshghi M (2015) Design an intelligent curve to reduce the accident rate. In: IEEE signal processing and intelligent systems conference
11. Mohamedaslam C, Roshan TA, Sahal MTM, Najeeb NA, Nisi K (2016) A smart vehicle for accident prevention using wireless blackbox and eyeblink sensing technology along with seat belt controlled ignition system. In: IEEE international conference on green engineering and technologies
12. Bavya R, Mohanamurali R (2014) Next generation auto theft prevention and tracking system for land vehicles. In: IEEE international conference on information communication and embedded systems
13. Bagavathy P, Dhaya R, Devakumar T (2011) Real time car theft decline system using ARM processor. In: IEEE 3rd international conference on advances in recent technologies in communication and computing
14. https://sites.ndtv.com/roadsafety/5-things-know-breathalysers-2061/
15. https://en.wikipedia.org/wiki/Heart_rate
16. https://www.dieselnet.com/standards/eu/ld.php
17. https://www.haoyuelectronics.com/Attachment/MQ-2/MQ-2.pdf

Computing Threshold Voltage Shift in Submicron MOSFET in Presence of Body Effect and Induced Lateral Field

Farnaz Khanam, Anwita Nath, and Arpan Deyasi

Abstract Shift of threshold voltage in submicron MOSFET is analytically computed in the presence of induced lateral electric field and body effect. Two-dimensional Poisson's equation is solved with appropriate boundary conditions, and shift of threshold is estimated as a function of channel length for different external bias, doping concentration, and dielectric properties. Effect of flatband voltage is also taken into account while calculating threshold, which reduces the voltage. It is found that the presence of high-K dielectric and lower dielectric thickness effectively reduces the percentage of shift, and it is only significant at lower channel length. Consideration of body effect significantly reduces the threshold voltage shift, which is crucial from design point of view. Results are critically important for ultra short- channel devices.

Keywords Induced field · Dielectric thickness · Dielectric constant · Doping concentration · External voltage · Flatband voltage

1 Introduction

Short-channel effect is one of long-standing problem for the VLSI engineers [1–3] over the last decade as it basically degrades the device performance along with simultaneous failure of gate control over the channel carriers. But very recently, a few landmark results are published where gate length is reduced beyond 20 nm [4–6], and several novel structures are proposed [7, 8] for increase in gate control. As far ITRS

F. Khanam · A. Nath · A. Deyasi (✉)
Department of Electronics and Communication Engineering, RCC Institute of Information Technology, Kolkata, India
e-mail: deyasi_arpan@yahoo.co.in

F. Khanam
e-mail: farnazkhantabish@gmail.com

A. Nath
e-mail: nathanwita@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_48

roadmap [9] is concerned, investigation of transistor properties when gate length reaches beyond 10 nm is the prime concern, and therefore, several new complex geometries [10, 11] as well as new material compositions [12, 13] are proposed to reduce short-channel effect. A combination of low-K and high-K materials is one of the major players in this design procedure, where threshold voltage computation remains the key concern [14] due to increased probability of leakage current.

DGMOSFET, triple-gate MOSFET, and CNTFET devices are more popular nowadays as it effectively lowers the threshold voltage at shorter gate length, because of added gate control. But conventional MOSFET suffers this problem, so far reported [15]. But it has the advantage of existing fabrication technologies, and it supported the established setups. Henceforth, a compromise is somewhat required at different stages, and it becomes the duty of researchers to measure the effects with single gate terminal so that complex geometries and multiple gate terminals can intentionally be added for specific requirements to innovate the designs.

Owing to the presence of lateral induced electric field, shift of threshold takes place, and it plays pivotal role in determining subthreshold current of the device. This effect is more critical at lower channel length, and this shifts the threshold point, as lateral diffusion is quite severe at lower gate length. Thus, it becomes customary to investigate the effect on both threshold and shift of threshold, in the presence of conventional dielectric material, as well as for high-K. In this present paper, shift of threshold is computed analytically after solving two-dimensional Poisson's equation, where several external factors are considered. Expression of induced field is carried out, and shift is measured as a function of that. Here, surface potential is very important, as it controls the induced field. In the next section, mathematical formulation is mentioned, whereas Sect. 3 contains results obtained analytically. Findings are summarized by conclusion.

2 Mathematical Formulation

Considering quantum confinement, expression of charge density can be obtained by solving 2D Poisson's equation.

$$\frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = -\frac{qN_A}{\varepsilon_s} \tag{1}$$

where all the field components are assumed inside depletion region only.

Introducing appropriate boundary conditions and neglecting the body effect coefficient, we can rewrite the expressions of potential expressions as [16]

$$\frac{\partial E_y}{\partial y} = \frac{E_y(0,z) - E_d(W,z)}{W} \tag{2a}$$

$$\frac{\partial E_z}{\partial z} + \frac{1}{l^2}[V_{GS} - V_{fb} - \phi_{sp}(z)] = -\frac{\eta q N_A}{\varepsilon_s} \tag{2b}$$

where 'W' is the depletion width and 'l' is the characteristic length of channel. Here, we have considered $E_y(0, z)$ is the field component along the semiconductor–dielectric interface, $E_d(0,z)$ is the field component confined inside dielectric region only.

Finally, the expression of surface potential may be put in the following form [16]

$$\begin{aligned}\phi_{sp}(z) = \phi_{sp}(L) &+ \left[V_{DS} + V_{bi} - \phi_{sp}(L)\right]\frac{\sinh(z/l)}{\sinh(L/l)} \\ &+ \left[V_{bi} - \phi_{sp}(L)\right]\frac{\sinh[(L-z)/l]}{\sinh(L/l)}\end{aligned} \tag{3}$$

and lateral induced field is given by [16]

$$\begin{aligned}E_z = E_s(L) &- \frac{1}{l}\left[V_{\mathrm{DS}} + V_{\mathrm{bi}} - \phi_{\mathrm{sp}}(L)\right]\frac{\cosh(z/l)}{\sinh(L/l)} \\ &- E_s(L)\frac{\sinh(z/l)}{\sinh(L/l)} + E_s(L)\frac{\sinh[(L-z)/l]}{\sinh(L/l)} \\ &\frac{1}{l}\left[V_{\mathrm{bi}} - \phi_{\mathrm{sp}}(L)\right]\frac{\cosh[(L-z)/l]}{\sinh(L/l)}\end{aligned} \tag{4}$$

In the presence of body effect, Eqs. (3) and (4) will be modified as

$$\begin{aligned}\phi_{\mathrm{sp}}(z) = \phi_{\mathrm{sp}}(L) &+ \left[V_{\mathrm{DS}} + V_{\mathrm{SB}} + V_{\mathrm{bi}} - \phi_{\mathrm{sp}}(L)\right] \\ &\times \frac{\sinh(z/l)}{\sinh(L/l)} + \left[V_{\mathrm{bi}} + V_{\mathrm{SB}} - \phi_{\mathrm{sp}}(L)\right]\frac{\sinh[(L-z)/l]}{\sinh(L/l)}\end{aligned} \tag{5}$$

$$\begin{aligned}E_z = E_s(L) &- \frac{1}{l}\left[V_{\mathrm{DS}} + V_{\mathrm{SB}} + V_{\mathrm{bi}} - \phi_{\mathrm{sp}}(L)\right] \\ &\times \frac{\cosh(z/l)}{\sinh(L/l)} - E_s(L)\frac{\sinh(z/l)}{\sinh(L/l)} \\ &+ E_s(L)\frac{\sinh[(L-z)/l]}{\sinh(L/l)} \\ &\frac{1}{l}\left[V_{\mathrm{bi}} + V_{\mathrm{SB}} - \phi_{\mathrm{sp}}(L)\right]\frac{\cosh[(L-z)/l]}{\sinh(L/l)}\end{aligned} \tag{6}$$

After locating the minimum point for surface potential, we can compute the threshold voltage shift as

$$\begin{aligned}\Delta V_T = &[2(V_{\mathrm{bi}} + V_{\mathrm{SB}} - \phi_{s\,\mathrm{min}}) + V_{\mathrm{DS}}] \\ &\times \left(\exp\left(-\frac{L}{2l}\right) + 2\exp\left(-\frac{L}{l}\right)\right)\end{aligned} \tag{7}$$

3 Results and Discussions

Using Eq. (7), shift of threshold voltage is computed and plotted as a function of normalized channel length. Results are plotted with different parametric variations and compared with the result when body effect was absent, as published in reputed literature [17].

Figure 1 shows the threshold voltage shift in the presence and absence of body effect in submicron channel length, where comparison is made for both low-K as well as for high-K dielectric materials. For low-K material, result is compared with that obtained following [17], and it is found that shift decreases when body effect is taken into account. Shift is monotonically decreasing with increase of channel length. This is due to the fact that in submicron length, more precisely, close to nanometric length, surface potential increases, which, in turn, makes a large change of threshold voltage. Higher drain bias also increases the shift.

In the presence of body bias, effect of high-K materials is computed and plotted in Fig. 2. From the plot, it is seen that increasing the gate dielectric constant nullifies the threshold voltage shift, and the shift is very significant when conventional dielectric material is used. Lowering the channel length, i.e., more submicron region increases the shift for low-K materials.

Next we investigated the effect of dielectric thickness for HfO_2, as it gives minimum shift. Higher thickness increases the shift, as it restricts the tunneling by decreasing the gate capacitive effect, and therefore, threshold voltage changes. Result is graphically represented in Fig. 3.

If source doping concentration is varied, then it affects the threshold. If the difference with the background doping, i.e., with channel doping is large, then shift increases. But if doping of both the regions is very close then shift can be tuned to negligible amount. Result is shown in Fig. 4. It may be noted that all the calculations are made considering HfO_2 as dielectric and also in presence of body effect.

4 Conclusion

Shift in mV range is observed when threshold voltage is calculated for submicron MOSFET. Results reveal that lower shift can be observed if HfO_2 [high-K] is considered as dielectric material, with nanometric thickness, and lower electric field. All the computations are carried out in the presence of induced field, where it is shown that intentionally if body effect of suitable magnitude can be added in the device, it helps to reduce the shift of threshold. But if lowering of threshold is required, then body effect should be excluded.

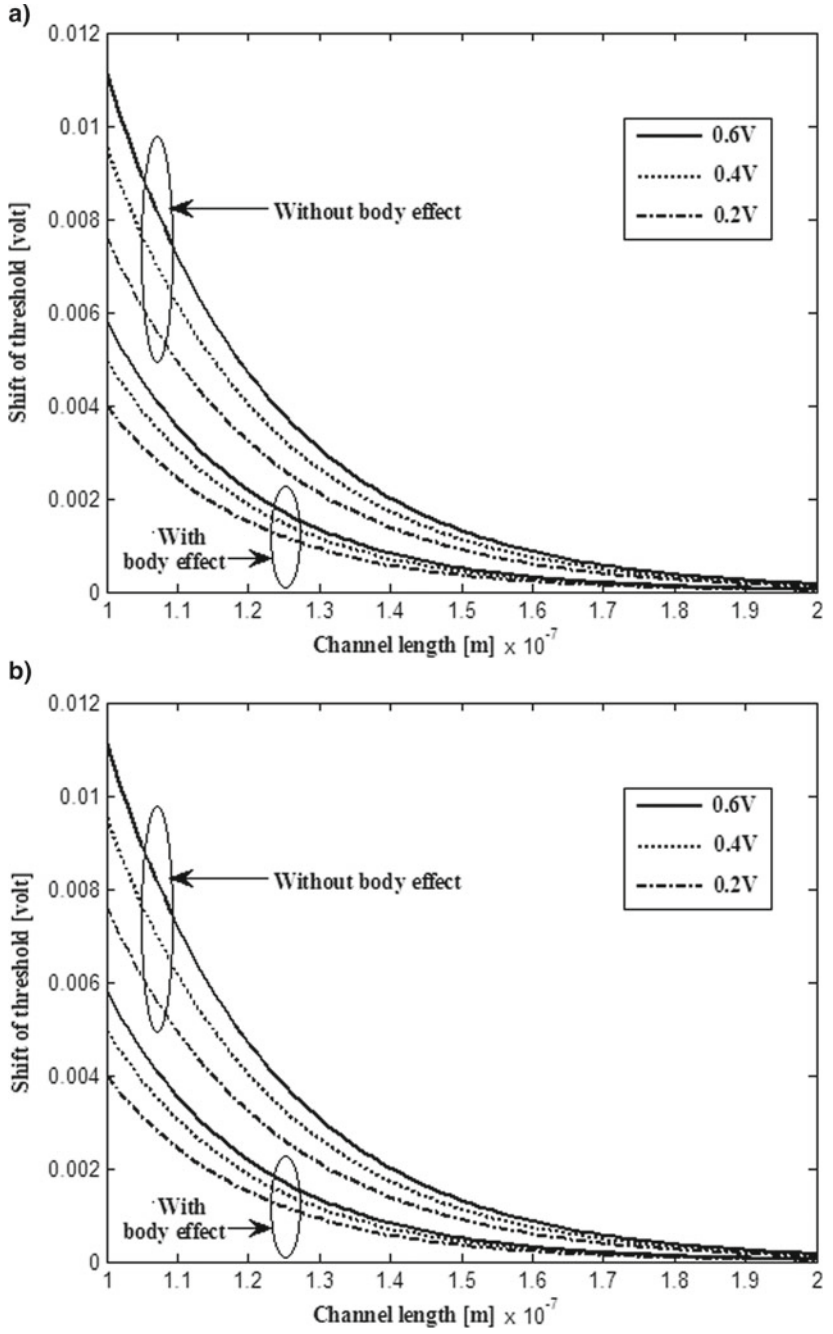


Fig. 1 **a** Comparative study of threshold voltage shift with channel length for different drain bias in the presence and absence of body effect using low-K dielectric; data in the absence of body effect is calculated following [17], **b** comparative study of threshold voltage shift with channel length for different drain bias in the presence and absence of body effect using high-K dielectric

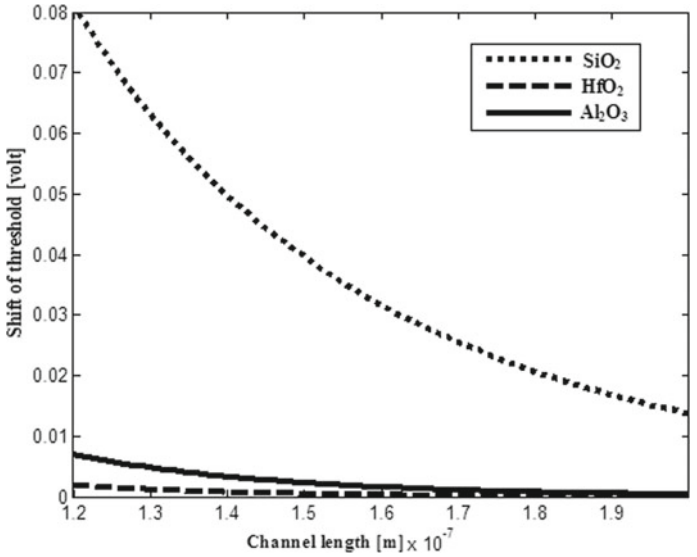


Fig. 2 Effect of threshold voltage shift for different dielectrics

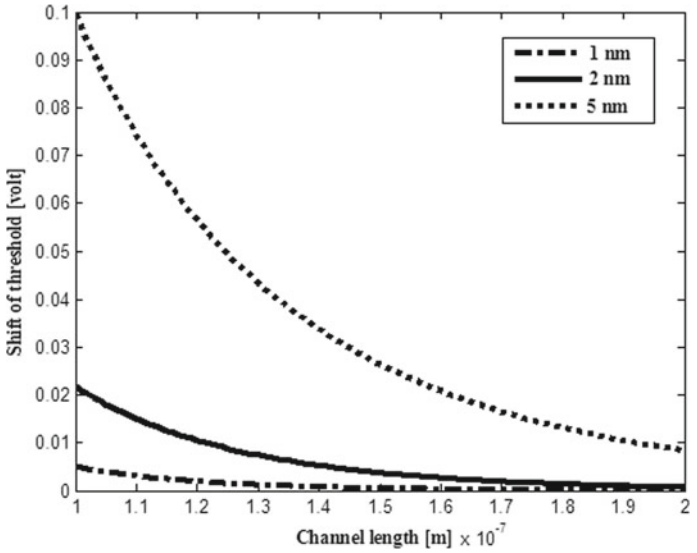


Fig. 3 Effect of threshold voltage shift for different dielectric thicknesses when HfO_2 is considered

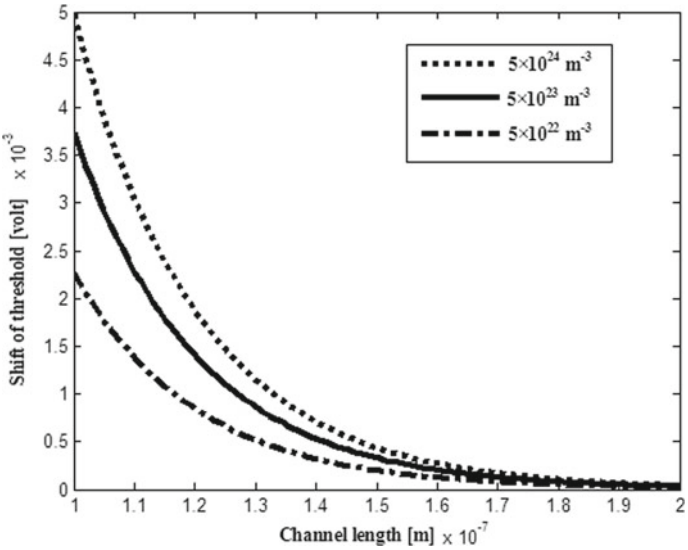


Fig. 4 Effect of threshold voltage shift for different doping concentrations when HfO_2 is considered

References

1. Kim YB (2010) Challenges for nanoscale MOSFETs and emerging nanoelectronics. Trans Electr Electron Mater 11(3):93–105
2. Gupta KA, Anvekar DK, Venkateswarlu V (2013) Comparative study and analysis of short channel effects for 180 nm and 45 nm transistors. Adv Intell Syst Comput Adv Comput Inf Technol 178:707–715
3. Djeffal F, Ghoggali Z, Dibi Z, Lakhdar N (2009) Analytical analysis of nanoscale multiple gate MOSFETs including effects of hot-carrier induced interface charges. Microelectron Reliab 49:337–381
4. Jagtap SM, Gond VJ (2017) Study the performance parameters of novel scale FINFET Device in nm Region. In: International conference of electronics communication and aerospace technology
5. Asthana PK, Goswami Y, Ghosh B (2016) A novel sub 20 nm single gate tunnel field effect transistor with intrinsic channel for ultra low power applications. J Semicond 37(5): 054002
6. Das UK, Eneman G, Shankar R, Velampati R, Chauhan YS, Jinesh KB, Bhattacharyya TK (2018) Consideration of UFET Architecture for the 5 nm node and beyond logic transistor. IEEE J Electron Devices Soc 6:1129–1135
7. Hiramoto T, Nagumo T (2006) Multi-Gate MOSFETs with back-gate control. In: IEEE international conference on ic design and technology
8. Rezgui H, Nasri F, Aissa MFB, Belmabrouk H, Guizani AA (2018) Modeling thermal performance of nano-GNRFET transistors using ballistic-diffusive equation. IEEE Trans Electron Dev 65(4):1611–1616
9. https://www.itrs2.net/
10. Woo HJ, Jin KC, Kyu CY (2008) Universal potential model in tied and separated double-gate MOSFETs with consideration of symmetric and asymmetric structure. IEEE Trans Electron Dev 55(6): 1472–1479
11. Deyasi A, Sarkar A (2018) Analytical computation of electrical parameters in GAAQWT and CNTFET using NEGF method. Int J Electron 105(12): 2144–2159

12. Singh D, Panda S, Mohapatra SK, Pradhan KP, Sahu PK (2014) Static performance analysis on UTB-SG and DG MOSFETs with Si and III–V Channel Materials. In: International conference on high performance computing and applications
13. Chen YH, Chan WM, Wu WC, Liao HJ, Pan KH, Liaw JJ, Chung TH, Li Q, Chang GH, Lin CY, Chiang MC, Wu SY, Natarajan S, Chang J (2014) A 16 nm 128 Mb SRAM in High-K Metal-Gate FinFET Technology with Write-Assist Circuitry for Low-V Applications. In: Proceeding of IEEE international solid-state circuits conference
14. Abdelhamid H, Anis AM, Aboulwafa ME, Eladawy MI (2019) Magnetic field effect on threshold voltage for ultrathin silicon gate-all-around nanowire field-effect-transistors. Silicon: 1–9
15. Frank DJ, Dennard RH, Nowak E, Solomon PM, Taur Y, Wong HSP (2001) Device scaling limits of Si MOSFETs and their application dependencies. Proc IEEE 89:259–288
16. Terril KW, Hu C, Ko PK (1984) An analytical model for the channel electric field in MOSFETs with graded-drain structures. IEEE Electron Dev Lett 5(11):440–442
17. Liu Z-H, Hu C, Huang J-H, Chan T-Y, Jeng M-C, Ko PK, Cheng YC (1993) Threshold voltage model for deep-submicrometer MOSFETs. IEEE Trans Electron Dev 40(1):86–95

Proficient Discrete Wavelet Transform Using Distributed Arithmetic Architecture on FPGA

K. B. Sowmya, Divya Jamakhandi, and Jose Alex Mathew

Abstract Discrete wavelet transform (DWT) is a filter where discrete samples are operated on to capture both spatial and frequency components of the input signal into the filter output. It is widely used tool in image processing for applications such as image compression. There are numerous wavelet transform most popularly the Haar transform, Daubechies transform, dual tree complex transform, etc. Its implementation involves multiplication operation that requires high hardware complexity. These wavelet transforms can also be implemented using a technique called distributed arithmetic (DA) which saves resource by using multiply and accumulate accelerators. But the generic DA architecture consumes large LUTs, to eliminate this disadvantage, this paper proposes a resource efficient and fast alternative architecture designed using Xilinx Vivado for implementation of Daubechies low pass filter on FPGA.

Keywords Discrete wavelet transform (DWT) · Daubechies filter · Distributed arithmetic (DA) · Lookup table (LUT) · Resource utilization · Hardware accelerator

1 Introduction

Implementation of digital signal processing algorithm on FPGA and their use as hardware accelerators is becoming the growing research area. Though these algorithms are easily implemented using softwares like MATLAB and Python, they are realized with highly expensive hardware and delay. Therefore, by implementing

K. B. Sowmya · D. Jamakhandi (✉)
Department of Electronics and Communication Engineering, R V College of Engineering, Bengaluru, Karnataka, India
e-mail: divyaaj.ec16@rvce.edu.in

J. Alex Mathew
Department of Electronics and Communication Engineering, Srinivas Institute of Technology, Mangaluru, Karnataka, India

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_49

them on FPGA, the resource utilization is reduced to a greater extent. One such digital signal processing algorithm is discrete wavelet transform (DWT) which is used extensively in image processing for applications such as image edge detection, image de-noising, pattern recognition and image compression. Due to its excellent correlation property, it is used in the JPEG image formatting. Of the many wavelet transforms, Daubechies wavelets are a family of orthogonal wavelets that define the DWT which are characterized by a high number of vanishing points for a given input. Every wavelet type under the Daubechies filter has a unique scaling function that generates orthogonal multi-resolution analysis. The architecture being implemented in this paper is that of a distributed architecture in which the high resource utilization operation of multiplication between the inputs and coefficients is replaced by bitwise multiply and accumulate architecture that uses very less hardware resources and also computes output with high throughput.

1.1 DWT

A wavelet is a mathematical tool to get important information notes from various kinds of data like audio, image, video, etc. Comparing with Cosine transform, wavelet transform provides excellent coding gain for applications in image processing [7]. The algorithm approached by distributed arithmetic (DA) is to eliminate the multipliers which consume large number of logic elements and LUTs and substitute it with LUTs alone or simple logic blocks.

1.2 Daubechies Wavelet Transform

Daubechies wavelet transform has problem with image edges as the computation requires inputs beyond the end of the image matrix. The smaller tap filters like Daub4 (Daubechies-4tap) and Daub8 (Daubechies-8tap) can be used to eliminate this edge problem since out of boundary requirements will be reduced. The expression of the scaling elements of Daub4 and Daub6 filters are shown in Eqs. (1) and (2), respectively.

$$\begin{aligned} h_0 &= \frac{1+\sqrt{3}}{4\sqrt{2}}, h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \\ h_2 &= \frac{3-\sqrt{3}}{4\sqrt{2}}, h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}} \\ g_0 &= h_3, g_1 = -h_2, g_2 = h_1, g_3 = -h_0 \end{aligned} \tag{1}$$

$$
\begin{aligned}
h_0 &= \frac{1+z_1+z_2}{16\sqrt{2}}, h_1 = \frac{5+z_1+3z_2}{16\sqrt{2}}, \\
h_2 &= \frac{10-2z_1+2z_2}{16\sqrt{2}}, h_3 = \frac{10-2z_1-2z_2}{16\sqrt{2}}, \\
h_4 &= \frac{5+z_1-3z_2}{16\sqrt{2}}, h_5 = \frac{1+z_1+z_2}{16\sqrt{2}} \\
g_0 &= h_5, g_1 = -h_4, g_2 = h_3, \\
g_3 &= -h_2, g_4 = h_1, g_5 = -h_0
\end{aligned} \tag{2}
$$

But the Daub4 and Daub6 filters require a long filter category. Therefore, this requires large depth LUTs when implementing generic distributed arithmetic architecture. This increases resource utilization and reduces hardware performance. To eliminate this disadvantage, this paper proposes an alternative to the use of LUTs using minimum logic elements.

2 Literature Survey

There have been various literatures on implementation of DWT and DA architectures on hardware using number of methods. One such architecture is presented by S. S. Divakara and team in the paper "High Speed Area Optimized Hybrid DA Architecture for 2D-DTCWT" [1] which presents a DA architecture where combinations of MUX are used to calculate the multiplication of coefficient and input bit. Further these outputs are used for CWT application. The disadvantage of this architecture is that mux uses redundant AND gates and an internal OR gate which can be reduced.

In the paper "Modified Distributive Arithmetic based 2D-DWT for Hybrid Image Compression", Mohan and Satyanarayana [2] also implement the DA architecture based DWT for image compression. To increase image compression rate, they are further using neural network. But this could be very heavy for the hardware when implemented on FPGA. To further increase compression rate using simplified technique, more layers of DWT could be added, or other compression methods like DCM and Huffman coding can be cascaded with our architecture.

The literature "High-performance FPGA Implementation of Discrete Wavelet Transform for Image Processing" by Huang and team [3] implements generic DA architecture with a serial to parallel converter do increase throughput. Both low pass and high pass DWT are computed, and results multiplexed. While this architecture has good throughput, the latency would be high due to serial to parallel conversion and also high usage of LUT. Fig. 1 shows the architecture presented by this paper.

Jagdamb Behari Srivastava, R. K. pandey and Jitendra Jain in their paper "Efficient Multiplier-less Design for 1-D DWT Using 9/7 Filter Based on NEDA Scheme" [4] propose an architecture for 1D-DWT which eliminated the use of multiplier by expanding the coefficients into binary bits which will be ANDed to give the addition of only selective inputs thus using only adders. The usage of this is cumbersome to

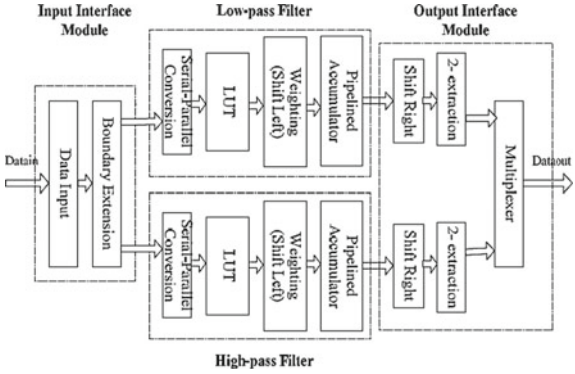


Fig. 1 Generic DA architecture with serial–parallel conversion for both with low pass and high pass DWT computation [3]

be implemented in 2D, also the coefficients and addends need to be pre-computed, and therefore, it cannot be used in a generalized manner.

In the paper "Result-Biased Distributed-Arithmetic-Based Filter Architecture for Approximately Computing the DWT" [5], Maurizio Martina and team proposes a result-biased method in which the probability of arrival of 1 is calculated which if near to absolute will be pre-fixed to one and the calculation will be done. The proposed design has been stated to reduce the hardware by around 20%. IT also implements butterfly adders to reduce the cascading which cause delay. But as this is based on probability, the accuracy would not be very high as also stated by the author. The architecture of butterfly adder is shown in Fig. 2.

The literature "Distributed Arithmetic Architecture of Discrete Wavelet Transform (DWT) with Hybrid Method" by Jafar and team [6] proposes a 3D-DWT for medical image processing where they cascade three N-point 1D filters and two transpose memories for a 3D volume. They have synthesized these results for Spartan-3

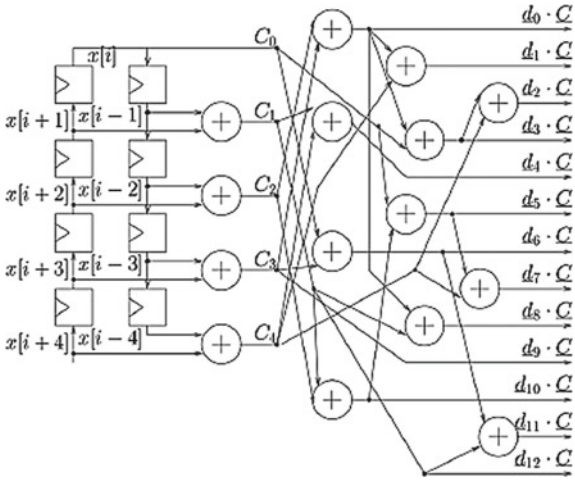


Fig. 2 Butterfly adder architecture design for higher operand numbers [5]

FPGA. They have implemented both Haar and Daubechies filters and given a comparative study between the two and also hardware requirements, and output image quality analysis has been done.

3 Mathematical Analysis

Consider the coefficients A_k from $k = 1$ to K for a K-coefficient filter and K consecutive inputs X_k each of N bits. Each value of output is equal to signed multiplication of the inputs with respective coefficients and the summation of the multiplication results as shown in (3).

$$Y = \sum_{k=1}^{K} A_k X_k \tag{3}$$

Each input X_k can be represented in terms of their bits as shown in Eq. (4) where b_{k0} is the MSB, and each bit is summed based on their binary place value.

$$X_k = -b_{k0} + \sum_{n=1}^{N-1} b_{kn} 2^{-n} \tag{4}$$

By substituting the Eq. (4) in (3), we get below Eq. (5).

$$Y = \sum_{k=1}^{K} A_k \left[-b_{k0} + \sum_{n=1}^{N-1} b_{kn} 2^{-n} \right] \tag{5}$$

By expanding the above Eq. (5), we get Eq. (6) which is the implementation of distributed arithmetic.

$$Y = -\sum_{k=1}^{K} (b_{k0} A_k) + \sum_{k=1}^{K} \sum_{n=1}^{N-1} (A_k b_{kn}) 2^{-n} \tag{6}$$

Here, for every bit of each input, the output of $A_k b_{kn}$ is calculated and right shifted starting from LSB, and finally, the result of $b_{k0} A_k$ which is of the MSB is subtracted from the resulted. This architecture uses simple AND array and MAC unit as hardware resources.

In this paper, Daubechies low pass filter is implemented which has nine coefficients, and each input is 8 bits. The Eq. (6) can be further represented as in (7).

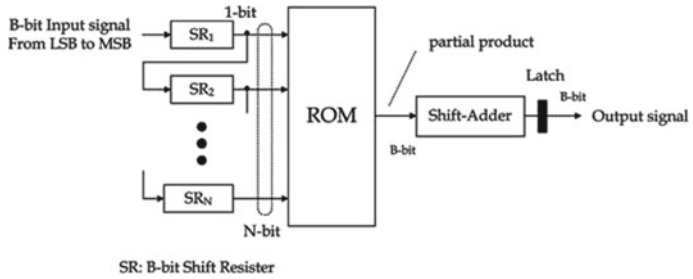


Fig. 3 Generic DA architecture

$$Y = -\sum_{k=1}^{9}(b_{k0}A_k) + \sum_{k=1}^{9}\sum_{n=1}^{7}(A_k b_{kn})2^{-n} \tag{7}$$

4 Implementation

4.1 Generic DA Architecture

A generic DA architecture is the one in which the Eq. (4) is implemented. Usually, generic DA substitutes LUTs for logical elements required by heavy multiplications. It calculates all the possible sums for the second term in Eq. (4) for input bits 0 and 1 and stores them in lookup table (ROM). Based on the input bits, a corresponding sum is selected and sent to the external MAC unit for further calculation and multiplication with 2^{-1} in each cycle (right shift does this operation) [8]. Also, the sum of products of MSB with all the coefficients is calculated separately and then added to the previous result. The generic distributed arithmetic architecture is shown in Fig. 3 [9].

4.2 Modified DA Architecture

In the generic DA architecture, the Usage of LUT is very high, for example, if there are nine coefficients as in Daubechies low pass filter, the depth of the LUT required will be 2^9, i.e., 512 each of large width. Therefore, the architecture implemented in this paper is modified DA architecture in which the LUT usage is also reduced by substituting the LUT block by an arithmetic block which instantaneously calculated the output value using an AND Array and Adder instead of retrieving values from the LUT. Thereby using a minimum number of logic elements, even the excessive LUT usage is reduced by this architecture. The architecture is shown in Fig. 4.

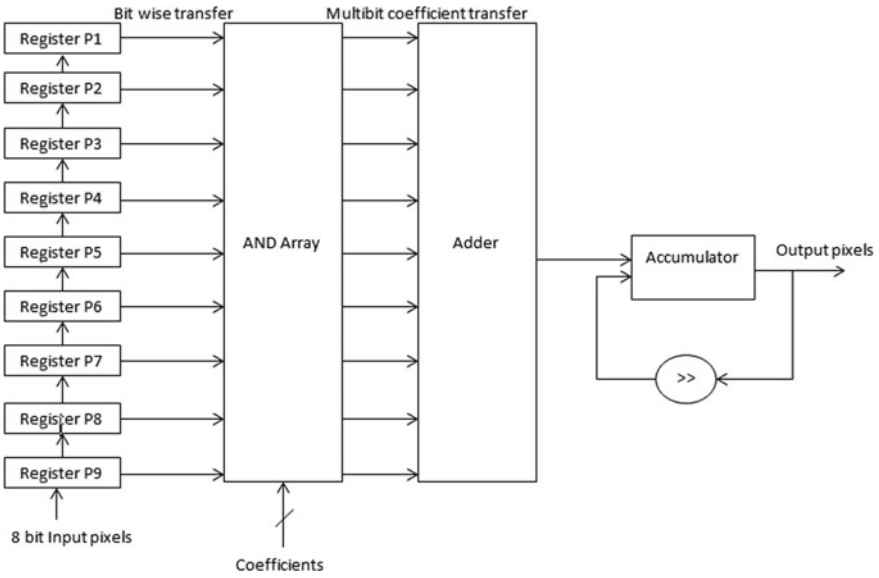


Fig. 4 Proposed architecture

4.3 Operation

Using the above architecture, the Daubechies 9/7 wavelet filter in which the low pass filter has nine coefficients is shown in Table 1.

Each of the above coefficients are scaled by a factor of 10^6 and given as input to the AND arrays. The other input to the AND gates is the bits of the nine consecutive inputs fed into the registers, respectively, to each block.

In every clock cycle, one pixel is taken as input into the module which is stored in the input shift register with all its bits entering parallel and every next cycle the bits are parallelly pushed into next shift register and new pixel enters the input register . Using a repeated architecture of AND array, the LSB of each of the nine shift registers are ANDed with every bit of the coefficient, and the results are added in the same cycle after shifting right after each bit operation using the MAC unit. Only the result of the MSB is subtracted to get the result. But since for the initial eight cycles

Table 1 Daubechies low pass filter coefficients

N	Low pass filter ($h[n]$)
±4	+0.026748
±3	−0.0168641
±2	−0.0782232
±1	+0.2668642
0	+0.6029490

Fig. 5 Left: input image, right: Daubechies low pass filter output

some of these registers will be empty, the output will be valid from the 9th cycle. Therefore, every cycle one output pixel value is produced, thus having a throughput of 1 pixel/cycle with a latency of eight cycles initially.

The nine operand adder can be realized using sequential cascade of adders or pipelined adders based on the timing requirement. Since the delay is within the tolerable limit and not affecting any skew, this paper uses a cascade of adders. Also, in the similar manner, high pass filter can be designed which has seven coefficients, needing 7 AND array blocks. It will have a latency of only six clock cycles with throughput remaining the same.

5 Results

5.1 DWT Output

The input to the module is a 64 × 64 image with 4096 pixels. The complete output is reconstructed after 4096 cycles with initial latency of eight cycles. The input and output of the design are shown in Fig. 5.

5.2 Resource Utilization

The graph given in Fig. 6 shows the summary of the resource utilization which are 3, 14, 1, 1% of BUFG, IO, flip-flops and LUTs which is quite minimal resource usage compared to other architectures.

The detailed usage of resources like slice logic elements, IO and clocking elements is given by Tables 2, 3 and 4.

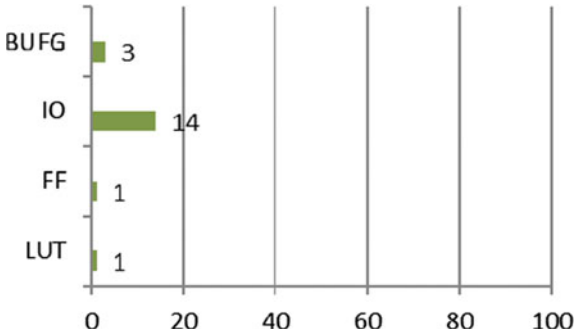


Fig. 6 Resource utilization in percentage on FPGA (zed board)

Table 2 Detailed usage of slice logic

Site type	Used	Fixed	Available	Util%
Slice LUTs*	281	0	53,200	0.53
LUT as logic	281	0	53,200	0.53
LUT as memory	0	0	17,400	0.00
Slice registers	15	0	106,400	0.01
Register as flip-flop	15	0	106,400	0.01
Register as latch	0	0	106,400	0.00
F7 muxes	0	0	26,600	0.00
F8 muxes	0	0	13,300	0.00

Table 3 IO and GT specifications

Site type	Used	Fixed	Available	Util%
Bonded IOB	28	0	200	14.00
Bonded IPADs	0	0	2	0.00
Bonded IOPADs	0	0	130	0.00
PHY_CONTROL	0	0	4	0.00
PHASER_REF	0	0	4	0.00
OUT_FIFO	0	0	16	0.00
IN_FIFO	0	0	16	0.00
IDELAYCTRL	0	0	4	0.00
IBUFDS	0	0	192	0.00
PHASER_OUT/PHASER_OUT_PHY	0	0	16	0.00
PHASER_IN/PHASER_IN_PHY	0	0	16	0.00
IDELAYE2_IDELAYE2_FINEDELAY	0	0	200	0.00
ILOGIC	0	0	200	0.00
OLOGIC	0	0	200	0.00

Table 4 Clocking utilization

Site type	Used	Fixed	Available	Util%
BUFGCTRL	1	0	32	3.13
BUFID	0	0	16	0.00
MMCME2_ADV	0	0	4	0.00
PLLE2_ADV	0	0	4	0.00
BUFMRCE	0	0	8	0.00
BUFHCE	0	0	72	0.00
BUFR	0	0	16	0.00

Table 5 Result analysis

Parameters	Slice logic elements	Flip-flops
DWT-DA architecture [7]	315	280
DWT hybrid architecture [1]	382	22
Developed architecture	281	15

5.3 Power Consumption

The total on-chip power consumption is 0.108 W, and the junction temperature is 26.2 °C which is much less than the thermal margin of 59.9 °C. Also, the entire design works without any timing errors.

5.4 Comparative Analysis

Table 5 gives the comparison of resource utilization mainly the use of slice logic elements and flip-flops of the architectures developed in papers [1, 7]. It can be seen that there is an improvement in resource utilization in this new developed architecture.

6 Conclusion

Therefore, the design proposed in this paper is a resource efficient, power efficient and fast architecture for a discrete wavelet transform. It adopts basic principles of distributed arithmetic alongside eliminating its disadvantage of high LUT usage using minimum logic elements.

This module can be packaged into IP for reusability and used as hardware accelerator in larger modules where DWT is required for compression, feature extraction or further processing.

The design can also be used for other wavelet filters such as Haar, Symlet and Orthogonal with different coefficients and little modification in number of AND array blocks.

References

1. Divakara SS, Patilkulkarni S, Prasanna Raj C (2018) High speed area optimized hybrid da architecture for 2d-dtcwt. Int J Image Graph 18(1): 1850004
2. Mohan M, Satyanarayana SP (2014) Modified distributive arithmetic based 2d-dwt for hybrid (neural network-dwt) image compression. Glob J Comput Sci Technol F Graph Vision 14(2) Version 1.0
3. Huang Q, Wang Y, Chang S (2011) High-performance FPGA implementation of discrete wavelet transform for image processing. 978-1-4244-6554-5/11/$26.00©2011. IEEE
4. Behari Srivastava J, Pandey RK, Jain J (2013) Efficient multiplier-less design for 1-d dwt using 9/7 filter based on neda scheme. Int J Innov Res Comput Commun Eng 1(4)
5. Martina M, Masera G, Roch MR, Piccinini G (2015) Result-biased distributed-arithmetic-based filter architectures for approximately computing the DWT. IEEE Trans Circ Syst I Regul Pap 62(8)
6. Ja'afar NH, Ahmad A, Amira A (2013) Distributed arithmetic architecture of discrete wavelet transform (dwt) with hybrid method. 978-1-4799-2452-3/13/$31.00©2013. IEEE
7. Manikandababu CS, Munira NJR Modified distributive arithmetic algorithm based 3d dwt processor with parallelism operation of 1d-dwt. Int J Adv Eng Technol. E-ISSN 0976-3945
8. Velukar SS, Parlewar MP (2014) FPGA implementation of fir filter using distributed arithmetic architecture for DWT. Int J Comput Appl 92(16): 0975–8887
9. Thirumala Selva C, Sudhakar R (2016) An efficient 2d dwt-a distributed arithmetic with rapid arithmetic coder for medical image compression. Asian J Inf Technol 15(14): 2371–2382

Propagation Characteristics of SIW and Waveguide: A Comparison

Sheelu Kumari, Vibha Rani Gupta, and Shweta Srivastava

Abstract Substrate-integrated circuit (SIC) technology has provided the most successful solutions for microwave and millimeter wave technology and expected to be the only technology useful in terahertz frequency range. Substrate-integrated waveguide (SIW) has proved itself the most used guided-wave structure in this technology, but includes one additional loss called leakage loss along with the other two losses found in conventional waveguides, i.e., dielectric loss and conduction loss. In this paper, different losses found in SIWs and conventional waveguides operating in different frequency bands from C band to K band are calculated and compared. It is found that the total losses in SIWs are comparable to that for the waveguides. Propagation characteristics of both the structures are found to be almost same. HFSS software is used to simulate the designed structures.

Index Terms Conduction loss · Dielectric loss · Leakage loss · Propagation characteristic · Substrate-integrated waveguide · Waveguide

1 Introduction

Microwave-integrated circuits have developed through different generations [1] with betterment in circuit size, density of integration, and low-cost techniques for fabrication. The first generation of the microwave circuits includes waveguides and coaxial lines, which can handle high-power and high-Q structures but are voluminous and bulky. The second generation replaced coaxial cables by printed transmission lines like microstrips lines, striplines, and externally added or surface-mounted active

S. Kumari (✉) · V. R. Gupta
Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India
e-mail: sheelusingh80@gmail.com

S. Srivastava
Department of Electronics & Communication Engineering, Jaypee Institute of Information Technology, Sector 128, Noida, India

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_50

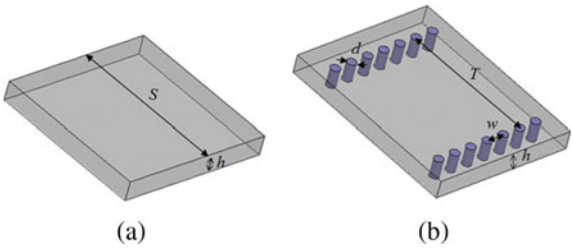


Fig. 1 Configuration of **a** waveguide, **b** SIW

devices constitute planar microwave-integrated circuits (MICs) and systems. These circuits and systems suffer from severe radiation losses especially at high frequencies and are found inappropriate for high-Q value structures like filters. Further integration and multilayered ceramic processing technique lead to the development of miniature hybrid microwave-integrated circuit (MHMIC), the third generation of the circuits. In these circuits, passive components like MIM resistors and capacitors are constituted and integrated into planar layers during the fabrication process. Next comes the fourth generation, in which both active and passive elements are integrated onto the same substrate.

SICs with special synthesized guided-wave structures and planar circuits like microstrip and CPW on the same substrate solved the whole problem of obtaining benefits of different generation of circuits and excluding their demerits. Different types of wave-guiding structures in SIC technology [2] are substrate-integrated slab waveguide (SISW), substrate-integrated waveguide (SIW), substrate-integrated image dielectric guide (SIIDG), substrate-integrated non-radiating dielectric (SINRD) guide, substrate-integrated insular guide (SIIG), and substrate-integrated inset dielectric guide (SIINDG). The most developed and popular SIC platform is SIW technology [3–16].

Waveguides, loaded with dielectrics, consist of dielectric with upper and lower metallic layers constituting upper and lower walls. Two metal layers on sides form two side walls. Figure 1a shows configuration of a waveguide with width S and height h. However, SIWs consist of dielectric with upper and lower metallic layers acting as upper and lower walls, and two side walls are replaced by two rows of vias. Figure 1b shows configuration of a SIW. Distance between the two rows of vias constituting two sidewalls is T. Distance between any two consecutive vias in either of the via-rows is pitch w. Diameter of each via is d. Height of the structure is h.

Conventional waveguides have only two types of losses: dielectric loss and conduction loss. But due to the vias, SIWs exhibit leakage loss, along with the above-mentioned two losses. SIWs can be practically used only if the value of leakage loss is very small. In this paper, different losses in waveguides and their equivalent SIWs operating in different frequency bands from C band to K band are calculated and compared.

2 Design Details

Dimensions of waveguides operating in different frequency bands from C band to K band with cut-off frequency f_c are calculated. Design rules mentioned in [17, 18] are followed to avoid band-gaps and leakage loss in operating frequency band. Pitch w is kept twice the via diameter d. Dimensions of the equivalent SIWs are calculated using design Eq. (1) mentioned in [18]. Substrate Duroid 5870 with $\varepsilon_r = 2.33$ and height 0.787 mm is used for structure simulation in HFSS software. Table 1 shows the dimensions of waveguides and their equivalent SIWs operating in different frequency bands.

$$S = T - \frac{d^2}{0.95w} \tag{1}$$

Waveguides have two types of losses [19]: dielectric loss (α_d) and conductor loss (α_c), which can be calculated using Eqs. (2) and (3), respectively.

$$\alpha_d = ((k^2 \tan\delta))/2\beta \tag{2}$$

$$\alpha_c = R_s \frac{(2h\pi^2 + S^3k^2)}{S^3h\beta k\eta} \tag{3}$$

$$\beta = \sqrt{k^2 - k_c^2} \tag{4}$$

$$k = 2\pi f\sqrt{\mu\varepsilon} \tag{5}$$

$$k_c = 2\pi f_c\sqrt{\mu\varepsilon} \tag{6}$$

$$\eta = \sqrt{\mu/\varepsilon} \tag{7}$$

$$R_s = \sqrt{\omega\mu_0/2\sigma} \tag{8}$$

where

Table 1 Dimensions of the waveguides and SIWs

Freq. band	f_c(GHz)	S (mm)	T (mm)	d (mm)	w (mm)
C	3	32.76	33.28	1	2
X	7	14.04	14.56	0.5	1
Ku	11	8.80	9.10	0.5	1
K	16	6.14	6.29	0.3	0.6

S and h are width and height of waveguide, respectively.

k is free space wave number.

β is phase constant.

$\tan\delta = 0.0012$ is the dielectric loss tangent.

μ and ε are absolute permeability and permittivity, respectively.

η is wave impedance in the dielectric.

$\sigma = 5 \times 10^7$ S/m is conductivity of metal.

R_s is surface resistivity of the conductors.

f_c is cut-off frequency, and f is operating frequency.

Total loss α_t, in a waveguide, is therefore given by

$$\alpha_t = \alpha_d + \alpha_c \tag{9}$$

SIWs have one addition loss: leakage loss (α_l) [20] along with the two other losses found in waveguides, i.e., dielectric loss (α_d) and conductor loss(α_c). Equations (10), (11), and (13) give values of α_d, α_c, and α_l, respectively, for SIWs.

$$\alpha_d = (k^2 \tan\delta)/2k_z \tag{10}$$

$$\alpha_c = \frac{R_s}{T_e \eta \sqrt{1 - \frac{k_c^2}{k^2}}} \left[\frac{T_e}{h} + \frac{2k_c^2}{k^2} \right] \tag{11}$$

$$k_z(f) = \sqrt{\left\{ k^2 - \left[\frac{2}{T_e} \cot^{-1}\left(\frac{f_c}{f} r_s (1 - j) \right) \right]^2 \right\}} \tag{12}$$

$$\alpha_l = \{(1/T)(d/T)^{2.84}[(w/d) - 1]^{6.28}\}/\{4.85[(2T/\lambda)^2 - 1]^{0.5}\} \tag{13}$$

where k_z is complex propagation constant, T_e is the effective value of T [17], $r_s$is real part of the surface wave impedance [17], and λ is wavelength in dielectric. All other designations are same as used for waveguides.

Total loss α_t in SIWs is therefore given by

$$\alpha_t = \alpha_d + \alpha_c + \alpha_l \tag{14}$$

3 Results

S parameter response for the designed C band waveguide and its equivalent SIW is shown in Fig. 2. It is observed that both the waveguide and SIW start transmission of waves from $f_c = 3.25$ GHz. S_{21} response is almost same for both the guides over operation bandwidth. Values of S_{11} parameter for the SIW are higher compared to S_{11} parameter for the waveguide, but are fairly good.

Comparison of different types of losses for C band waveguide and its equivalent SIW is shown in Figs. 3, 4, and 5. It is observed that the leakage loss in SIW is negligible as compared to other losses.

Figures 6, 7, 8, and 9 show comparison of total losses in the waveguides and their respective equivalent SIWs operating in C, X, Ku, and K band. They also show the phase constant β for both the guides. Graphs showing total loss constitute the real part of complex propagation constant for different waveguides and their respective SIWs. Equation (4) can be used to calculate the phase constant for both waveguide and its equivalent SIW, and therefore, imaginary part of complex constant is same for both.

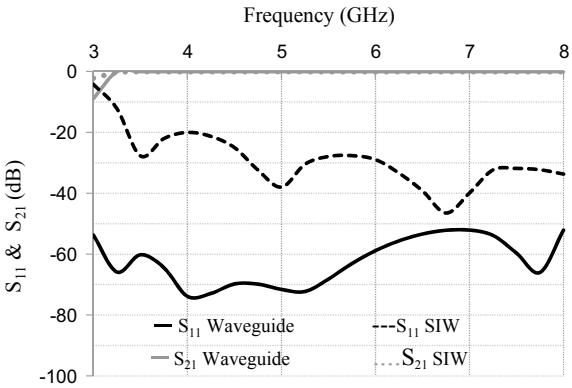


Fig. 2 S parameter response for the C band waveguide and SIW

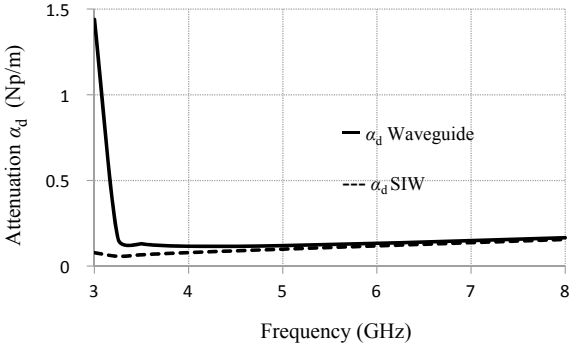


Fig. 3 Attenuation α_d for the C band waveguide and SIW

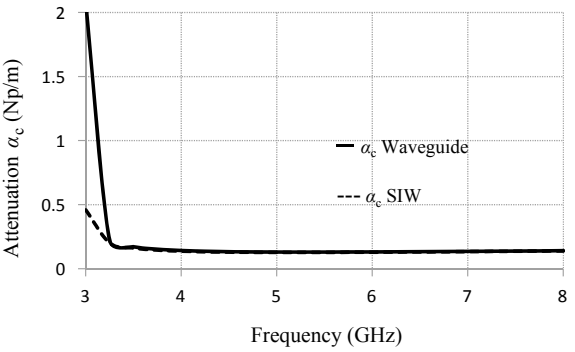


Fig. 4 Attenuation α_c for the C band waveguide and SIW

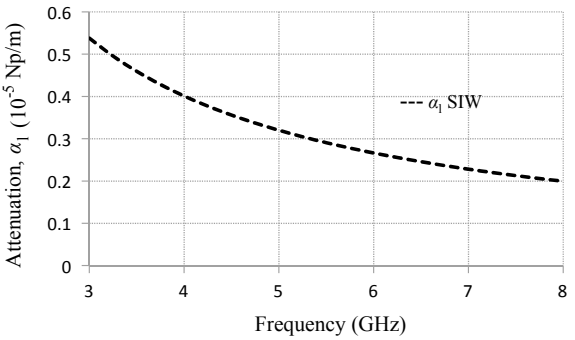


Fig. 5 Attenuation α_l for the C band SIW

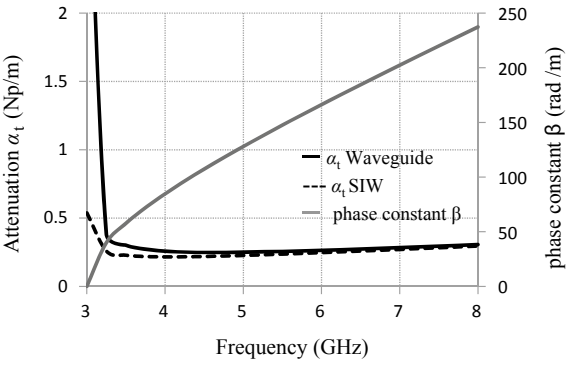


Fig. 6 Total attenuation α_t and phase constant β for the C band waveguide and SIW

The total loss increases for both types of the guides operating in higher frequency bands. For C band total loss is almost equal to that for the waveguide, but for higher frequency bands, total loss in SIW is lower than that for the waveguide.

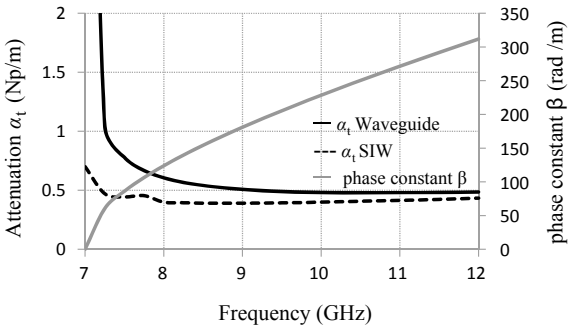


Fig. 7 Total attenuation α_t and phase constant β for the X band waveguide and SIW

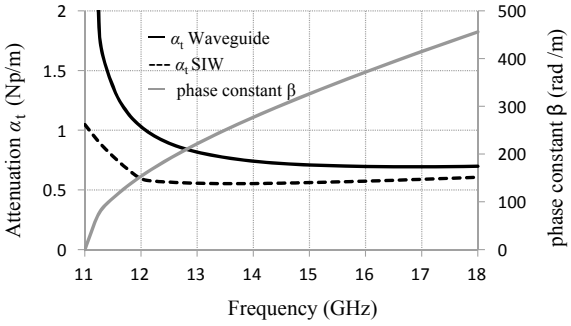


Fig. 8 Total attenuation α_t and phase constant β for the Ku band waveguide and SIW

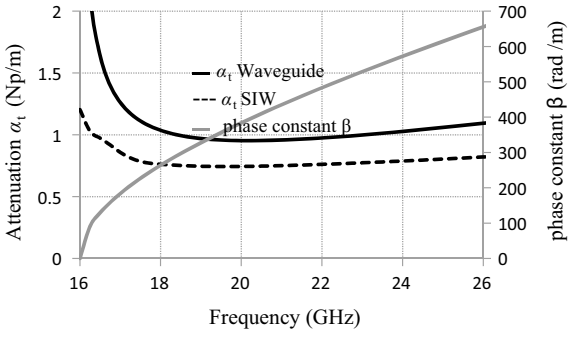


Fig. 9 Total attenuation α_t and phase constant β for the K band waveguide and SIW

4 Comparison with Previous Work

Calculated values of total loss for the designed SIW (Fig. 8) are compared with the extracted values of total SIW loss [17, Fig. 12] in Table 2. Frequency range common in [17, Fig. 12] and the present work (Fig. 8) is only presented for the comparison. Low loss of the designed SIW is evident from Table 2.

Table 2 Comparison of total loss for the SIW (Fig. 8) and the referred SIW [17, Fig. 12]

Frequency (GHz)	α_t (Calculated) [17, Fig. 12]	α_t (Measured) [17, Fig. 12]	α_t Present work
15	2.15	2.25	0.56
15.5	1.7	1.7	0.56
16	1.5	1.5	0.57
16.5	1.2	1.1	0.58
17	1.2	1.2	0.58
17.5	1	1	0.59
18	0.95	0.85	0.6

Unit of total loss α_t is Np/m

5 Conclusion

Leakage loss is negligible in comparison with the other two losses found in SIWs, and therefore, the main constituents of total loss are dielectric loss and conduction loss. Total loss in SIWs is comparatively less than their respective waveguides at higher frequency bands and hence is a better candidate for upper frequency band in microwave and millimeter technology.

References

1. Wu K (2010) Towards the development of terahertz substrate integrated circuit technology. SiRF, pp 116–119
2. Wu K, DeslandesD, Cassivi Y (2003) Substrate integrated circuits—a new concept for high frequency electronics & optoelectronics. In: Proceedings of 6th international telecommunication modern satellite cable, broadcast services Conference, pp 3–10
3. Bozzi M, Georgiadis A, Wu K (2011) Review of substrate integrated waveguide circuits and antennas. IET Microwaves Antennas Propag 5(8):909–920
4. Bozzi M, Perregrini L, Wu K, Arcioni P (2009) Current and future research trends in substrate integrated waveguide technology. Radio Eng 18(2):201–209
5. Song K, Lu W, Guo S, Fan Y (2019) Four-way hybrid SIW/microstrip-line power divider with improved output isolation. Electron Lett 55(1): 36–38
6. Huang L, Zhang S (2018) Ultra-wideband ridged half-mode folded substrate- integrated waveguide filters. IEEE Micro Wirel Comp Lett 28(7):579–581
7. Chen RS, Wong S-W, Zhu L (2015) Wideband bandpass filter using U-slotted substrate integrated waveguide (SIW) cavities: IEEE Microwave Wirel Comp Lett 25(1)
8. Dong K, Mo J, He Y, Ma Z, Yang X (2016) Design of a millimetre-wave wideband bandpass filter with novel slotted substrate integrated waveguide. Microwave Opt Technol Lett 58(10):2406–2410
9. Hosseini-Fahraji A, Mohammadpour-Aghdam K, Faraji-Dana R (2016) Design of wideband millimeter-wave bandpass filter using substrate integrated waveguide. In: IEEE Iranian Conference on Electrical Engineering (ICEE), Shiraz, Iran, May 2016
10. Liu C, An X (2017) A SIW-DGS wideband bandpass filter with a sharp roll-off at upper stopband. Microw Opt Technol Lett 59(4):789–792

11. Wu Y-D, Li G-H, Yang W, Yang X (2017) Design of compact wideband QMSIW band-pass filter with improved stopband. Progr Electr Res Lett 65:75–79
12. Hu H-T, Chen F-C, Chu Q-X (2016) A wideband U-shaped slot antenna and its application in MIMO terminals. IEEE Antennas Wirel Propag Lett 15:508–511
13. Rhbanou A, Sabbane M, Bri S (2015) Design of K-band substrate integrated waveguide bandpass filter with high rejection. J Microw Optoelectr Electromag Appl 14(2):155–169
14. Xu S, Ma K, Meng, Meng F, Yeo KS (2015) Novel defected ground structure and two-side loading scheme for miniaturized dual-band SIW bandpass filter designs. IEEE Microw Wirel Compon Lett 25(4):217–219
15. Li M, Jiang S, Chen C, Chen W, Zhang H (2018) Miniaturized dual-band filter using triple folded substrate-integrated waveguide resonators. Microw Opt Technol Lett 60:2038–2043
16. Kordiboroujeni Z, Bornemann J (2014) New wideband transition from microstrip line to substrate integrated waveguide. IEEE MIT 62(12):2983–2989
17. Deslandes D, Wu K (2006) Accurate modeling, wave mechanism & design considerations of substrate integrated waveguide. IEEE Trans MIT 54:2516–2526
18. Xu F, Wu K (2005) Guided wave and leakage characteristics of substrate integrated waveguide. IEEE transactions on microwave theory and techniques, vol 53, no 1
19. Pozar DM (2009) Microwave engineering, 3rd edn. Wiley, India, pp 106–112
20. Pasian M, Bozzi M, Perregrini L (2014) A Formula for Radiation Loss in Substrate Integrated Waveguide. IEEE Trans Microw Theory Tech 62(10):2205–2213

Predicting Software Cost Through Entity–Relationship Diagrams: An Empirical View

Mansi Gupta, Kumar Rajnish, and Vandana Bhattarcharjee

Abstract In order to effectively monitor and handle the entire software development process, software cost estimation plays an important role. The most popular tool for software cost estimation nowadays is Constructive Cost Model II (COCOMO II). This tool uses lines of code (LOC) and function points (FPs) to calculate software size. However, the accurate prediction of software development costs become critical as well as difficult as these LOC & FP are actual implementation details which cannot be gathered in early stages. Thus, ER model—a concept-based model which is constructed during requirement phase (early phase) of the SDLC—is explored for the software cost estimation. In this paper, an attempt has been made to propose a multiple regression model for software cost estimation based on number of entities, attributes, relationship, and path complexity of an ER diagram. The MATLAB tool has been used for analysis of data.

Keywords ER diagram · Entity · Relationship · Attributes · Software cost estimation

1 Introduction

It is well known and documented that the software industry suffers from regular overruns of costs, and the calculation of software costs remains a problem. So, software cost estimation [1] plays a vital role if done in the initial stages of the software development process. As it can help in effectively monitoring and handling the entire software development process, many contributors have made efforts over the past four decades to develop parametric and few non-parametric modeling techniques.

M. Gupta (✉) · K. Rajnish · V. Bhattarcharjee
Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India
e-mail: jv.mansi@gmail.com

K. Rajnish
e-mail: krajnish@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_51

For instance, Farr and Zagorski model [2], Aron model, Walston and Felix model, Wolverton model, Kustanowitz model,Putnam SLIM, GRC model, Doty model, Jensen model, Bailey and Basili Model [3], Nikolaos Mittas [4], Emilia Mendes [5], Sweta Kumari [6] and [7–9] have been proposed.

Today, the most popular cost estimation model is the Constructive Cost Model (COCOMO II) [10]. It is the latest major extension released in 1981 to the original COCOMO [11, 12] template. It is a framework for estimating the cost, effort, and schedule of planning a new software development project. It works in three phases: prototyping using application composition model capabilities, early design estimation, post-architecture estimation. This tool uses lines of code (LOC) and function points (FPs) to calculate software size. However, at the early stage of software development, lines of code cannot be accessed or calculated. Ultimately, the function points (FPs) are the outcome of design phase of SDLC. As a result, it leads to many software cost-related problems [13–15]. This research proposal suggests the effectivity of using ER diagram, for software cost estimation. ER diagrams are a valuable graphical method for describing the ER model. ER model is a conceptual modeling approach constructed in requirement analysis phase. The arrangement of this paper is as follows: Sect. 2 shows the basic concept regarding ER model and the factors affecting software cost. Section 3 illustrates the proposed cost estimation model. Section 4 concludes the new estimating model for software cost and discusses future work.

2 Background

The ER model is a database design or prototype that can be used later to be implemented as a database. ER model's main components are entity set and relationship set. Most of the practices of design and development of software are focused on the concept of ER model. Consequently, the ER model tends to have the most readily available data from the requirements capture and analysis phase of SDLC [16]. Although, we consider that the ER diagram is a non-directional graph after examining the ER diagram. However, it becomes simple to map an ER diagram to a directional graph; for this, the entity is considered as the vertex in a graph and the relationship is being considered as the edge in a graph.

3 Proposed Model

3.1 Description of Proposed Model

This section proposes a model to estimate software cost. For the purpose of cost estimation, the following metric is being used:

- **NOE**: the total count of entities that appear in entity–relationship diagram.
- **NOR**: the total count of relationships that appear in entity–relationship diagram.
- **NOA**: the total count of attributes that appear in entity–relationship diagram
- **NOP**: the count of path complexity of an entity–relationship diagram

Path complexity of an ER Diagram is the sum total of all the paths taken by one entity to access the other entities (when each path has some length. Graph theoretic concepts shall be used to estimate this path complexity.

The proposed predictive cost regression model is named as regression model for entity–relationship diagram (RMERD), which is defined as follows:

$$\text{RMERD} = a + b * \text{NOE} + c * \text{NOR} + d * \text{NOA} + e * \text{NOP}$$

where values for the constant $a = 2.974$, and weight $b = 10.866$, $c = -3.899$, $d = -0.005$, $e = -0.499$.

3.2 Data Collection

The twelve ER diagram datasets (refer Table 1) used here for the modeling and calculation of the cost estimation model are being referenced from the paper titled "Software Cost Estimation through Conceptual Requirement" [17]. All the data are actual industry data fetched from AKEMCO Technology Pte Ltd, IPACS e-Solutions (S) Pte Ltd and Singapore Computer Systems Ltd. These projects are across varying applications such as billing, management of freight, quotation and order processing.

Table 1 Actual data gathered from industry

NOE	NOR	NOA	NOP	Actual cost
6	5	112	9.33	48
6	5	75	8.33	38
21	22	512	63.17	81
3	2	86	2.67	29
3	2	86	2.67	34
25	33	656	118.03	92
4	3	66	5	33
8	7	212	19	70
14	14	126	31.43	80
16	17	441	33.93	85
64	69	1524	204.15	322
38	38	779	84.36	235

3.3 Results and Interpretation

For software cost estimation, regression-based technique, i.e., multiple linear regression model, is being used. The data were available for the parameters: NOE, NOR, NOA, and NOP. And man-day is used as a unit to measure software cost. The correlation coefficient (r) is used to calculate the magnitude of the relationships between these factors and the estimated cost. Higher values of r indicate high positive relationship. The correlation values for various metrics are given in Table 2.

All of these above parameters, according to correlation coefficient, have strong linear relationship with the software cost. The below graph shows the correlation between NOE, NOR, NOA, NOP and predicted software cost (Figs. 1, 2, 3 and 4).

The calculated multiple regression equation from the dataset is as follows:

RMERD : Regression model for ER Diagram

$$\text{RMERD} = 2.974 + 10.866\ \text{NOE} - 3.899\ \text{NOR} - 0.005\ \text{NOA} - 0.499\ \text{NOP}$$

Table 2 Correlation of predictive cost with NOE, NOR, NOA, NOP

	NOE	NOR	NOA	NOP
Predictive cost	0.982	0.956	0.949	0.893

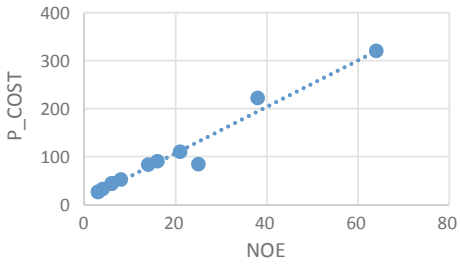


Fig. 1 Graph showing positive correlation between NOE and P_COST (predictive cost)

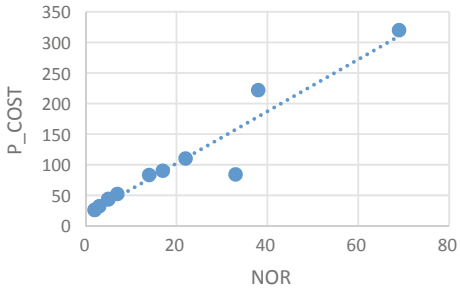


Fig. 2 Graph showing positive correlation between NOR and P_COST (predictive cost)

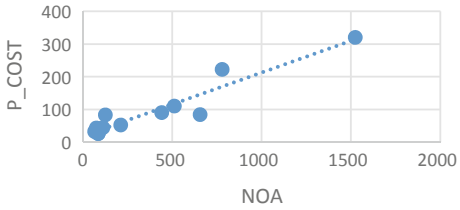


Fig. 3 Graph showing positive correlation between NOA and P_COST (predictive cost)

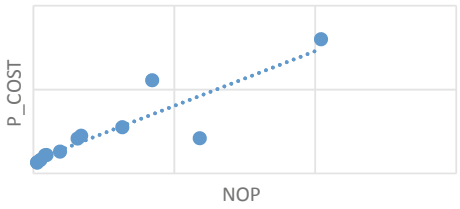


Fig. 4 Graph showing positive correlation between NOP and P_COST (predictive cost)

Table 3 Regression statistics for RMERD

Regression statistics	
Multiple *R*	0.9907
R square (R^2)	0.9814
Adjusted *R* square	0.9708
Observations	12

Further to determine the accuracy and goodness of this RMERD model, the following regression statistics were also calculated which is mentioned in Table 3.

R^2 statistics evaluate the strength of the multiple regression relationship. The value of R^2 obtained here is 0.9814 which shows that the RMERD model has high regression relationship. The comparison of actual cost and estimating cost is represented in Table 4 (Fig. 5).

4 Conclusion

Estimating software costs is an important process that cannot be neglected in software development. Proper measures must be taken to properly manage every task. As said by George Odiorne "*If you can't measure it, you can't manage it*." In this paper, an attempt has been made to propose a multiple regression model for software cost estimation based on number of entities, attributes, relationship, and path complexity of an ER diagram. The paper began with a concise overview of the cost estimate perception which used COCOMO II and the factors that affect it. Although COCOMO II is an enhanced approach, it is effective after certain stage of SDLC.

Table 4 Comparison between predictive cost and actual cost

NOE	NOR	NOA	NOP	Actual cost	Predictive cost
6	5	112	9.33	48	44
6	5	75	8.33	38	43
21	22	512	63.17	81	110
3	2	86	2.67	29	26
3	2	86	2.67	34	26
25	33	656	118.03	92	84
4	3	66	5	33	32
8	7	212	19	70	52
14	14	126	31.43	80	83
16	17	441	33.93	85	90
64	69	1524	204.15	322	320
38	38	779	84.36	235	222

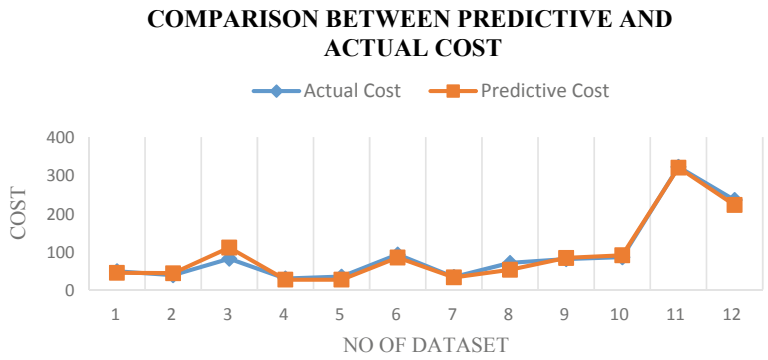


Fig. 5 Graph showing comparison between predictive cost and actual cost

But to gain the knowledge of software cost at the initial stages, a very simple model RMERD is proposed here. From Tables 3 and 4, the results show that the proposed RMERD leads to be good measure for software cost estimation. Although at some places of the result predictive cost is either overestimated or underestimated, it needs to be verified and is considered for future scope.

5 Future Work

This research work focuses on software cost estimation based on. ER diagram parameters for mid and small projects. Here, we have worked upon a small dataset, but a more clear and refined software cost can be gathered through large datasets. For

future extension, this area can be further improved by using other techniques such as neural network for and deep learning techniques for improved cost estimation. It is our belief that with the power of deep networks and proper hyperparameter tuning, very good cost estimation models can be obtained.

References

1. Keshta M (2017) Software cost estimation approaches: a survey. J Softw Eng Appl 10(10):824–842
2. Londeix B (1987)Cost estimation for software development. In: STC Telecommunications, Addison-Wesley Publishing Company, UK
3. Abbas SA, Liao X, Rehman AU, Azam A (2012) Cost estimation: a survey of well-known historic cost estimation techniques. J Emerg Trends Comput Inf Sci 3(4)
4. Mittas N, Angelis L (2013) Ranking and clustering software cost estimation models through a multiple comparisons algorithm. IEEE Trans Softw Eng 39(4)
5. Usman M, Mendes E, Weidt F, Britto R (2014) Effort estimation in agile software development. In: Proceedings of the 10th international conference on predictive models in software engineering
6. Sweta K, Pushker S (2013) Comparison and analysis of different software cost estimation methods. Int J Adv Comput Sci Appl 4(1)
7. Yadav VK (2017) Software cost estimation: process, methods and tools. Global Sci-Tech 9(3):138
8. Chirra SMR, Reza H (2019) A survey on software cost estimation techniques. J Softw Eng Appl 12(06):226–248
9. Zaffar Iqbal S (2017) Comparative analysis of common software cost estimation modelling techniques. Math Model Appl 2(3):33
10. Siva Suryanarayana Ch, Satya Prakash S (2018) An assessment on software effort estimation. Int J Adv Res Ideas Innov Technol 4(3)
11. Satyananda (2009) An improved fuzzy approach for COCOMO's effort estimation using Gaussian membership function. J Softw 4:452–459
12. Boehm BW, Abts C, Brown AW, Chulani S, Clark BK, Horowitz E, Madachy R, Reifer DJ, Steece B (2000)Software Cost Estimation with COCOMO II. Prentice Hall PTR
13. Puja D (2017) Intelligent method for cost estimation during software maintenance. Int J Eng Comput Sci
14. Rajper S, Shaikh ZA (2016) Software development cost estimation: a survey. Ind J Sci Technol 9(31)
15. Kazim M (2018) Comparison of various software cost estimation models. Global Sci-Tech 10(3):132
16. Mishra A, Dubey D (2013) A comparative study of different software development life cycle models in different scenarios. Int J Adv Res Comput Sci Manage Stud 1(5)
17. Zhao Y, Tan HBK (2003) Software cost estimation through conceptual. In: Proceedings of third international conference on quality software

OpenCL™ Implementation of Fast Frequency Domain Image Denoisification Kernels for Parallel Environments

Ashutosh Satapathy and L. M. Jenila Livingston

Abstract From the last couple of decades, image denoisification is one of the challenging areas in the image processing and computer vision domains that adds clarity to images by removing noise and makes them suitable for further processing. Images or videos captured by various visual sensors are degenerated during various circumstances such as sensor temperature, light intensity, target environment, erroneous electric instruments, and interference in the communication channels and required to be enhanced by various filtering techniques. Most of the real-world applications like student monitoring, vigilance system, traffic analysis, target detection, and classification use image smoothing to reduce artifacts present in the captured images and additionally demand its quicker execution. In this paper, we have explained different types of noise models normally degrade the quality of an image and the three most popular frequency domain smoothing filters (ideal, Gaussian, and Butterworth) are available to remove those noises. In addition to that, OpenCL kernels of those filters have been implemented for rapid denoisification and can be easily portable on wide verities of parallel devices. At last, performance evaluation has been carried out for both the CPU and GPU implementations to seek out their productiveness on those noise models in terms of time and precision. Various performance assessment metrics like entropy, root-mean-squared error, peak signal-to-noise ratio, and mean absolute error are applied to determine the rightness of the above filters.

Keywords Image smoothing · Noise models · Frequency domain transformation · Low-pass filter · OpenCL architecture · Performance metrics

A. Satapathy (✉) · L. M. Jenila Livingston
School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, Chennai 600127, India
e-mail: ashutosh.satapathy1990@gmail.com

L. M. Jenila Livingston
e-mail: jenila.lm@vit.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_52

1 Introduction

An image can be a two- or a three-dimensional pictorial representation of visual perception that speaks millions of words, and it is captured by various objects or phenomena like telescope, microscope, cameras, water, glass, or human eyes. The images captured by different devices in real-world scenarios are mostly affected by certain additive and multiplicative noises due to low light intensity, sensor temperature, erroneous electro-optical devices, target environment, and interference in the communication channels, which need to be preserved. Image denoisification is one of the significant preprocessing steps in the field of image processing and computer vision, and its quicker implementation is the challenging one too. Most of the real-world applications like student monitoring, vigilance system, traffic analysis, target detection, and classification use image smoothing to reduce artifacts present in the captured images and additionally demand its quicker execution [1–3].

Image smoothing needs greater interpretation of the noise distribution throughout an image, and the intention behind it is to reduce the noise present in the image without losing any details that make the image suitable for subsequent processing [4, 5]. In smoothing, pixels having higher values are reduced because they are the most probable points modified due to noise while the pixels having lower values than its neighbor are increased, leading to a smoothed image. So, most of the image smoothing techniques use the low-pass filters to discard the higher-frequency information that is normally contaminated by the heterogeneous noises and selection of the cutoff frequency decides the quality of the final image. A noisy image is affected by the speckle noise with variance 0.05 and its smoothed image by Gaussian low filter, as shown in Fig. 1.

Image smoothening can be done in the spatial domain or frequency domain. All the low-pass filters are normally classified into two types based on their operations: One is linear, and the other one is nonlinear. In linear filters, the output of a pixel depends on linear combinations of its neighbor pixels and it is represented as a straight line equation in a higher dimension whereas a nonlinear filter can be represented as a curve of two or higher degrees in the same dimension. We have reviewed different types of noise models and developed OpenCL kernels for the frequency domain low-pass filters for their faster implementation.

Fig. 1 An example of image smoothing: **a** noisy image and **b** its restored image

2 Related Work

In 2014, Dyre and Sumathi analyzed the frequency domain Gaussian bandpass and high-pass filters to enhance different grayscale fingerprint images of size 480 × 320 pixels each. Here, the MATLAB implementations of the Gaussian bandpass and high-pass filters consumed 0.8644 and 0.8892 s CPU time, respectively, to enhance the same image [6]. Chinnasamy and Vanitha implemented and examined different restoration filters to remove Brownian motion noise from the brain MRI images, whose execution needed at least 2.2 s for their MATLAB implementations on a CPU, which made them unsuitable for time-bound real-time applications [7]. On the other hand, MATLAB 7.8 implementation of Gaussian low-pass filter on a four-core CPU took on an average 0.0985 s to clean salt and pepper, speckle, Poisson, and Gaussian noises from a single channel 128 × 128 Lena image, while such filter required 0.09978 s to enhance the same size Cameraman image [8]. Ghosh and Chaudhury from the Indian Institute of Science proposed a fast approximation of Gaussian kernel using raised cosines filter and Monte Carlo sampling, whose enhancement time for a 512 × 512 RGB image is about 17.21 s on an Intel four cores @3.40 GHz CPU machine using MATLAB 8.4 [9]. In the same year, Yano and Kuroki approximated the Gaussian filter with the help of addition and shift operators to implement multilayer convolution of binominal filter and its GPU implementation on the 2688 cores NVIDIA GTX TITAN needed approximately 90 s to enhance an image of size 4096 × 1712 pixels [10]. As a bilateral filter is computationally expensive, approximating such filter using polynomial and trigonometric functions and applying them in the discrete Fourier transform domain minimizes the computational cost to 0.987 s for a 512 × 512 pixels sized image on a 3.4 GHz quad-core CPU [11]. Image enhancement using the 5 × 5 Gaussian filter consumes 12.5 s on the Intel Core i5 7th generation CPU in the spatial domain, while CUDA implementation of it requires only 7 ms on 384 cores NVIDIA GPU, but its unique implementation makes it unsuitable for other parallel platforms [12]. Rakhshanfar and Amer proposed a cascading network of Gaussian filters from multiple domains that utilize unrelated features to increase image quality at the expense of 49.65 s computational cost for a 768 × 512 gray image [13].

Chen et al. used the Lagrange multiplier technique to optimize the tight bound of the quaternion mean square error between the original and enhanced images and obtain the optimal weight values for the non-local mean filter, whose CPU @2.40 GHz execution time is about 176 s to remove the Gaussian noise from a 512 × 512 gray image [14]. The MATLAB implementation of naïve Butterworth bandpass and high-pass filters in the frequency domain takes 0.8268 s and 0.7020 s CPU time to enhance the size of a 480 × 320 pixel gray fingerprint image [6]. Shukla and Singh used Butterworth low-pass filter to remove the salt and pepper, speckle, Poisson, and Gaussian noises from a 128 × 128 grayscale Lena image in 0.072 s using quad-core CPU. Concurrently, the same filter took around 0.085 s to remove such noises from the 128 × 128 pixels grayscale Cameraman image [8]. Later, in 2017, Fadloullah, Mechaqrane, and Ahaitouf reconstructed the Butterworth filters using artificial bee

colony algorithms to get the optimal value in less execution time and compared its performance with the same filter implementation using ant colony and particle swarm optimization algorithms, but natural slowness of such evolutionary algorithms made it unusable for limited-time applications [15]. In the same year, Fan et al. applied the single-scale Retinex algorithm on the image representations of a colored image in HSV and RGB space separately and the resulted images were fused together to generate the final enhanced image, but the Butterworth filter built inside the Retinex algorithm let an image to be defogged in 0.648 s [16].

3 Noise Models

Most digital images are contaminated by various noises, which introduce unwanted signals and produce unpleasant outcomes such as unknown artifacts, broken edges and lines, unseen corners, faded objects, and distorted backdrop in the image. These noises are introduced due to various circumstances such as sensor temperature, light intensity, target environment, erroneous electric instruments, and interference in the communication channels. Noise in a digital image is modeled either by probability density function (PDF) or histogram which is overlapped on PDF of the actual image [17]. Noise is defined as an action 'η' which is not present in the original signal 'f' but contaminates the acquired image 'g', and the mathematical expressions of both additive and multiplicative noises are given in Eqs. (1) and (2).

$$g(x, y) = f(x, y) + \eta(x, y) \tag{1}$$

$$g(x, y) = f(x, y) \times \eta(x, y) \tag{2}$$

Depending upon the noise PDF, the influence on images is to be of different extents. Modeling of some of the popular noises present during real-time image acquisition is discussed in the upcoming subsection.

3.1 Gaussian Noise

In the spatial domain, Gaussian noise is one of the statistical independent noises that satisfies the properties of the normal distribution (Gaussian) function and it is generated due to various factors such as poor illumination, high temperature, object radiation, conversion of signals from optical to electrical, and signal amplification during acquisition [18, 19]. As most of the cases, Gaussian noise in an image disturbs the distribution of intensity throughout the image, and it is modeled as a PDF or a normalized histogram with respect to intensity values. The PDF of Gaussian noise w.r.t. noise intensity 'u' is given in Eq. (3).

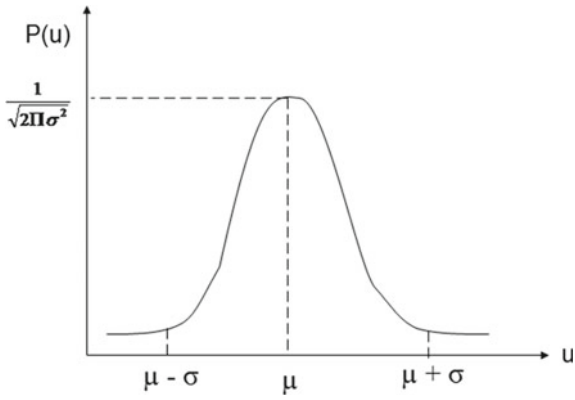


Fig. 2 PDF of Gaussian noise

$$P(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(u-\mu)^2}{2\sigma^2}} \tag{3}$$

Independent and identically distributed Gaussian noise throughout the image is represented by a bell-shaped curve as shown in Fig. 2. Seventy percentage of pixels of the noisy image are in the noise range $[\mu - \sigma, \mu + \sigma]$, whereas 90% of pixels of the noisy image are in the noise range $[\mu - 2\sigma, \mu + 2\sigma]$. Here, mean '$\mu$' is the expectation of noise intensities and 'σ' is the standard deviation of the noise throughout the image.

3.2 Localvar Noise

Localvar noise is the special case of white Gaussian noise with zero mean, and the standard deviation is dependent on pixel intensities throughout the image [20]. Unlike Gaussian noise, Localvar noise has both +ve and –ve components which randomly increase and decrease pixel values of the actual image to produce the degraded one. The PDF of Localvar noise w.r.t. noise intensity 'u' is given in Eq. (4).

$$P(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-u^2}{2\sigma^2}} \tag{4}$$

Independent and identically distributed Localvar noise throughout the image is represented by a bell-shaped curve centered at origin as shown in Fig. 3. Seventy percentage of pixels of the noisy image are in the noise range $[-\sigma, \sigma]$, whereas 90% of pixels of the noisy image are in the noise range $[-2\sigma, 2\sigma]$. Here, σ is the standard deviation of the noise throughout the image and u represents noise intensity of the input image.

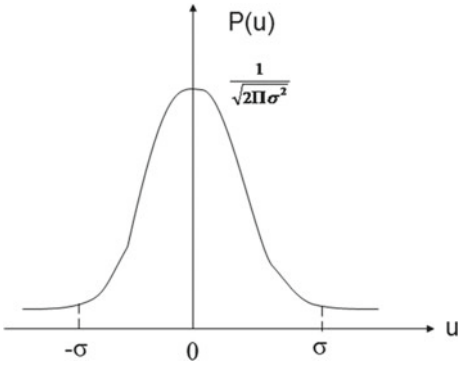


Fig. 3 PDF of Localvar noise

3.3 Uniform Noise

Uniform noise, so-called quantization noise, is one of the statistical noises that satisfies the properties of the uniform distribution function [19]. The PDF of uniform noise w.r.t. noise intensity 'u' is given in Eq. (5). It is introduced during the quantization process while converting an analog signal to a digital one, i.e., quantizing pixel values of an analog image to different discrete levels in the spatial domain. Signal-to-quantization noise ratio (SQNR) of a degraded image is restricted by minimum and maximum pixel values in the image, and it is directly proportional to the difference between these two which is given in Eq. (7).

$$P(u) = \begin{cases} \frac{1}{(b-a)} \text{if} a \leq u \leq b \\ 0 \text{otherwise} \end{cases} \tag{5}$$

$$\mu = \frac{a+b}{2} \text{ and } \sigma^2 = \frac{(b-a)^2}{2} \tag{6}$$

$$\text{SQNR} = 20 \log_{10} \frac{I_{\max} - I_{\min}}{\sigma} \tag{7}$$

As quantization noise is equally distributed throughout the image, and it is represented as a rectangle as shown in Fig. 4. Here, μ and σ represent mean and variance of noise intensities added during image degradation where 'a' and 'b' are the minimum and maximum noise intensities present in the image. In SQNR, Imax and Imin stand for maximum and minimum pixel values of a noisy image.

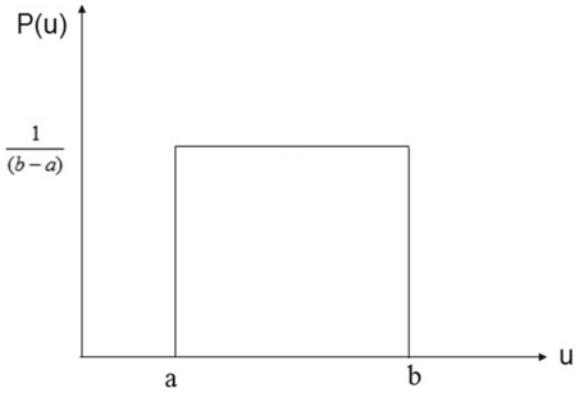


Fig. 4 PDF of uniform noise

3.4 Salt and Pepper Noise

Salt and pepper noise is a special case of bipolar impulse noise which statistically drops the actual pixel values of the original image and replaces them with two intensity levels 0 or 255 [21]. As intensity values 255 and 0 seem like a white and black spot on the image, it is called as salt and pepper noise. This is introduced due to errors in analog-to-digital converter orbit error during data transmission, and sometimes it is inserted by uncleansed sensor lens or erroneous storage space. The PDF of salt and pepper noise w.r.t. noise intensity 'u' is given in Eq. (8) and shown in Fig. 5.

$$P(u) = \begin{cases} P_{\mathrm{a}} & \text{if } u = a \\ P_{\mathrm{b}} & \text{if } u = b \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Figure 5 represents the salt and pepper noise PDF with zero mean and variance of 0.05. It consists of two impulse distributions at level 'a' and level 'b' which are

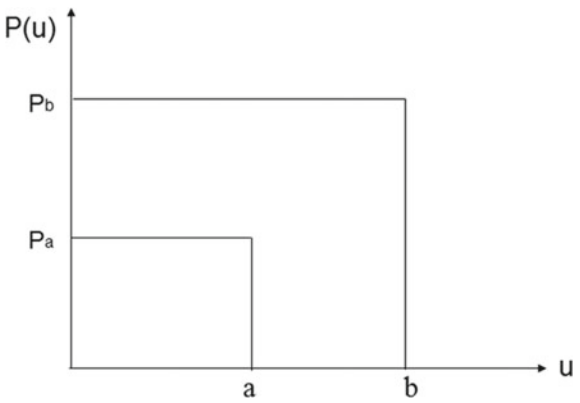


Fig. 5 PDF of salt and pepper noise

also known as dark level and bright level, respectively. It clearly shows that the distribution of bright noise is more compared to dark noise throughout the image. However, salt and pepper noise does not corrupt the full image; instead, it modifies some pixel values in the image.

3.5 Impulse Noise

Impulse noise is unipolar that statistically drops the actual pixel values of the original image and replaces them with a particular intensity level '*a*' as shown in Fig. 6. Impulse noise is also called short noise which degenerates the image quality, and it is occurring due to errors in image acquisition, quantization, or transmission or due to atmospheric or sensor heat [22]. Signal interference in communication medium and environmental disruption during data transmission sometimes caused impulse noise to be added. Unlike the salt and pepper noise, salt or pepper noise is also considered as impulse noise that introduced either bright or dark spots over the entire image.

$$P(u) = \begin{cases} P_{\mathrm{a}} & \text{if } u = a \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Equation (8) with P_{a} or P_{b} equal to zero is considered as unipolar impulse noise. Impulse noise is arbitrarily scattered all over the image, and its values are independent and uncorrelated to image pixel values. However, unlike Gaussian noise, impulse noise does not corrupt the full image; instead, it modifies some pixel values in the image while keeping the rest of the pixel values as they are.

Fig. 6 PDF of impulse noise

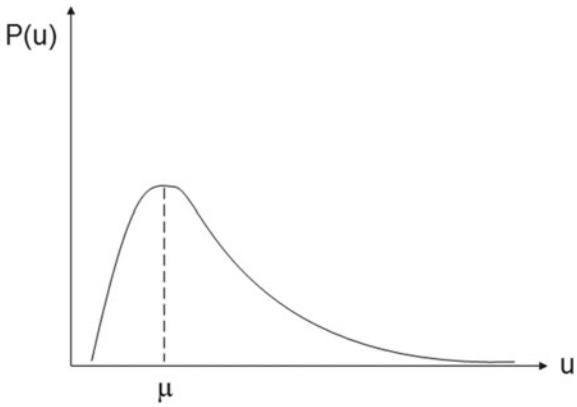


Fig. 7 PDF of Poisson noise

3.6 Poisson Noise

In the spatial or temporal domain, Poisson noise is also called as quantum noise and shot noise that satisfies the properties of the Poisson distribution function and it is created due to discrete nature of the electric charge or various electromagnetic waves such as X-ray, gamma ray, or visible light emitted from optical devices [23]. As this light source injects photons on the object, a random number of photons per unit time emitted from these light sources introduce noise in the final image. In the case of electric charge or intensity of light, the absolute value of this noise is directly proportional to the root over of the average amount of occurrences (Fig. 7).

$$P(u) = \frac{e^{-\mu}\mu^{u}}{u!} \tag{10}$$

In Eq. (10), μ represents mean and variance of noise distribution while u stands for intensity levels. If the intensity of light increases quickly, the proportion of the noise decreases whereas signal-to-noise ratio increases rapidly. Thus, shot noise is often occurred due to low light intensities or little current flow that has been amplified.

3.7 Speckle Noise

Speckle noise is considered as a random and deterministic noise that occurs in radar, SAR, ultrasound, and the OCT system, and it degrades the quality of images produced from it. During image acquisition, the imaging sensor transmits a signal to the target body and the reflected signal from the target is captured by it. During transmission, the signal is exposed to various additive noises throughout the medium. The asymmetric and rough surface of the target body makes the magnitude of the reflected signal vary along with noise magnitude as noise is also reflected from the target surface. As

the magnitude of the speckle noise is directly proportional to the pixel intensity, it is also considered as a multiplicative noise [24]. The PDF of speckle noise is given in Eq. (11).

$$P(u) = \frac{1}{\sigma^2} e^{\frac{-u}{\sigma^2}} \tag{11}$$

$$g(x,\ y) = f(x,\ y) \times u(x,\ y) + \delta(x,\ y) \tag{12}$$

Here, σ^2 and u stand for variance and noise intensity level, respectively. In Eq. (12), $f(x, y)$ and $g(x, y)$ are the actual image and the degraded image due to speckle noise, respectively, where $u(x, y)$ and $\delta(x, y)$ represent the multiplicative and additive components of the speckle noise whose values are varying based on the signal strength.

3.8 Rayleigh Noise

Unlike the Gaussian noise, Rayleigh noise is one of the asymmetric statistical noises that obeys Rayleigh distribution which is skewed to its right. Normally radar and velocity images are contaminated by this noise, and it is heavily affected by how quickly transmitter and/or receiver are moving [25]. Satellite images are also affected by this noise due to the target environment and ionospheric and tropospheric propagation of radio signals. It is happened due to no transmission influence across a line of sight between the sensor and the target. The PDF of Rayleigh noise with respect to noise intensity 'u' is given in Eq. (13) and shown in Fig. 8.

$$P\ (u)\ = \begin{cases} \frac{2}{b}(u-a)e^{\frac{-(u-a)^2}{b}} & \text{if } u \geq a \\ 0 \ \text{ if } u < a \end{cases} \tag{13}$$

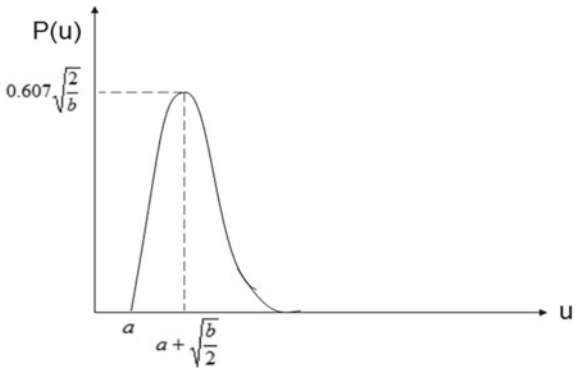


Fig. 8 PDF of Rayleigh noise

$$\mu = a + \sqrt{\frac{\pi b}{4}} \text{ and } \sigma^2 = \frac{b(4-\pi)}{4} \tag{14}$$

Normally this noise is found in magnetic resonance imaging (MRI) images as it is recorded as complex images but displayed as magnitude images. As pixel values are Rayleigh distributed in an MRI image, image intensities are contaminated by Rayleigh noise which satisfies the above distribution where μ and represents mean and variance of Rayleigh PDF.

3.9 Gamma Noise

Like the Rayleigh noise, gamma or Erlang noise is also one of the asymmetric statistical noises that obeys gamma distribution and normally seen in laser-based images. The gamma-injected charge diminishes the signal handling ability and coverts the origin of the acute noise. The gamma radiation-introduced noise is larger than the optical shot noise because each gamma produced a Compton electron that generates lots of electron–hole pairs. So, the amount of noise carriers is directly proportional to the shot noise on the arbitrary arrival of the Compton electrons multiplied by the expected amount of electron–hole pairs introduced by Compton electrons [26].

$$P(u) = \begin{cases} \dfrac{a^b u^{b-1}}{(b-1)!} \mathrm{e}^{-au} & \text{if } u \geq 0 \\ 0 & \text{if } u < 0 \end{cases} \tag{15}$$

$$\mu = \frac{b}{a} \text{ and } \sigma^2 = \frac{b}{a^2} \tag{16}$$

The PDF of Rayleigh noise with respect to noise intensity 'u' is given in Eq. (15) and shown in Fig. 8. The magnitude of gamma radiation noise is 30–60 times larger than the Poisson noise on an optically generated image of the same size. Here, μ and σ^2 are mean and variance of the gamma distribution function whose values are dependent on gamma radiation background of the target environment (Fig. 9).

3.10 Exponential Noise

Exponential noise is an independent statistical noise that has been seen in various applications such as radar imaging, and neurological and astronomical imaging. These images often consist of random noise that does not come from the target environment, but the sources in electronic instrumentation during image acquisition. Light intensity and sensor temperature are the crucial factors that introduce noise in

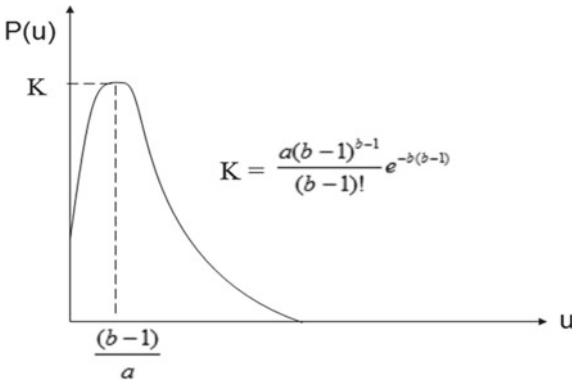


Fig. 9 PDF of gamma noise

the image during its acquisition, whereas interference in the communication channels also degrades image quality during its transmission. These noises can be dependent or independent, and exponential noise is one of the independent noises appeared in the image [27].

$$P(u) = \begin{cases} a\mathrm{e}^{-au} & \text{if } u \geq 0 \\ 0 & \text{if } u < 0 \end{cases} \tag{17}$$

$$\mu = \frac{1}{a} \text{ and } \sigma^2 = \frac{1}{a^2} \tag{18}$$

The PDF of exponential noise is the special case of Erlang noise with $b = 1$ as shown in Fig. 10 where mean 'μ' and variance 'σ^2' are dependent on the control parameter 'a'. The control parameter 'a' manages the expandability of the exponential curve. Smaller the value of 'a' makes the curve expanded larger, and the noise is scattered better than the larger value of 'a'.

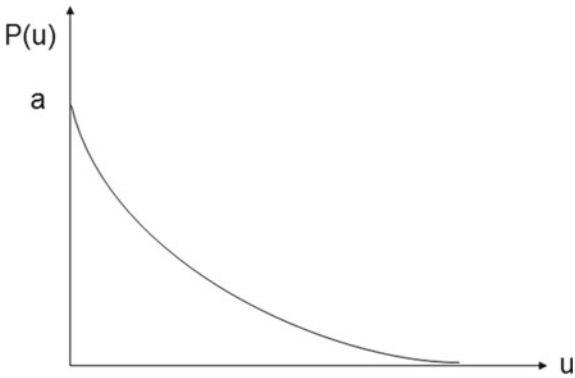


Fig. 10 PDF of exponential noise

4 Image Smoothening

As we discussed earlier, image smoothing is one of the image enhancement techniques used to remove the noise and make the image suitable for further processing. Image smoothening can be done in the spatial domain or frequency domain [28]. Smoothening in the spatial domain requires mask operations where an image region of mask size is multiplied by weighted values of the mask to compute the final value of its centered pixel. This process is repeated until the final values for all the pixels are computed in an image. In the frequency domain, various transformation techniques are used to transfer the image and this transformed image is multiplied with a frequency space filter to compute the final values in the transformed domain. As smoothening in frequency domain follows point operations instead of mask operations, it makes them faster and we have discussed various frequency domain smoothening filters in the upcoming section.

4.1 Frequency Domain Transformation

Frequency domain analysis is one of the important operations in the field of image processing, mostly used in image enhancement, restoration, compression, and feature extraction. The spatial domain represents the change in pixel intensities w.r.t. scene, whereas the frequency domain represents the rate of change in pixel intensities in the spatial domain. It transforms an image of size N × N into an N × N array of frequency coefficients that represents the actual image [29]. Operation in frequency space requires an image transformation, performs an operation on the transformed image and transforms it back to get the final image in the spatial domain.

$$F(k,\ l) = T[f(x,\ y)] \tag{19}$$

Here, $f(x,\ y)$ and $F(k,\ l)$ are the pixel values in spatial and frequency domain, respectively, where T is the transformation technique. In the frequency space, each pixel value indicates a specific frequency involved in the spatial image. The two most important frequency domain transformations used for image enhancement, compression, and feature extraction are discussed below.

4.1.1 Discrete Fourier Transformation

Discrete Fourier transformation (DFT) is one of the most used transformation techniques that extracts frequency components from the image and represents them as a combination of sine and cosine functions. This transformation of a two-dimensional vector in the spatial domain is also called as two-dimensional DFT (2D-DFT) whose

frequency domain image represents absolute values of its sine and cosine components at a particular point [30]. A discrete Fourier transform of an image extracts sufficiently visible frequency components that dominate the image in the spatial domain without containing all the frequency values present in an image. Forward and inverse discrete Fourier transformations of an image are given in Eqs. (20) and (21).

$$F(k,\ l) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,\ y) \mathrm{e}^{-j2\pi \left(k \frac{x}{M} + l \frac{y}{N}\right)} \tag{20}$$

$$F(x,\ y) = \frac{1}{\sqrt{MN}} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} F(k,l) \mathrm{e}^{j2\pi \left(k \frac{x}{M} + l \frac{y}{N}\right)} \tag{21}$$

$$0 \leq x,\ k \leq M-1 \text{ and } 0 \leq y,\ l \leq N-1 \tag{22}$$

Here, $f(x,\ y)$ is the pixel intensity of an image in the spatial space and the exponential function is the unitary matrix responsible for creation of each point $F(k,\ l)$ in the frequency domain where M and N represent total number of row and column pixels in a digital image, respectively.

4.1.2 Discrete Cosine Transformation

Like DFT, discrete cosine transformation (DCT) also converts a 1D or 2D information in the temporal or spatial space to represent its transformed values in frequency space. Unlike the DFT, it generates only real numbers which make it much faster than DFT to operate. The transformation of a 2D matrix in the spatial domain is also called as two-dimensional DCT (2D-DCT) whose transformed domain image represents integer values of the real components at a particular point [31]. The DC components of a DCT transformed image are located at (0, 0) pixel whose value is extremely higher than the rest of the image. Forward and inverse discrete cosine transformations of an image are given in Eqs. (23) and (24).

$$F(k,\ l) = \frac{2}{\sqrt{MN}} C(k)C(l) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,\ y) \cos\left[\frac{(2x+1)k}{2M}\right] \cos\left[\frac{(2y+1)l}{2N}\right] \tag{23}$$

$$f(x,\ y) = \frac{2}{\sqrt{MN}} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} C(k)C(l)F(k,l) \cos\left[\frac{(2x+1)k}{2M}\right] \cos\left[\frac{(2y+1)l}{2N}\right] \tag{24}$$

$$0 \leq x,\ k \leq M-1,\ 0 \leq y,\ l \leq N-1 \text{ and } C(\upsilon) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if} \upsilon = 0 \\ 1 & \text{if} \upsilon > 0 \end{cases} \tag{25}$$

Here, $f(x, y)$ is the pixel intensity of an image in the spatial space and the cosine function is the unitary matrix responsible for creation of each point $F(k, l)$ in the frequency domain where $C(k)$ and $C(l)$ are the control parameters whose values are varied w.r.t. pixel position in the frequency space.

4.2 Frequency Domain Filters

Filtering in the frequency domain requires a point-to-point multiplication of frequency components with suitably chosen weight function which attenuates higher-frequency components by keeping lower-frequency components as they are [32, 33]. The process of selective lower-frequency exclusion from an image is termed as 2D frequency domain high-pass filtering. The value of the filter kernel is dependent on the distance between the pixel positions from its DC components in frequency space. The computation of distance $D(k, l)$ is differed for both DFT and DCT transformations and given in Eqs. (28) and (29).

$$F'(k,\ l) = G(k,\ l) \times H(k,\ l) \tag{26}$$

$$f'(x,\ y) = F^{-1}\left[F'(k,\ l)\right] \tag{27}$$

$$\text{In DFT image: } D(k,\ l) = \sqrt{\left(\frac{x}{2} - abs\left(\frac{k}{2} - \frac{x}{2}\right)\right)^2 + \left(\frac{y}{2} - abs\left(\frac{l}{2} - \frac{y}{2}\right)\right)^2} \tag{28}$$

$$\text{In DCT image: } D(k,\ l) = \sqrt{k^2 + l^2} \tag{29}$$

Here, $G(k, l)$ and $H(k, l)$ are the transformed noisy image and filter kernel at position (k, l), respectively. Final noise-free image $f'(x, y)$ is constructed by taking the inverse transformation of a smoothed image in the frequency domain. Two types of filter kernels, non-separable and separable filter kernels used for image smoothening, are discussed below.

4.2.1 Ideal Filter

The ideal low-pass filter removes all the frequencies above the cutoff frequency while keeping the lower frequencies unchanged. The frequency response of an ideal low-pass filter is a rectangular function where transition region does not exist between

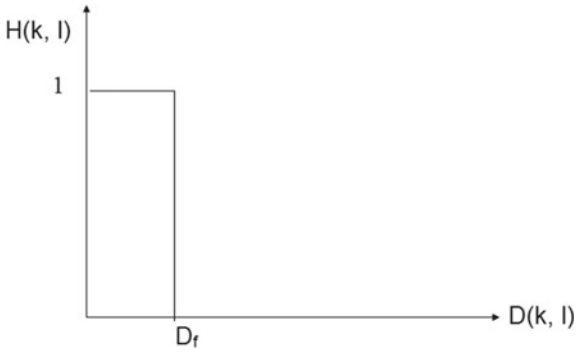


Fig. 11 Ideal low-pass filter

pass band and stop band in contrast to practical filter; i.e., the value of the transfer function (TF) after the cutoff frequency immediately falls to zero [34]. The point spread function (PSF) of a 2D low-pass ideal filter of cutoff frequency D_f is given in Eq. (30) (Fig. 11).

$$H(k,\ l) = \begin{cases} 1 & D(k,l) \leq D_f \\ 0 & D(k,l) \geq D_f \end{cases} \tag{30}$$

Frequency response from a 2D ideal low-pass filter is mathematically realized by multiplying frequency map of an input image matrix with its rectangular function in the frequency domain which is similar to convolving its impulse response with image pixel values in the spatial domain.

4.2.2 Gaussian Filter

Gaussian low-pass filter is a practical filter having frequency response alike a Gaussian function, used to enhance image quality in the frequency domain by reducing noise during image preprocessing stage. Unlike an ideal filter, the PSF is a monotonically decreasing function that decreases gradually and reduces to zero after the cutoff frequency whose response transition between its pass band and stop band increases as with its cutoff frequency [35]. The transfer function of a two-dimensional low-pass Gaussian filter is given in Eq. (31) where the cutoff frequency Df decides smoothness of its transition region in this filter (Fig. 12).

$$H(k,\ l) = \mathrm{e}^{\frac{-D^2(k,l)}{2D_f^2}} \tag{31}$$

The output from a Gaussian low-pass filter is realized mathematically by multiplying transformed values of an image with its Gaussian distributed value in the frequency domain which is similar to convolving its 2D discrete response with image pixels in the image space. Amplification of its cutoff frequency widens the transition

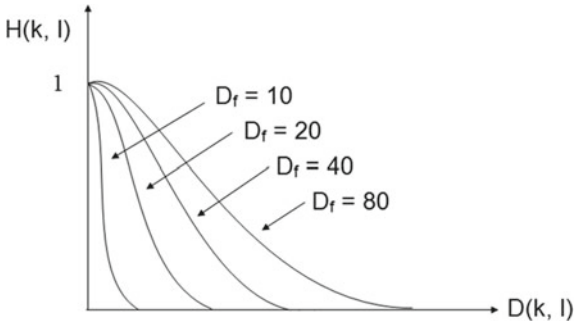


Fig. 12 Gaussian low-pass filter

between pass band and stop band that allows a tiny range of frequencies even after the cutoff frequencies.

4.2.3 Butterworth Filter

Butterworth low-pass filter is one of the practical filters whose frequency response is uniform in the pass band and gradually decreases to zero in the stop band. As the PSF of this low-pass filter is a smooth transfer function without any sharp cutoff frequency, it makes the filter suitable for many applications and the range of frequencies allowed by the filter is directly dependent on its cutoff frequency and order of the filter. The transfer function of a two-dimensional low-pass Butterworth filter is given in Eq. (32) where 'n' is the order of the filter and $D(k, l)$ is the distance from the origin (Fig. 13).

$$H(k,\ l)\ =\ \frac{1}{1+\left[\frac{D(k,l)}{D_f}\right]^{2n}} \tag{32}$$

A first-order Butterworth filter's response transition between pass band and stop band is at −6 dB per octave, while its value decreases as the order of this filter

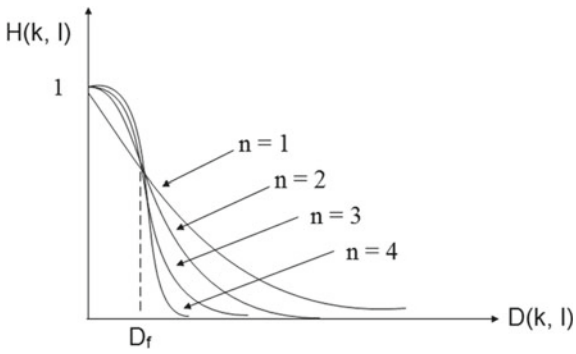


Fig. 13 Butterworth low-pass filter

increases, i.e., −12 and −18 dB per octave in a second-order and third-order filters, respectively [36]. Instead of generating non-monotonic ripples at pass band and/or stop band like other filters, its magnitude decreases monotonically with respect to an increase in frequency '*D*'.

5 OpenCL Architecture

Parallel processing allows multiple computing units or devices to work together and perform a particular task. Due to a limited amount of computing units, processing power, memories and deficiency of performing task rapidly, made available of a new computing device called graphics processing unit (GPU). Later, various programming standards were defined to make utilization of these computing devices. Some of them are OpenMP, MPI, OpenACC, CUDA, Pthreads, RenderSscript, and OpenCL. These programming standards or APIs are of two types: Some of them are heterogeneous, and others are architecture specific.

OpenCL programming standards and APIs allow professionals to execute their programs in various heterogeneous environments [37]. OpenCL provides low-level programming abstraction like 'C' and supports various device architectures. It is also included in many libraries and packages such as JACKET, MAGMA, BOLT C++, clAMDFFT, and clAMDBLAS for acceleration. It uses an extended version of C to define kernels and APIs for creating and managing them. Runtime compilation of kernels makes host applications to take benefit of all the computing devices. OpenCL architecture, mainly divided into four sophisticated models, is discussed below [38].

5.1 OpenCL Platform Model

The OpenCL platform model provides useful information about the architecture of different computing devices which benefits the programmer for kernel execution later. A host machine is connected to more than one computing device such as CPUs, GPUs, FPGAs, or any low-cost microcontrollers. Each computing device is the collection of more than one computing units, and each computing unit has many processing elements. Intel HD Graphics P530 is integrated with Xeon E3 1225 V5 microprocessor, uses mid-tier Gen9 microarchitecture that consists of 24 compute units, and 256 processing elements are combined to make one compute unit [39]. NVIDIA GTX 1050 Ti consists of twelve computing units, and each computing unit has thirty-two processing elements [40]. OpenCL kernels targeted to computing devices efficiently utilize processing elements to make applications run faster.

5.2 OpenCL Execution Model

The execution model is defined over two units of execution: One is host program execution, and the other one is kernel execution. The host program is responsible for query information about platforms or devices, create, build, and manage kernels, create and delete host variables, create and release context, create and release command queues, etc. During the submission of the kernel, the host program also creates an N-dimensional index space where the value of 'N' is either one, two, or three. In index space, the processing element is referred to as a work item and each coordinate of index space represents a processing element that executes a single instance of the kernel. One-, two-, and three-dimensional index spaces are shown in Fig. 14.

To provide modularity and make execution faster, OpenCL introduces the concepts of *work-item*, *work-group*, *global-size*, *group-size*, *global-id*, *local-id,* and *work-group-id*. *global-size* identifies the index space size in each dimension. *work-item* in index space represents processing elements, and *global-id* represents its coordinate in index space. To provide modularity, index space is logically divided into several *work-groups. work-items* in a *work-group* have assigned to a single task. *local-id* represents a *work-items* coordinate in a *work-group*, whereas *work-group-id* and

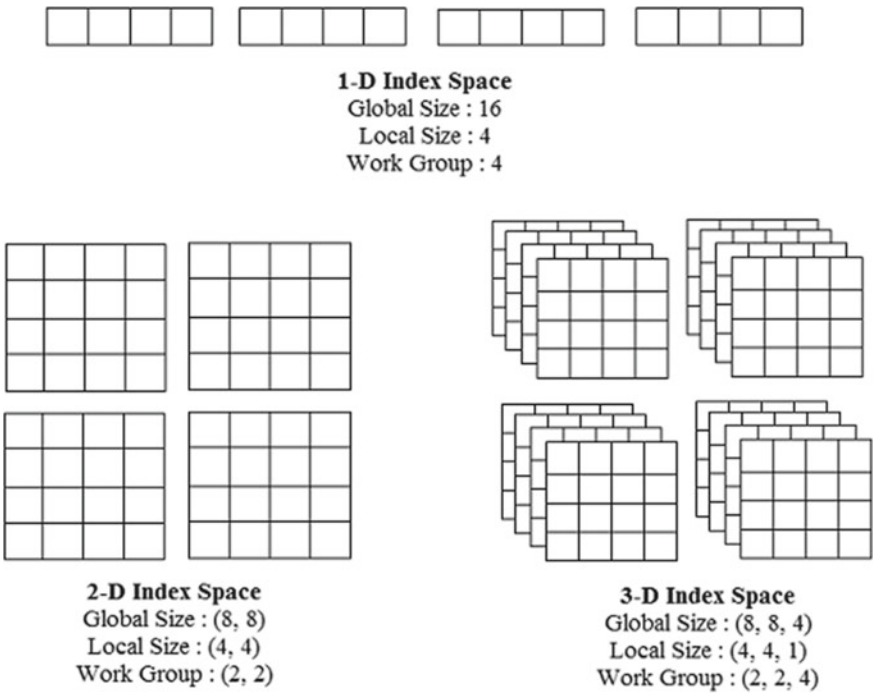


Fig. 14 OpenCL index space

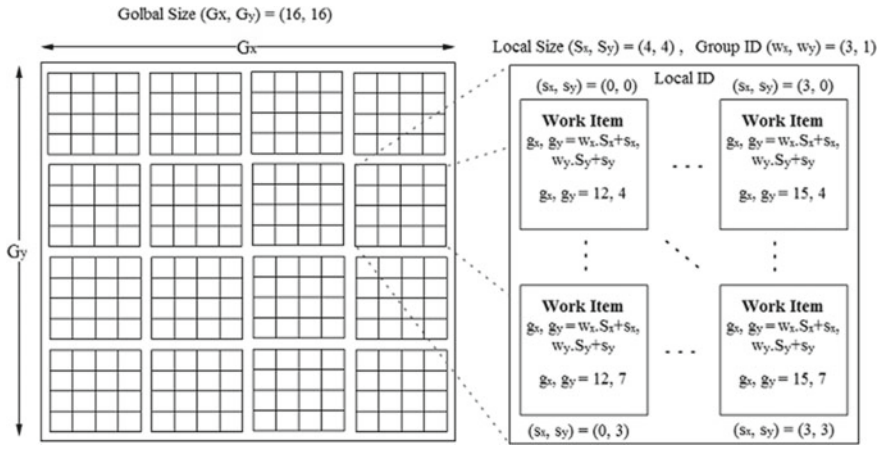


Fig. 15 OpenCL execution model

work-group-size represent a *work-group* coordinate and *work-group* size in index space, respectively. All these values are assigned by the host program at runtime. At last, Fig. 15 shows a pictorial representation of the above attributes and their calculations.

As shown above, the *global-size* of index space (G_x, G_y) is of (16, 16) and the *local-size* of each *work-group* (S_x, S_y) is (4, 4). (s_x, s_y) represents *local-id* of a work item in a *work-group. global-id* of a *work-item* inside a *work-group* of *group-id* (w_x, w_y) is calculated below.

$$\text{global - id}(g_x,\ g_y) = (w_x \cdot S_x + s_x,\ w_y \cdot S_y + s_y) \tag{33}$$

5.3 OpenCL Memory Model

The OpenCL memory model divides memory into four disjoint memory sections logically as shown in Fig. 16, and each section has its working principles and advantages. Each OpenCL device has its global memory, and it is the largest memory section that is shared among all the *work-items* in index space. This memory can be used as buffers based on program requirements. *__global* or *global* keywords specify buffers created in this section. Constant memory is another section of memory that is also shared among all the work items, but values are immutable throughout operations. *const* keyword used in a kernel specifies buffers in this region, and it is created by the host program by declaring OpenCL buffers as memory read-only. Each *work-group* has its local memory which is shared among its *work-items*, whereas each work item in index space has its private memory. *__local* or *local* keyword specifies

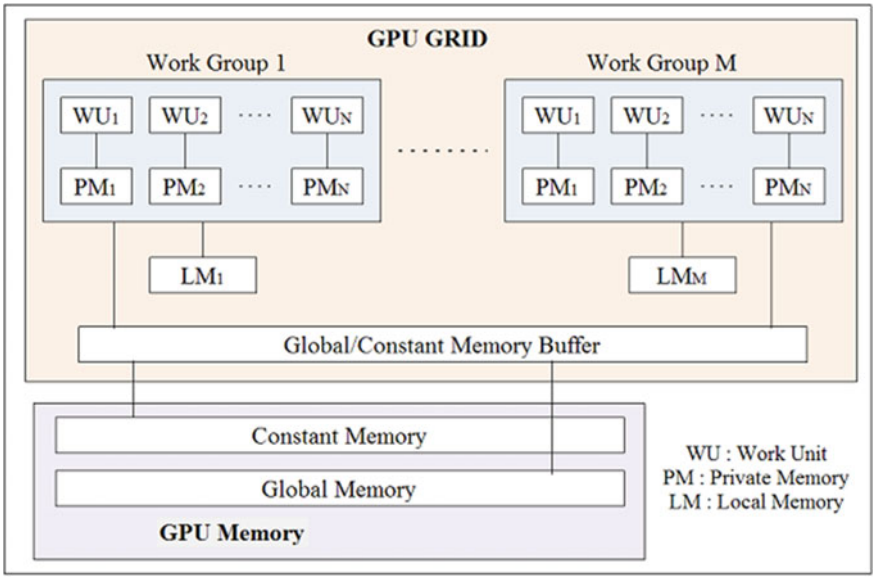


Fig. 16 OpenCL memory model

buffer created in local memory, and no specifiers or *__private* indicates items are created in private memory. In the OpenCL memory model, the size of memory is inversely proportional to the speed of memory. The memory model allows 1 KB to each work item as private memory, and speed of the private memory is faster than local and global memory. Global memory is the slowest memory section of memory size in some GBs. The memory regions are synchronized with each other to provide data consistency between *work-items* in index space.

5.4 OpenCL Programming Model

It not only allows multiple OpenCL platforms to reside in a single host machine but also allows user programs to select one of them during runtime [41]. OpenCL Installable Client Driver loads proper library files for the selected platform by the application. If the programmer wants to use multiple platforms at the same time, it also helps to do that. As the OpenCL programming model creates a separate context for all devices in a platform, it is not possible to pass data from one context to another context. The host program takes the responsibility to synchronize data transfer between multiple contexts. The application program distributes kernels across GPU cores, which are run independently from each other. So, the GPU of less core takes more time to execute than the GPU of a higher number of cores.

Here, we have used NVIDIA GTX 1050 Ti consisting of 768 cores for performance analysis.

6 OpenCL Kernels

OpenCL kernel for a computing device creates groups of threads that run by work-groups independently. It is created by the host program with the help of program objects and kernel objects which are also responsible for compilation and execution of kernels. The separate OpenCL program object is created for each context, and each program object can carry multiple kernel objects for different computing devices. OpenCL kernels take inputs from the host program, and the configuration of index space makes these kernels execute faster which ultimately makes the application faster. The portability nature of OpenCL kernels makes them easy to transfer between multiple platforms. However, writing an OpenCL kernel requires special attention because errors in the kernel are difficult to identify as compilation happens at runtime. At last, OpenCL kernel implementation of those above filters with respective *global_size* and *local_size* of each *work-group* is reported below.

6.1 DFT Kernel

```
#pragma OPENCL EXTENSION cl_khr_fp32 : enable
__kernel void DFT_Kernel (__global float2* data, __global float2* out,
                          __local float2* sharedMemory, int width, float norm)
{
       //Get the index of work items
       int globalID = get_global_id(0); int localID = get_local_id(0);
       int groupID = get_group_id(0); int groupSize = get_local_size(0);
       __local float2* sharedArray = sharedMemory;
       __private int v = globalID % width; __private float param = (-2.0*v)/width;
       // -2.0 is replaced by 2.0 in case of IDFT
       __private float c, s;__private float2 valueH, valueV;
       sharedArray[localID] = (0.0, 0.0);
       // Horizontal DFT Transformation
       for(int i = 0; i<groupSize; i++){
       valueH = data[groupSize*groupID+i];s = sinpi(param * i); c = cospi(param * i);
              sharedArray[localID] += (float2)(valueH.x*c-valueH.y*s,
                                               valueH.x*s+valueH.y*c);
       } out[groupSize*groupID+localID]=dot(sharedArray[localID],float2(norm, norm));
       sharedArray[localID] = (0.0, 0.0);
       // Vertical DFT Transformation
       for(int i = 0; i<groupSize; i++){
       valueV = out[groupSize*i+groupID];s = sinpi(param * i); c = cospi(param * i);
       sharedArray[localID] += (float2)(valueH.x*c-valueH.y*s,
                                               valueH.x*s+valueH.y*c);
       } out[groupSize*localID+groupID]=dot(sharedArray[localID],float2(norm, norm));
       //* Replace ‘data’ buffer with ‘out’ buffer and vice versa in IDFT
```

6.2 Ideal Filter

```
#define EXP 2.72
#pragma OPENCL EXTENSION cl_khr_fp32 : enable
#pragma OPENCL EXTENSION cl_khr_printf : enable
__kernel void ideal_kernel (__global float2* data,
                                    int height,
                                    int width,
                                    int CUTOFF)
{
      // Get the index of work items
      uint index = get_global_id(0);
      __private int U = index / width; __private int V = index % width;
      __private float imgx = data[index].x;__private float imgy = data[index].y;
      __private float D = pow(height/2 - abs(U - height/2), 2.0) +
                              pow(width/2 - abs(V - width/2), 2.0);
      __private float H = (sqrt(D) < CUTOFF)? 1.0 : 0.0;
      data[index].x = imgx * H; data[index].y = imgy * H;
}
```

6.3 Gaussian Filter

```
#define EXP 2.72
#pragma OPENCL EXTENSION cl_khr_fp32 : enable
#pragma OPENCL EXTENSION cl_khr_printf : enable
__kernel void gaussian_kernel (__global float2* data,
                                    int height,
                                    int width,
                                    int CUTOFF)
{
      // Get the index of work items
      uint index = get_global_id(0);
      __private int U = index / width; __private int V = index % width;
      __private float imgx = data[index].x; __private float imgy = data[index].y;
      __private float D = pow(height/2 - abs(U - height/2), 2.0) +
                              pow(width/2 - abs(V - width/2), 2.0);
      __private float H = pow(EXP, (-1.0 * D / (2.0 * pow(CUTOFF, 2.0))));
      data[index].x = imgx * H; data[index].y = imgy * H;
}
```

6.4 Butterworth Filter

```
#define EXP 2.72
#pragma OPENCL EXTENSION cl_khr_fp32 : enable
#pragma OPENCL EXTENSION cl_khr_printf : enable

__kernel void butterworth_kernel (__global float2* data,
                                  int height, int width,
                                  int CUTOFF,
                                  float Ord)
{
      // Get the index of work items
      uint index = get_global_id(0);
      __private int U = index / width; __private int V = index % width;
      __private float imgx = data[index].x;__private float imgy = data[index].y;
      __private float D = pow(height/2 - abs(U - height/2), 2.0) +
                          pow(width/2 - abs(V - width/2), 2.0);
      __private float H = 1 / (1 + pow (sqrt(D) / CUTOFF, Ord));
      data[index].x = imgx * H; data[index].y = imgy * H;
}
```

6.5 Global Size and Local Size

```
size_t global_size = Image.rows * Image.cols, size_t local_size = 256; /*DFT & IDFT*/
size_t global_size = Image.rows * Image.cols, size_t local_size = 256; /*Filtering*/
clStatus = clEnqueueNDRangeKernel(command_queue, kernel, 1, NULL, &global_size,
&local_size, 0, NULL, NULL);
```

7 Results

In this paper, we have used the above low-pass filters to remove unwanted signals introduced by various noises such as Gaussian, Localvar, uniform, salt and pepper, impulse, Poisson, and speckle. These frequency domain filters are implemented in C, OpenCV, and OpenCL programming languages. Results show that filter implementation in C and OpenCL is faster than C and OpenCV without any changes in the accuracy. Different performance metrics used to evaluate the accuracy of the above techniques are discussed below [42].

7.1 Entropy

Entropy describes the amount of information required to code an input image by any compression techniques. Image with lower contrast value or long runs of similar

pixels requires lesser information to compress than a high contrast image. So, an image with greater details has greater entropy value.

$$E = -\sum_{i} P_i \log_2 P_i \tag{34}$$

The entropy of a color image is calculated over three channels. Here, P_i is the probability of intensity level i. Higher the entropy means smaller –ve value and vice versa.

7.2 Mean Absolute Error

Mean absolute error (MAE) defines how the output image is closely related to the reference image. It is calculated by taking the average of the absolute value of pixels' difference over three channels. Lower the value indicates greater image enhancement. MAE of an RGB image is given below.

$$\text{MAE} = \frac{1}{3\text{MN}} \sum_{c} \sum_{i} \sum_{j} (|I_{\text{cs}}(i, j) - I_{\text{cr}}(i, j)|) \tag{35}$$

I_{cs}, I_{cr}, and c stand for the original image, enhanced image, and channel numbers. For a gray image, $c = 1$ and $\frac{1}{3\text{MN}}$ is replaced with $\frac{1}{\text{MN}}$.

7.3 Root-Mean-Squared Error

It is calculated by taking the square root of mean-squared error (MSE). Like MAE, the lower RMSE value indicates better quality improvement. It is also used to find the similarity between the reference and final images. The RMSE of an RGB image is computed as below.

$$\text{RMSE} = \sqrt{\frac{1}{3\text{MN}} \left(\sum_{c} \sum_{i} \sum_{j} (I_{\text{cs}}(i, j) - I_{\text{cr}}(i, j))^2 \right)} \tag{36}$$

RMSE is the positive component of the square root of the MSE. For a gray image, MSE is calculated for a single channel and RMSE is computed by taking the square root of the MSE.

7.4 Standard Deviation

Standard deviation (SD) measures the quality of a smoothed image from its original input image. It is mainly affected by the presence of noise in images, and its value is accurate without any noise. A higher standard deviation indicates greater image quality; i.e., an image with high contrast value has a higher standard deviation than a low contrast image. The standard deviation of an RGB image is given below.

$$\sigma = \sqrt{\frac{1}{3\mathrm{MN}}\left(\sum_{c}\sum_{i}\sum_{j}(I_{\mathrm{cr}}(i,j)-\mu_{\mathrm{c}})^2\right)} \tag{37}$$

σ^2 is computed by taking the average of the square of the difference between the pixel value and the corresponding channel mean. Here, μ_c is the mean value of channel c and I_{cr} represents the smoothed image.

7.5 Signal-to-Noise Ratio

Signal-to-noise ratio (SNR) is calculated over the source image and its smoothed image. It finds the signal to noise between source and final images, and the higher value indicates better image enhancement. Unit to represent signal-to-noise ratio is decibel. Signal-to-noise ratio of an RGB image is given below.

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_c\sum_i\sum_j[I_{\mathrm{cs}}(i,j)]^2}{\sum_c\sum_i\sum_j[I_{\mathrm{cs}}(i,j)-I_{\mathrm{cr}}(i,j)]^2} \tag{38}$$

A value greater than 0 dB indicates there is less noise than signal, and a value greater than 30 dB indicates good image quality. It generates very high value when both the source and final images are indistinguishable.

7.6 Peak Signal-to-Noise Ratio

Peak signal-to-noise ratio (PSNR) is also used to measure the quality of output image with respect to its input image. It is inversely proportional to RMSE. Like SNR, higher the value indicated better image enhancement. PSNR of a colored image is given below.

$$\mathrm{PSNR} = 20\log_{10}\frac{255}{\mathrm{RMSE}} \tag{39}$$

A value greater than 0 dB indicates there is less noise than signal, and a value greater than 30 dB indicates good image quality. Images of equally likely provide high peak signal-to-noise ratio.

7.7 Percentage Fit Error

Percentage fit error (PFE) is the percentage of the norm of the difference between an actual image and its denoised image with respect to the norm of the noisy image. If the smoothed image differs from the source image, the percentage of error also increases and images of equally likely provide a negligible error rate.

$$\mathrm{PFE} = \frac{\left(I_r - I_f\right)}{I_f} * 100 \tag{40}$$

Lower the PFE indicates better contrast enhancement. In the case of an RGB image, the norms are calculated over three channels, and then the percentage is computed with respect to the input image.

7.8 Structural Similarity Index Measure

Structural similarity index measure (SSIM) is a tool to evaluate the degree of similarity between the input image and its processed one that indirectly points out the amount of image distortion that happens during information transmission and processing. Contradict to PSNR, the quality of an image is based on the extent of correlation between the observable structures present in both images, whose range varies from 0 (completely divergent) to 1 (equal).

$$\mathrm{SSIM} = \left(\frac{2\mu_{\mathrm{cs}}\mu_{\mathrm{cr}}}{\mu_{\mathrm{cs}}^2 + \mu_{\mathrm{cr}}^2}\right) * \left(\frac{2\sigma_{\mathrm{cscr}}}{\nu_{\mathrm{cs}} + \nu_{\mathrm{cr}}}\right) = L_I * L_V \tag{41}$$

Here, $(\mu_{\mathrm{cs}}, \mu_{\mathrm{cr}})$, $(\nu_{\mathrm{cs}}, \nu_{\mathrm{cr}})$ stand for the mean and variance of the given input image and its enhanced image, respectively, where σ_{cscr}, L_I, and L_v symbolize standard deviation, local luminance, and local variance between them. Larger similarity index value indicates greater similarity between input and its produced one.

7.9 Correlation

Image correlation is a method to find a statistical relationship between two images, i.e., how an enhanced image is linearly dependent on its initial source image in

image smoothing applications. When a source image is completely identical to its processed image, it provides 1 and 0 generated when both the images are completely different from each other. Like SSIM, a higher correlation value indicates better image improvement and vice versa.

$$\text{CORR} = \frac{2C_{\text{cscr}}}{C_{\text{cs}} + C_{\text{cr}}} \tag{42}$$

$$C_{\text{cscr}} = \sum_{c} \sum_{i} \sum_{j} [I_{\text{cs}}(i, j) * I_{\text{cr}}(i, j)] \tag{43}$$

$$C_{\text{cs}} = \sum_{c} \sum_{i} \sum_{j} [I_{\text{cs}}(i, j)]^2 \text{ and } C_{\text{cr}} = \sum_{c} \sum_{i} \sum_{j} [I_{\text{cr}}(i, j)]^2 \tag{44}$$

Infrastructure is used for the analysis of both C-OpenCV and C-OpenCL implementations as described in Table 1. Here, 2D-DFT extracts the frequency components of a source image and after filtration, 2D inverse DFT converts back the image from its frequency domain to the spatial domain. Different noisy images and their smoothed images using the above low-pass filters with various cutoff frequencies are shown in Figs. 17 and 18.

Tables 2, 3, and 4 show, as the cutoff frequency D_{f} increased, there was remarkable reduction in MAE, RMSE, and PFE while dramatically improvement in entropy, SD, SNR, PSNR, SSIM, and correlation values, and these metrics were same up to three decimal points for CPU and GPU implementations. Here, time consumption by C-OpenCV implementation comprised the total time required to create a filter plus applying the filter on the transformed image and transformed it back to the spatial domain for display purposes. Total C-OpenCL code execution time included transferring the image data from host memory to device memory, transforming it from spatial to the frequency domain, creation of filters to the inverse transformation of filtered data, and transforming the data from device memory to host memory. All three channels of an RGB image were taken into consideration for analysis

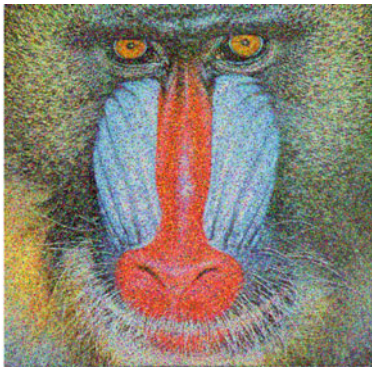
Table 1 Infrastructure specification

Components	Specification
Processor	Intel® Xeon® E3-1200 v5
Memory	DDR4 8 GB 2133 MHz
Storage	HDD 1 TB
Graphics processor (integrated)	Intel HD Graphics P530
Graphics processor (dedicated)	NVIDIA GTX 1050 Ti
Packages	C, OpenCV, OpenCL
Image	Size = 256 × 256, bit depth = 24.

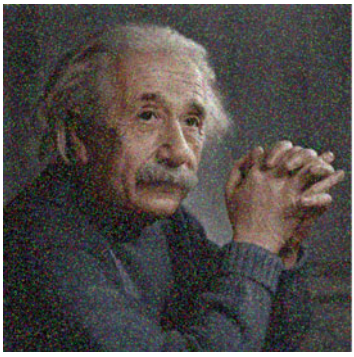
(a) Gaussian noise, Mean (μ) = 0.109, Variance (σ2) = 0.0124.

(b) Local variance noise, Mean (μ) = 0, Window Size (k) = 5

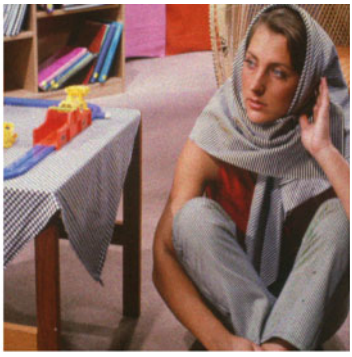
(c) Uniform noise, Range (A = 0, B = 1), Noise Scale Factor (S) = 0.3



(d) Salt and Pepper noise, ($P_{(A=0)}$ = 0.5, $P_{(B=255)}$ = 0.5), 20% of total pixels



(e) Impulse noise, Range (A = 0, B = 255), 20% of total pixels



(f) Poisson noise, Dependent on image (I). Pixel Scale Factor (S) = e^{12}

Fig. 17 Generated noisy images using various noise models with their specified parameters

(g) Speckle noise, Mean (μ) = 0, Variance (σ^2) = 0.04

(h) Rayleigh noise, Range (A = 0, B = 1), Noise Scale Factor (S) = 0.3

(i) Gamma noise, Range (A = 2, B = 5), Noise Scale Factor (S) = 0.1

(j) Exponential noise, A =1, Noise Scale Factor (S) = 0.2

Fig. 17 (continued)

purposes, and the time specifies denoising of all three channels of an RGB image in the frequency domain.

Table 2 shows ideal low-pass filter code execution time of OpenCL kernel on Intel, and NVIDIA GPU environments are approximately four times and ten times lower than its CPU implementation using C and OpenCV on CPU. Image enhancement completion time at $D_f = 50$ is lower for all kinds of noisy images, and it increases as with cutoff frequency for both of its CPU and GPU implementations. In contrast to that, there is also a slight improvement in entropy, SNR, PSNR, and correlation values, but those values are not up to the mark, which makes the ideal low-pass filter unsuitable for most of the practical applications.

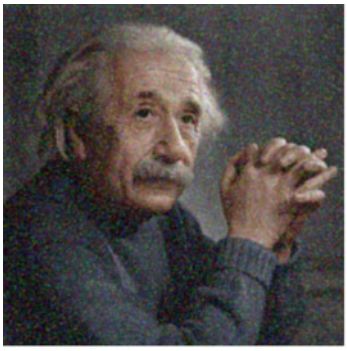
A Gaussian low-pass filter rendered better accuracy than an ideal low-pass filter, but its computation induces an extra bit of time compared to an ideal filter for both of its OpenCV and OpenCL implementations. By using this filter, we bring

(a) Ideal filter, Cutoff Frequency (D_f) = 70

(b) Ideal filter, Cutoff Frequency (D_f) = 90

(c) Ideal filter, Cutoff Frequency (D_f) = 50

(d) Gaussian filter, Cutoff Frequency (D_f) = 50



(e) Gaussian filter, Cutoff Frequency (D_f) = 70

(f) Gaussian filter, Cutoff Frequency (D_f) = 90

Fig. 18 Smoothed images generated from the respective noisy images using ideal, Gaussian, and Butterworth filters in reference to their cutoff frequency and order of the filter

(g) Butterworth filter, Cutoff Frequency (D_f) = 50, Order (n) = 1

(h) Butterworth filter, Cutoff Frequency (D_f) = 70, Order (n) = 1.

(i) Butterworth filter, Cutoff Frequency (D_f) = 90, Order (n) = 2

(j) Butterworth filter, Cutoff Frequency (D_f) = 90, Order (n) = 2.

Fig. 18 (continued)

down the RMSE values dramatically while showing remarkable growth in its PSNR, SSIM, and correlation metrics as shown in Table 3. Apart from that, this filter at any cutoff frequency, yields better compression with lower standard deviation than the ideal filter, which means less amount of information is needed to code those processed images, as well as the relationship between the neighborhood pixels is fairly maintained, compared to the filtered images from the ideal filter.

Table 4 shows Butterworth filter with order 1 performs better than ideal and Gaussian filters at a particular cutoff frequency and its performance is degraded with an increase in the order of this filter; i.e., a Butterworth filter of order two is performed worse than a Gaussian filter at a given D_f as shown in Tables 3 and 4. However, the computation time of a Butterworth filter is marginally lower than the Gaussian filter but higher than the ideal filter. This filter allows some of the frequencies after the cutoff frequency due to smoothness of its transition curve from pass band to stop

Table 2 Performance metrics of an ideal filter in C-OpenCV and C-OpenCL implementations

C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (sec) OpenCV	Time (sec) OpenCL-HD Gh. P530	Time (sec) OpenCL-GTX 1050 Ti
($D_f = 50$)	17.a	7.59	27.3	31.6	56.5	13.75	18.13	17.6	0.72	0.954	0.181	0.054	0.018
	17.b	7.66	16.7	24.3	66.5	16.12	20.41	6.23	0.88	0.937	0.172	0.053	0.018
	17.c	7.72	50.8	53.8	56.7	06.85	13.51	50.6	0.62	0.984	0.186	0.055	0.019
	17.d	7.38	22.8	29.0	40.6	13.57	18.88	6.15	0.54	0.852	0.174	0.053	0.018
	17.e	6.89	16.9	19.9	33.0	10.73	22.16	25.6	0.58	0.953	0.188	0.054	0.019
	17.f	7.65	11.8	18.0	51.0	16.81	23.02	3.11	0.71	0.943	0.197	0.055	0.018
	17.g	7.11	10.2	15.9	35.9	13.84	24.06	4.56	0.63	0.912	0.171	0.054	0.018
	17.h	7.34	68.8	70.8	48.8	05.21	11.13	68.9	0.77	0.957	0.183	0.054	0.018
	17.i	7.37	63.5	65.5	44.2	04.39	11.80	73.5	0.66	0.955	0.188	0.053	0.018
	17.j	7.72	51.7	55.9	57.6	6.46	13.18	55.4	0.61	0.958	0.175	0.055	0.018
($D_f = 70$)	17.a	7.60	26.7	30.6	57.3	14.03	18.41	17.5	0.73	0.962	0.180	0.056	0.019
	17.b	7.65	14.3	20.9	68.1	17.41	21.70	5.66	0.90	0.957	0.187	0.056	0.020
	17.c	7.73	50.7	53.7	57.2	06.87	13.53	50.5	0.62	0.987	0.195	0.055	0.019
	17.d	7.46	23.0	29.3	42.6	13.48	18.79	6.09	0.55	0.844	0.189	0.055	0.019
	17.e	6.98	17.2	21.1	34.1	10.21	21.66	25.6	0.48	0.931	0.192	0.054	0.020
	17.f	7.65	10.8	16.9	51.4	17.36	23.57	3.08	0.74	0.950	0.199	0.055	0.019
	17.g	7.15	10.2	15.1	37.1	14.33	24.54	3.91	0.62	0.922	0.202	0.056	0.019
	17.h	7.34	68.8	70.6	49.5	05.22	11.14	68.9	0.77	0.959	0.203	0.055	0.019
	17.i	7.40	63.4	65.4	44.9	04.41	11.82	73.5	0.66	0.960	0.205	0.055	0.019
	17.j	7.74	51.7	55.8	58.6	06.47	13.19	55.3	0.62	0.961	0.191	0.054	0.020

(continued)

Table 2 (continued)

C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (sec) OpenCV	Time (sec) OpenCL-HD Gh. P530	Time (sec) OpenCL-GTX 1050 Ti
($D_f = 90$)	17.a	7.62	26.4	30.2	57.9	14.16	18.53	17.5	0.72	0.965	0.205	0.053	0.019
	17.b	7.64	12.8	18.8	69.2	18.33	22.62	5.14	0.92	0.967	0.214	0.055	0.020
	17.c	7.74	50.8	53.8	57.5	06.86	13.52	50.6	0.62	0.988	0.208	0.056	0.020
	17.d	7.53	23.7	30.3	44.9	13.18	18.49	6.07	0.56	0.829	0.207	0.056	0.020
	17.e	7.06	17.9	22.8	35.4	9.53	20.98	25.7	0.40	0.902	0.208	0.055	0.020
	17.f	7.65	10.2	16.2	51.7	17.72	23.94	3.07	0.76	0.954	0.225	0.054	0.019
	17.g	7.19	10.6	15.1	38.3	14.32	23.53	3.48	0.59	0.923	0.220	0.056	0.020
	17.h	7.34	68.8	70.7	50.0	05.21	11.14	68.9	0.77	0.956	0.223	0.056	0.020
	17.i	7.42	63.4	65.4	45.5	04.41	11.82	73.5	0.67	0.961	0.215	0.056	0.020
	17.j	7.74	51.6	55.9	59.6	06.45	13.18	55.3	0.62	0.960	0.219	0.055	0.020

Table 3 Performance metrics of a Gaussian filter in C-OpenCV and C-OpenCL implementations

C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (s) OpenCV	Time (s) OpenCL-HD Gh. P530	Time (s) OpenCL-GTX 1050Ti
($D_f = 50$)	17.a	7.51	26.8	30.5	55.3	14.08	18.45	17.6	0.77	0.967	0.188	0.055	0.019
	17.b	7.73	13.6	20.5	65.0	17.61	21.89	5.63	0.91	0.959	0.183	0.055	0.020
	17.c	7.70	50.8	53.5	56.0	06.90	13.56	50.6	0.63	0.989	0.201	0.056	0.020
	17.d	7.36	21.6	27.4	40.0	14.07	19.38	6.10	0.59	0.876	0.194	0.054	0.020
	17.e	6.86	16.7	19.7	32.6	10.77	22.22	25.6	0.58	0.957	0.199	0.056	0.020
	17.f	7.61	10.5	16.7	50.0	17.53	23.69	3.03	0.75	0.951	0.185	0.056	0.020
	17.g	7.07	8.57	13.7	35.0	15.18	25.36	3.92	0.71	0.936	0.203	0.053	0.019
	17.h	7.35	69.0	70.7	48.0	05.22	11.14	69.0	0.78	0.965	0.196	0.055	0.020
	17.i	7.33	63.5	65.2	43.5	04.44	11.84	73.5	0.68	0.967	0.190	0.056	0.020
	17.j	7.70	51.7	55.2	56.8	06.56	13.28	55.4	0.63	0.970	0.191	0.055	0.020
($D_f = 70$)	17.a	7.56	26.3	29.6	56.4	14.33	18.71	17.5	0.77	0.973	0.223	0.054	0.020
	17.b	7.71	11.7	17.7	66.8	18.85	23.14	5.09	0.93	0.972	0.208	0.056	0.020
	17.c	7.71	50.8	53.5	56.6	06.90	13.56	50.6	0.63	0.990	0.211	0.056	0.022
	17.d	7.44	21.3	27.3	42.1	14.10	19.41	6.00	0.61	0.871	0.213	0.056	0.021
	17.e	6.96	16.8	20.9	33.9	10.25	21.70	25.6	0.48	0.933	0.217	0.054	0.020
	17.f	7.63	9.34	15.0	50.7	18.39	24.60	2.72	0.79	0.961	0.207	0.053	0.020
	17.g	7.12	8.78	13.1	36.4	15.57	25.77	3.33	0.68	0.941	0.216	0.055	0.020
	17.h	7.35	69.0	70.6	49.0	05.23	11.15	69.0	0.78	0.965	0.205	0.055	0.020
	17.i	7.37	63.5	65.1	44.4	04.45	11.85	73.5	0.69	0.970	0.218	0.055	0.020
	17.j	7.72	51.7	55.2	58.0	06.57	13.29	55.3	0.64	0.971	0.209	0.056	0.020

(continued)

Table 3 (continued)

C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (s) OpenCV	Time (s) OpenCL-HD Gh. P530	Time (s) OpenCL-GTX 1050Ti
($D_f = 90$)	17.a	7.59	26.1	29.2	57.2	14.43	18.81	17.5	0.75	0.975	0.227	0.056	0.021
	17.b	7.79	10.8	16.5	68.0	19.50	23.78	4.82	0.94	0.977	0.225	0.056	0.021
	17.c	7.73	50.8	53.6	57.0	06.89	13.54	50.6	0.62	0.989	0.220	0.055	0.022
	17.d	7.51	21.5	28.2	44.4	13.81	19.12	5.94	0.62	0.856	0.217	0.056	0.021
	17.e	7.05	17.0	22.7	35.2	09.56	21.01	25.6	0.41	0.903	0.224	0.056	0.022
	17.f	7.64	8.36	13.2	51.2	19.49	25.68	2.26	0.84	0.970	0.224	0.054	0.021
	17.g	7.16	9.56	13.7	37.5	15.21	25.41	2.99	0.64	0.936	0.226	0.055	0.020
	17.h	7.35	68.9	70.7	49.7	05.22	11.14	68.9	0.77	0.961	0.223	0.056	0.021
	17.i	7.40	63.4	65.2	45.0	04.44	11.84	73.5	0.68	0.968	0.225	0.056	0.020
	17.j	7.75	51.6	55.5	59.0	06.52	13.25	55.3	0.63	0.967	0.221	0.055	0.021

Table 4 Performance metrics of a Butterworth filter in C-OpenCV and C-OpenCL implementations

Order and C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (s) OpenCV	Time (s) OpenCL-HD Gh. P530	Time (s) OpenCL-GTX 1050 Ti
(n = 1, D_f = 50)	17.a	7.51	26.8	30.6	54.4	14.05	18.42	17.6	0.77	0.967	0.192	0.055	0.018
	17.b	7.75	14.3	21.0	63.6	17.40	21.69	5.86	0.91	0.957	0.189	0.055	0.019
	17.c	7.69	50.8	53.5	55.4	06.91	13.56	50.6	0.63	0.988	0.191	0.054	0.018
	17.d	7.33	21.3	26.9	39.4	14.22	19.53	6.15	0.61	0.884	0.185	0.055	0.018
	17.e	6.84	16.8	19.8	32.3	10.72	22.16	25.6	0.57	0.956	0.180	0.056	0.018
	17.f	7.60	10.4	16.0	49.3	17.80	24.02	2.77	0.77	0.955	0.188	0.055	0.018
	17.g	7.04	8.71	13.8	34.3	15.13	25.34	4.06	0.71	0.936	0.179	0.055	0.019
	17.h	7.36	69.0	70.7	47.3	05.21	11.13	69.0	0.78	0.965	0.182	0.055	0.018
	17.i	7.32	63.5	65.2	43.0	04.43	11.84	73.5	0.68	0.967	0.184	0.056	0.019
	17.j	7.69	51.7	55.2	56.1	06.56	13.28	55.4	0.64	0.970	0.180	0.055	0.018
(n = 1, D_f = 70)	17.a	7.55	26.4	29.7	55.7	14.31	18.68	17.6	0.77	0.973	0.195	0.056	0.019
	17.b	7.73	12.3	18.4	65.6	18.51	22.81	5.27	0.93	0.969	0.208	0.055	0.019
	17.c	7.70	50.8	53.5	56.2	06.90	23.56	50.6	0.63	0.990	0.207	0.056	0.019
	17.d	7.41	20.8	26.9	41.2	14.23	19.54	6.01	0.63	0.878	0.198	0.053	0.018
	17.e	6.93	16.7	20.9	33.6	10.28	21.73	25.6	0.48	0.935	0.206	0.055	0.019
	17.f	7.62	9.18	14.3	50.2	18.81	25.02	2.51	0.82	0.965	0.201	0.055	0.019
	17.g	7.09	8.91	13.2	35.7	15.49	25.70	3.44	0.68	0.940	0.204	0.055	0.019
	17.h	7.36	69.0	70.6	48.4	05.22	11.15	68.9	0.79	0.965	0.202	0.054	0.018
	17.i	7.36	63.5	65.1	43.9	04.44	11.85	73.5	0.69	0.969	0.198	0.055	0.019
	17.j	7.71	51.7	55.2	57.4	06.57	13.30	55.4	0.64	0.971	0.197	0.053	0.019

(continued)

Table 4 (continued)

Order and C. Freq.	Images	Entropy	MAE	RMSE	SD	SNR	PSNR	PFE	SSIM	Correlation	Time (s) OpenCV	Time (s) OpenCL-HD Gh. P530	Time (s) OpenCL-GTX 1050 Ti
($n = 2$, $D_f = 90$)	17.a	7.59	26.2	29.4	57.4	14.39	18.76	17.5	0.76	0.973	0.223	0.056	0.020
	17.b	7.66	11.2	17.1	68.4	19.16	23.45	4.92	0.94	0.974	0.213	0.056	0.020
	17.c	7.72	50.8	53.6	57.2	06.89	13.54	50.6	0.63	0.990	0.216	0.055	0.020
	17.d	7.49	21.5	27.8	43.5	13.94	19.25	5.95	0.61	0.862	0.210	0.055	0.019
	17.e	7.01	16.9	21.7	34.7	09.95	21.40	25.6	0.45	0.920	0.225	0.055	0.020
	17.f	7.64	9.02	14.5	51.4	18.67	24.88	2.68	0.80	0.963	0.214	0.055	0.019
	17.g	7.16	9.16	13.3	37.4	15.41	25.62	3.15	0.66	0.939	0.215	0.054	0.020
	17.h	7.33	68.9	70.6	49.7	05.23	11.15	68.9	0.78	0.963	0.224	0.056	0.019
	17.i	7.39	63.4	65.2	45.0	04.44	11.85	73.5	0.68	0.969	0.219	0.056	0.020
	17.j	7.74	51.7	55.3	58.9	06.55	13.27	55.3	0.63	0.970	0.220	0.053	0.020

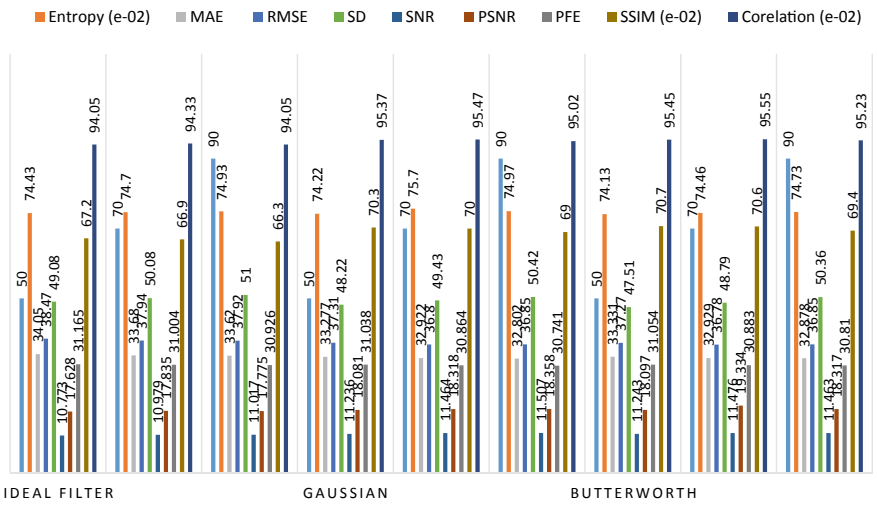


Fig. 19 Performance metrics of ideal, Gaussian, and Butterworth filters with respect to their cutoff frequencies at 50, 70, and 90

band, which is large at its lowest order and decreases with the increase of its order, and that makes the filter not reliable as Gaussian filter in some of the image processing and computer vision practices.

Figure 19 shows that a Butterworth filter (order = 1, Df = 70) performed well compared to an ideal and Gaussian filter at 70 and 90, but its performance gets worse with the increase of its order as shown at (Df = 90) for order 2. At the same time, entropy and standard deviation are increased as with cutoff frequencies but declined gradually for Gaussian and Butterworth filters, which means a change of the filter brings better encoding of information with the relationship between the pixels highly maintained. In addition to that, while maintaining those metrics, the average GPU computation time on Intel and NVIDIA GPUs is reduced by one-fourth and one-tenth of CPU computation time due to the deployment of OpenCL kernels for those above filters as shown in Fig. 20. Instead, CPU execution increases with respect to the cutoff frequency, and there is a tiny change in execution time on GPU. As we have discussed earlier, the MATLAB implementations of both the Gaussian and Butterworth filters consume 0.09981 and 0.079 s for enhancing a grayscale 128 × 128 pixel image on a four-core CPU [8], whereas enhancing a 480 × 320 pixel size grayscale image by those filters requires 0.8644 and 0.7020 s, respectively [6]. Similarly, MATLAB implementation of the Gaussian approximation kernel using cosine filter and Monte Carlo sampling requires 17.21 s to enhance a 512 × 512 RGB image [9], but the approximated Gaussian filter using multilayer convolution of binomial filter needs only 90 s to enrich a 4096 × 1712 pixel image on 2688 cores NVIDIA GPU [10]. The other two implementations such as enhancing an HD image using the 5 × 5 Gaussian filter in the spatial domain take 12.5 s on an Intel Core i5 7th generation system [12], while its cascade implementation fetches 49.65 s CPU time to

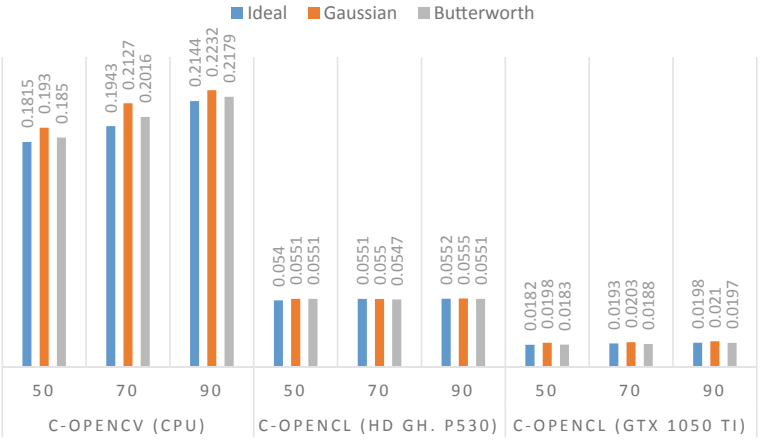


Fig. 20 Time consumption by C-OpenCV and C-OpenCL implementations of frequency domain filters with respect to the cutoff frequencies at 50, 70, and 90

enhance a 768 × 512 gray image [13]. Butterworth filter optimization using artificial bee colony algorithms and multi-scale Retinex algorithm built on Butterworth filter gains high accuracies, but the execution times are compromised due to slowness of the evolutionary algorithm and the process of lengthy sequential execution, respectively [15, 16]. However, the overall computational cost for processing an RGB image of size 256 × 256 pixels by the OpenCL Gaussian and Butterworth kernels is (0.0552, 0.0549 s) and (0.0203, 0.0189 s) on Intel and NVIDIA GPUs, respectively, which are quite less compared to the previous implementations. So, the Gaussian OpenCL kernel is the preferable one among these filters as it provides better results in terms of accuracy and time consumption. Platform independent nature of these OpenCL kernels makes them widely portable on heterogeneous devices like ARM, FPGA, and GPU.

8 Conclusion

Image denoisification is one of the image enhancement techniques used to strengthen the visual quality of an image by removing unwanted signals from the images that are introduced due to various factors, such as sensor temperature, light intensity, target environment, erroneous electric instruments, and interference in the communication channels. In this paper, we have not only presented the noise model and various types of noises that are affecting image degradation but also explained frequency domain filtering and mostly used separable and non-separable filters for image denoisification. OpenCL kernels of the above low-pass filers are developed for faster enhancement, which has a wide range of applications, especially in real-time implementation. From the performance analysis, we have concluded that Gaussian and first-order

Butterworth low-pass filters supply better results than other techniques in terms of time consumption and accuracy. At last, more experimentation has to be done later to improve the accuracy and reduce time utilization that supports the needs according to current demands.

References

1. Gonzalez RC, Woods RE (2008) Introduction. In: Horton MJ (ed) Digital image processing. Prentice-Hall, Upper Saddle River, NJ, pp 1–33
2. Jayaraman S, Esakkirajan S, Veerakumar T (2015) Introduction to image processing system. In Jha S (ed) Digital image processing, vol 7. Tata McGraw Hill Education, New Delhi, pp 1–46
3. Meena SK, Potnis A, Mishra M (2017) Review and application of different contrast enhancement technique on various images. In: First international conference on electronics, materials engineering and nano-technology (IEMENTech), pp 1–6
4. Gonzalez RC, Woods RE (2008) Image enhancement in the spatial domain. In Horton MJ (Ed) Digital image processing. Prentice-Hall, Upper Saddle River, NJ, pp 75–146
5. Jayaraman S, Esakkirajan S, Veerakumar T (2015) Image enhancement. In Jha S (ed) Digital image processing, vol 7. Tata McGraw Hill Education, New Delhi, pp 243–323
6. Dyre S, Sumathi CP (2014) Hybrid approach to enhancing fingerprint images using filters in the frequency domain. In: 2014 IEEE international conference on computational intelligence and computing research. IEEE, pp 1–6
7. Chinnasamy G, Vanitha S (2015) Implementation and comparison of various filters for the removal of fractional brownian motion noise in brain MRI images. Int J Trend Eng Technol (IJTET), ISSN: 2349-9303
8. Shukla A, Singh RK (2015) Performance analysis of frequency domain filters for noise reduction. e-J Sci Technol (e-JST) 9(5):167–178
9. Ghosh S, Chaudhury KN (2016) Fast bilateral filtering of vector-valued images. In: 2016 IEEE international conference on image processing (ICIP), pp. 1823–1827. IEEE
10. Yano T, Kuroki Y (2016) Fast implementation of Gaussian filter by parallel processing of binominal filter. In 2016 international symposium on intelligent signal processing and communication systems (ISPACS). IEEE, pp 1–5
11. Nair P, Popli A, Chaudhury KN (2017) A fast approximation of the bilateral filter using the discrete Fourier transform. Image Proces Line 7:115–130
12. Preethi K, Vishvaksenan KS (2018) Gaussian filtering implementation and performance analysis on GPU. In 2018 international conference on inventive research in computing applications (ICIRCA), pp 936–939. IEEE
13. Rakhshanfar M, Amer MA (2019) Efficient cascading of multi-domain image Gaussian noise filters. J Real-Time Image Proces 1–13
14. Chen B, Liu Q, Sun X, Li X, Shu H (2014) Removing Gaussian noise for colour images by quaternion representation and optimization of weights in non-local means filter. IET Image Proc 8(10):591–600
15. Fadloullah I, Mechaqrane A, Ahaitouf A (2017) Butterworth low pass filter design using evolutionary algorithm. In: 2017 international conference on wireless technologies, embedded and intelligent systems (WITS), pp 1–6. IEEE
16. Fan T, Li C, Ma X, Chen Z, Zhang X, Chen L (2017) An improved single image defogging method based on Retinex. In: 2017 2nd international conference on image, vision and computing (ICIVC), pp 410–413. IEEE
17. Boncelet C (2009) Image noise models. In: Bovik A (ed) The essential guide to image processing. Elsevier, Burlington, pp 143–167

18. Al-Ghaib H, Adhami R (2014) On the digital image additive white noise Gaussian noise estimation. In International conference on industrial automation, information and communications technology (IAICT), pp 90–96
19. Lopez-Rubio E (2009) Restoration of images corrupted by Gaussian and uniform impulsive noise. Pattern Recogn 43:1835–1846
20. Zubair AR, Fakolujo OA (2013) Development of statistics and convolution as tools for image noise suppression: statistical performance analysis of Spatial filters. Afr J Comput ICT 6(5):53–66
21. Deivalakshmi S, Sarath S, Palanisamy P (2011) Detection and removal of Salt and Pepper noise in images by improved median filter. In: IEEE conference on recent advances in intelligent computational systems (RAICS), pp 363–368
22. Harikiran J, Saichandana B, Divakar B (2010) Impulse noise removal in digital images. Int J Comput Appl 10(8):39–42
23. Rodrigues I, Sanches J, Bioucas-Dias J (2008) Denoising of medical images corrupted by Poisson noise. In: Fifteenth international conference on image processing (ICIP), pp 1756–1759
24. Benzarti F, Amiri H (2012) Speckle noise reduction in medical ultrasound images. Int J Comput Sci Iss 9(2):187–194
25. Boyat AK, Joshi BK (2015) A review paper: noise models in digital image processing. Signal & Image Process: Int J (SIPIJ) 6(2):63–75
26. Huang YM, Yan HY, Wen YW, Yang X (2017) Rank minimization with applications to image noise removal. Inf Sci 429:147–163
27. Schettini R, Corchs S (2010) Underwater image processing: state of the art of restoration and image enhancement methods. EURASIP J Adv Signal Process 2010:1–14
28. Buades A, Coll B, Morel JM (2004) On image denoising methods. Centre de Mathematiques et de Leurs Applications, Cachan, France. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.81&rep=rep1&type=pdf. Accessed 20 Jan 2018
29. Das A (2015) Interpretation and processing of image in frequency domain. Guide to signals and patterns in image processing. Springer, Cham, pp 93–147
30. Burger W, Burge MJ (2008) The discrete fourier transform in 2D. Digital image processing: an algorithmic introduction using Java. Springer, London, pp 343–366
31. Burger W, Burge MJ (2008) The discrete cosine transform (DCT). Digital image processing: an algorithmic introduction using java. Springer, London, pp 367–373
32. Marques O (2011) Frequency domain filtering. Practical Image and Video Processing Using MATLAB. Wiley, Hoboken, pp 235–264
33. Shaikh MS, Choudhry A, Wadhwani R (2016) Analysis of digital image filters in frequency domain. Int J Comput Appl 140(6):12–19
34. Fu-jun H, Han K (2012) The application of low-pass filtering to pretreatment in thermal wave NDT. In: International conference on measurement, information and control (MIC), pp 590–594
35. Xu J, Ling Y, Zheng X (2015) Forensics detection of Gaussian low-pass filtering in digital images. In: Eighth international congress on image and signal processing (CISP), pp 819–823
36. Lyra M, Ploussi A, Rouchota M, Synefia S (2014) Filters in 2D and 3D cardiac SPECT image processing. Cardiol Res Pract 2014(2014):1–11
37. Banger R, Bhattacharyya K (2013) Hello OpenCL. In: Dsouza W, Pandey K, Colaco K (eds) OpenCL programming by example. Packt Publishing, Birmingham, pp 7–34
38. Banger R, Bhattacharyya K (2013) OpenCL architecture. In: Dsouza W, Pandey K, Colaco K (eds) OpenCL programming by example. Packt Publishing, Birmingham, pp 35–58
39. Kirsch N (2015, August 17) Intel HD Graphics 530 has 24 execution units—Intel gen9 graphics architecture detailed. Retrieved from https://www.legitreviews.com/intel-hd-graphics-530-has-24-execution-units-intel-gen9-graphics-architecture_170869
40. Mann P (2016) Review: EVGA GeForce GTX 1050 Ti SC gaming. http://hexus.nettech/reviews/graphics/98329-evga-geforce-gtx-1050-ti-sc-gaming/

41. Banger R, Bhattacharyya K (2013) OpenCL program and kernel objects. In: Dsouza W, Pandey K, Colaco K (eds) OpenCL programming by example. Packt Publishing, Birmingham, pp 109–136
42. Naidu VPS, Raol JR (2008) Pixel-level image fusion using wavelets and principal component analysis. Def Sci J 58(3):338–352

FPGA Implementation of Radix-4-Based Two-Dimensional FFT with and Without Pipelining Using Efficient Data Reordering Scheme

Ashutosh Kumar, Shashank Gavel, and Ajay Singh Raghuvanshi

Abstract The multi-dimensional (multi-D) discrete Fourier transform (DFT) has found a major application in analyzing the field of signal and image processing. The reconstruction of image plays a primary role in the field of signal processing for exploring different fields such as remote sensing, photography, industrial inspection, robotics, and medical diagnosis. The advancement in DFTs has given birth to fast Fourier transform (FFT) by making the system much faster in both 2D and multi-D. In this article, we present an efficient 64 × 64 point 2D FFT architecture with a radix-4 block structure. The designed structure is a parallel architecture that takes less computation time. The proposed design is implemented using the Vitex-7 FPGA tool with supplementary hardware.

Keywords 2D FFT · Radix-4 · Virtex-7 · Slices · Pipeline

1 Introduction

Digital signal processing (DSP) has become an important segment for the current generation. The use of DSP in different applications has provided a platform to various researchers for analyzing different fields. This field has majorly found applications in robotics, industrial inspections, television systems, etc. Among them, many of the applications of DSP are only possible due to the advancement in VLSI technology. VLSI has provided evolving hardware solutions that have made the field of DSP mobile. Discrete Fourier transform (DFT) plays a major role in designing

A. Kumar (✉) · S. Gavel · A. S. Raghuvanshi
Department of Electronics and Communication Engineering, National Institute of Technology, Raipur, Raipur, Chhattisgarh, India
e-mail: ashutoshkumrr@gmail.com

S. Gavel
e-mail: sgavel.phd2016.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_53

a communication system that utilizes DSP. Fast Fourier transform (FFT) is an efficient structure of DFT that provides better performance by taking less number of operations. In general, the complexity of DFT is reduced by utilizing FFT [1].

FFT has found major applications in the field of image and signal processing due to its less complex structure and processing. One of the applications is the reorganization of the image from raw data by using 2D FFT. This necessitates systematic implementations to practical applications. Since to support the size of an image, the processor requires more memory. So, an efficient architecture is needed in order to optimize the memory as well as to achieve better performance.

To implement FFT, many hardware and software solutions have been utilized [2,3]. In this article, we present an efficient FPGA-based 2D FFT architecture. We have used the structured behavior of FFT to create a parallel-pipelined model of architecture that is executed in FPGA. 2D FFT is evaluated from $2\times M$ 1D FFT. Thus, the performance of 2D FFT is directly influenced by 1D FFT [4].

To implement 64 $\times$ 64 FFT architecture, we have used radix-4 algorithm that is built on a novel radix-4 butterfly unit [5]. 64 $\times$ 64 point FFT is computed using a cascade of two radix-4^3 blocks. The reordering of outputs is not needed as it is already ordered in our design that improves memory optimization [6]. The main contributions of this study are summarized below:

- 2D FFT architecture built on radix-4 algorithm with and without pipelining with maximum frequency using less number of slices and registers.
- A systematic output reordering technique implemented using six-bit up-counter.
- Memory curtailment for twiddle factors.
- FPGA implementation of 2D FFT.
- Collation of our work with other works shows significant improvement in terms of frequency, slices and registers.

The remaining paper is organized as shown. Section 2 provides brief idea of 1D and 2D FFT algorithms. A short study of the available literature is also described. Section 3 presents the architecture. Section 4 presents VLSI execution and results. Section 5 presents the conclusion.

2 Related Study and Framework

2.1 Related Study

There are many architectures and algorithms available on FFT. Pipeline-based architecture consumes more power and provides more area overhead [8,9]. Yin et al. [10] presents radix-2 real-valued FFT through rigorous formula deduction. Memory-based designs provide an advantage in terms of power consumption but require more area. Xia et al. [11] present several memory access architectures.

For 2D FFT also, different 2D architectures have been proposed. Chen and Prasanna [13] present a spiral software solution which provides high performance but also consumes more power. Hardware solutions are also available in [5]. A constructive architecture to calculate the 2D DFT for huge sized input is implemented in [14]. Hardware acceleration platform for reconstruction of the image is presented in [15]. Mixed radices are used for 1D FFT implementation. Yang et al. [16] present two-dimensional FFT where more datasets occupy off-chip in DRAM. The inner loop unrolling technique is employed to design two-dimensional FFT on FPGA. Kee et al. [4] present FFT in which M^2 intermediate memories are needed for the first phase, but in our work, M intermediate memories are needed. In 1D FFT itself, RAM requirement is less.

When designing FFT, output reordering is a big issue. With reordering, we produce output in natural order, i.e., in a reordered way. In parallel-pipelined FFT, the reordering of output is complex and is introduced in [17].

Even and odd outputs are saved using two buffers having size M. Rodríguez-Ramos et al. [18] present one memory of size M alternately for even and odd outputs. $3 \times M$ memories are needed for data reordering for feedback pipelined architectures having a multi-path delay.

2.2 Framework

The evolution of computationally systematic algorithm is made possible if we adopt a divide-and-conquer approach. Cooley–Tukey is the usual algorithm to calculate FFT [1]. Due to this algorithm, its complexity reduces from $O(M\log_2 M)$ to M^2. M-point sequence of $g(m)$ is given by-

$$G(r) = \sum_{m=0}^{M-1} g(m) W_m^{mr}, \quad r = 0, 1, 2, ..., M-1 \tag{1}$$

where W_M is the twiddle factor and is described as-

$$W_M = \mathrm{e}^{-2\pi i/M} \tag{2}$$

M-point inverse DFT is calculated as

$$g(m) = \frac{1}{M} \sum_{r=0}^{M-1} G(r) W_M^{-mr}, \quad m = 0, 1, 2, ..., M-1 \tag{3}$$

1D FFT can be used to calculate the 2D FFT. Using row–column deposition algorithm, a 1D FFT can be used to calculate the 2D FFT as shown below:

$$G(r_1, i_2) = \sum_{i_1=0}^{M-1} g(i_1, i_2) W_M^{r_1 i_1} \tag{4}$$

where $r_1 = 0, 1, 2, ..., M - 1$.

$$Q(r_1, r_2) = \sum_{i_2=0}^{M-1} G(r_1, i_2) W_M^{r_2 i_2} \tag{5}$$

where $r_2 = 0, 1, 2, ..., M - 1$.
Parallel radix-4 butterfly unit is employed in this study [7].

3 Proposed 2D FFT Architecture

This segment deals with radix-4^3 algorithm to implement the two-dimensional FFT. If input is $g(i_1, i_2)$, then 2D FFT for size 64×64 is expressed by-

$$q(r_1, r_2) = \sum_{i_2=0}^{63} \left[\sum_{i_1=0}^{63} g(i_1, i_2) W_{64}^{r_1 i_1} \right] W_{64}^{r_2 i_2} \tag{6}$$

where $r_1, r_2 = 0, 1, 2, ..., 63$.

Taking into account the inner block of summation

$$\sum_{i_1=0}^{63} g(i_1, i_2) W_{64}^{r_1 i_1}$$

The above summation is 64-point FFT which is calculated using radix-4^3 algorithm. Likewise, we can also compute the outer summation. So, cascade of two radix-4^3 blocks produces 2D FFT. 1D FFT is computed with the help of radix-4^3 architecture shown in [7]. Radix-4^3 consists of three levels of radix-4 butterfly units. Henceforward, radix-4^3 is denoted as R4^3 and radix-4 as R4.

R 4^3 algorithm is based on the signal flow graph [1]. The signal flow graph follows decimation in frequency pattern. The first phase consists of 16 radix-4 butterfly operations, the second phase employs 4 radix-4 operations, and the third phase consist of single radix-4 operation. $g(0)$, $g(1)$, $g(2)$, $g(3)$ are inputs of first R4 unit producing $G(0)$, $G(1)$, $G(2)$, $G(3)$ as outputs. A fully unrolled $R4^3$ architecture using a parallel R4 butterfly unit is our primary block [6]. The radix-4 block consists of four inputs, producing single output at a time depending on a two-bit control signal. Depending upon the two-bit control signal, i.e., mode select, outputs are produced in a specified order. In standard radix-4 operation, outputs are produced in a fixed sequence. The R4 butterfly component equation is shown in (7).

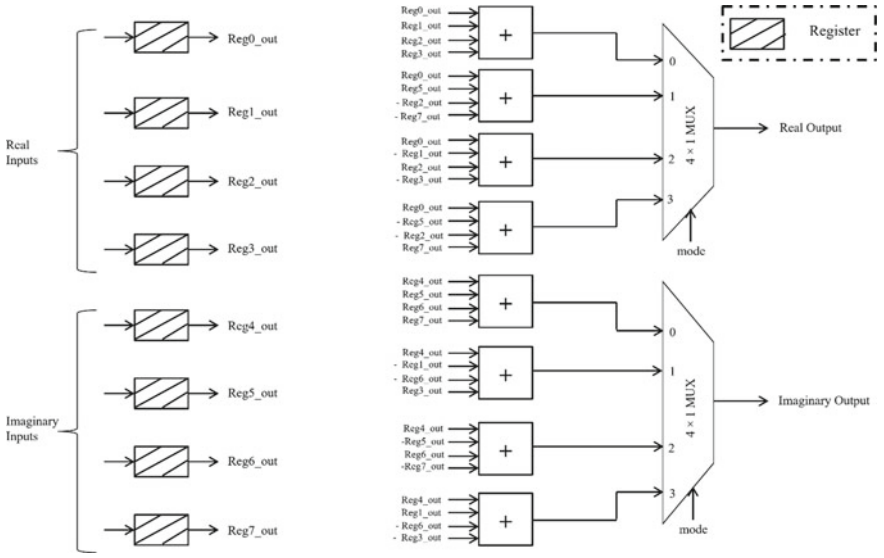


Fig. 1 Radix-4 butterfly architecture without pipelining

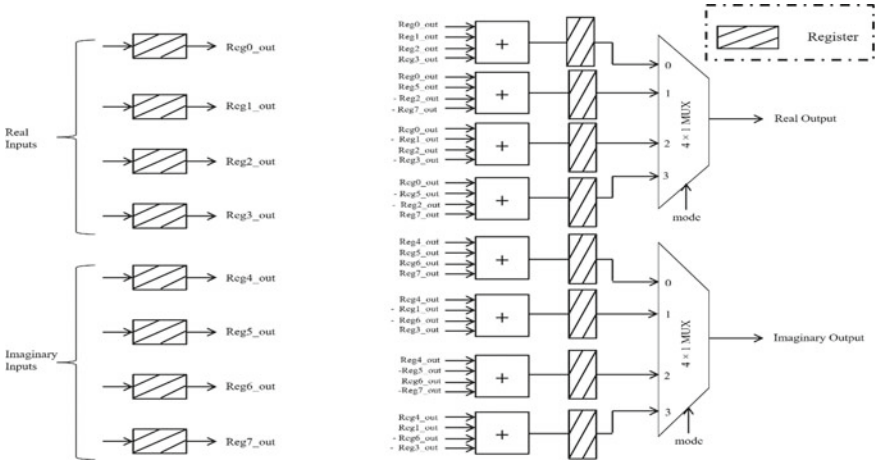


Fig. 2 Radix-4 butterfly architecture with pipelined adders

$$
\begin{aligned}
G(0) &= g(0) + g(1) + g(2) + g(3) \\
G(1) &= g(0) - jg(1) - [g(2) - jg(3)] \\
G(2) &= g(0) - g(1) + g(2) - g(3) \\
G(3) &= g(0) + jg(1) - [g(2) + jg(3)]
\end{aligned} \tag{7}
$$

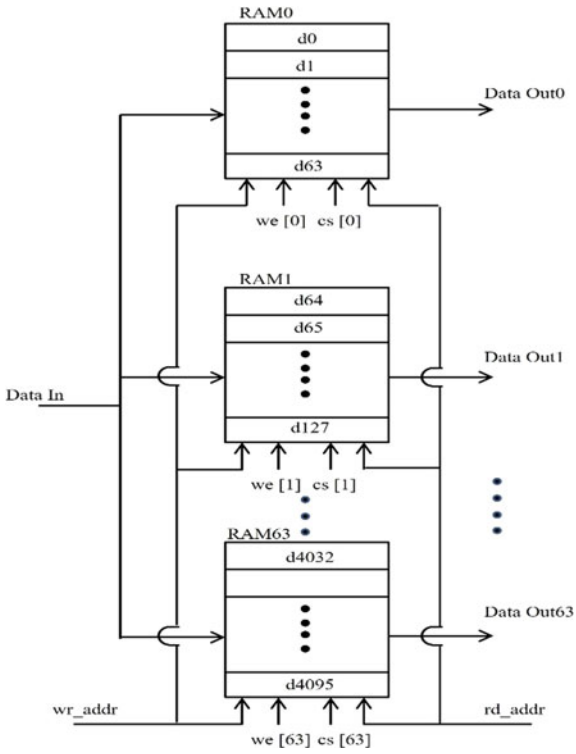


Fig. 3 Input memory

Figures 1 and 2 show the architecture of radix-4 butterfly without pipelining and with pipelined adders, respectively.

Outputs are produced in an ordered way using six-bit control signal which is implemented using a six-bit up-counter. Each phase is controlled by two-bit of up-counter which is shown as *mode* in Figs. 1 and 2 (Fig. 3).

The parallel unrolled architecture is shown in Figs. 4 and 5 for non-pipelined and pipelined architecture, respectively. The first phase is having 16 R4 blocks, the second phase is having 4 R4 blocks, and the third phase is having a single R4 block. W1 and W2 represent the twiddle factors of the first and second phase, respectively. The twiddle factors are not stored in memory; rather, it is generated and used directly. Wherever twiddle factor is having unity value, we do not need to use complex multipliers for that particular case. So, we will be using less number of multipliers. In each phase consisting of M twiddle factor values, $M/4$ will be having unity twiddle factor values, so no need for multipliers while doing complex multiplication. Suppose 64×64 input, which we have stored in memory, is described by matrix X, where 'p' will denote the row number and 'q' will denote the column number. In cycle 1, the mode is set to be 0 and the first column of X is provided as

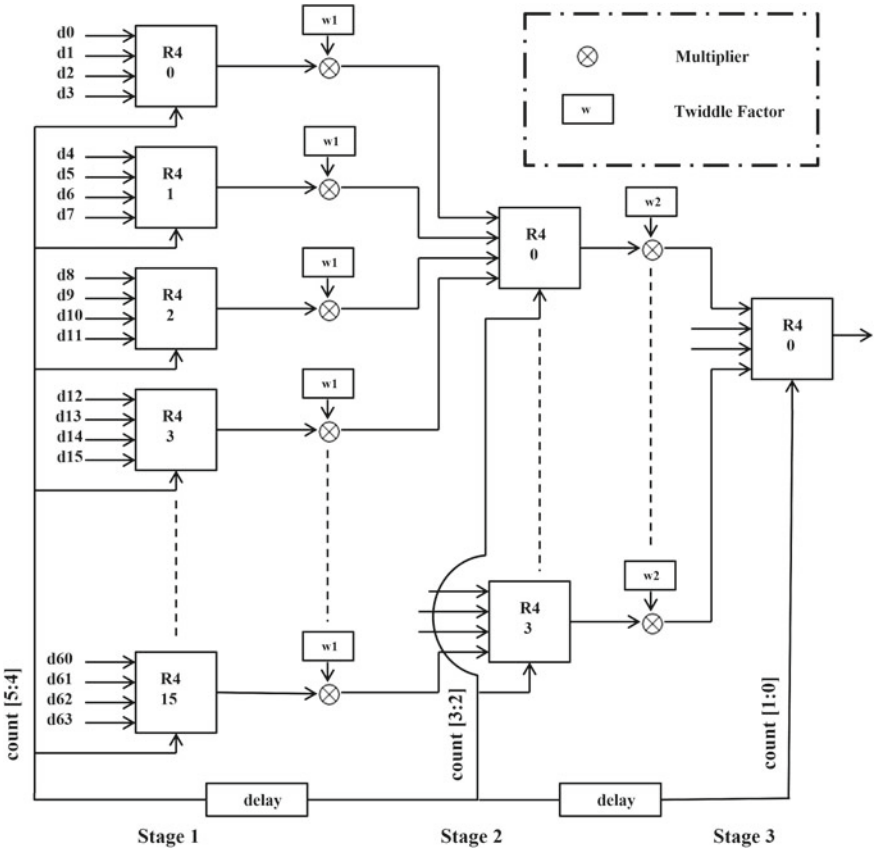


Fig. 4 Radix-4^3 architecture without pipelining

input. Its FFT will be calculated, and it represents the first element of the first row of matrix Y. In cycle 2, the mode is set to be 0 and the second column of X is provided as input. Its FFT will be calculated, and it represents the second element of first row of matrix Y. In this way, all the elements of matrix Y are produced.

In the 65th cycle, when the entire first row of matrix Y is produced, it is provided to the second $R\,4^3$ block. Registers are used to store the intermediate results. In each cycle of the second block, the mode is changed from 0 to 63 resulting in the first row of the matrix Z.

To get the 2D FFT output in an ordered way, for the first $R4^3$ block, the modes are provided as 0, 1, 2,…, 63. This produces the output of the first $R4^3$ block in a bit-reversed order. When bit-reversed input is provided to the second $R4^3$ block, 2D FFT produces output in an ordered way.

A parallel unrolled architecture is used to implement the $R4^3$ block. The ordering of output is controlled by a six-bit up-counter. The implementation in this way is using less memory and improved latency. First, $R4^3$ block is used to implement the 1D FFT and its output is provided to the next $R4^3$ stage to produce 2D FFT. The $R4^3$

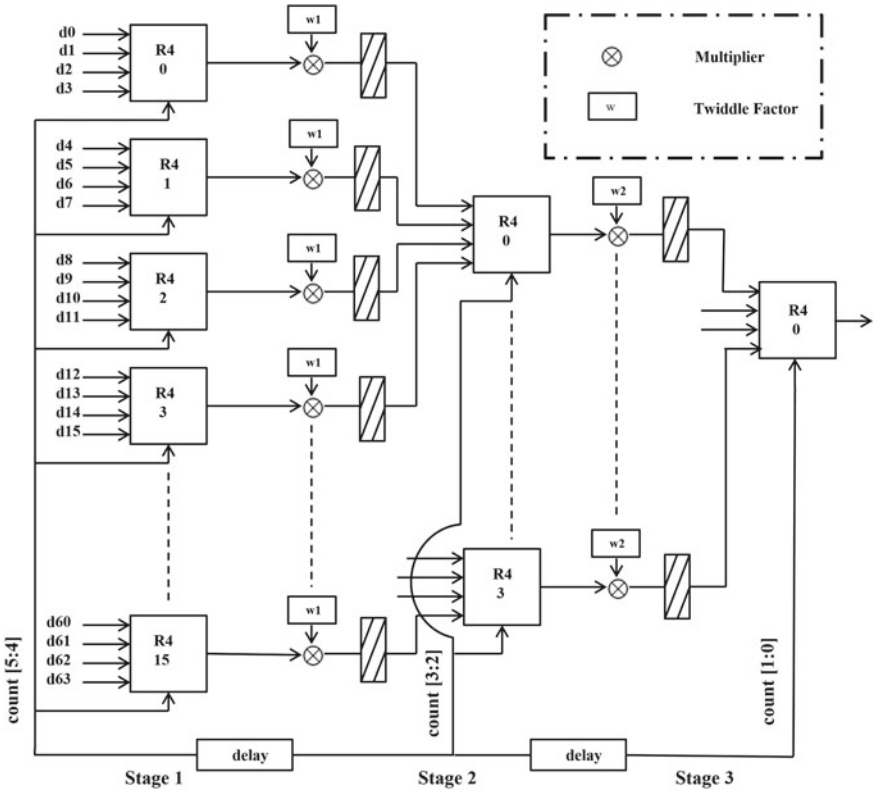


Fig. 5 Radix-4^3 architecture with pipelined multipliers

blocks which are used for the first phase and second phase are alike. The 2D FFT architecture block diagram is shown in Fig. 6.

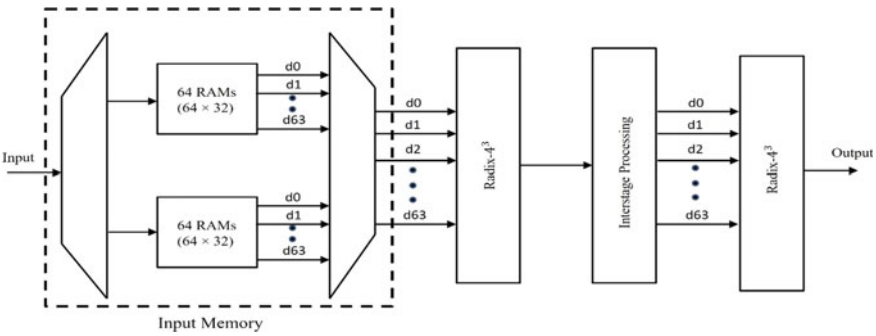


Fig. 6. 2D FFT architecture

We have used two banks of 64 RAMs for the input memory. One memory is used to read-in the data while the other one is used to read-out the data. One bank of RAM is shown in Fig. 3. RAM0 receives the first 64 inputs; the next 64 inputs are received by RAM1 and so on. The RAMs are read-in a parallel way. For all RAMs, the read address is the same.

The intermediate stage processing uses two banks of registers. In each bank, there are 64 registers. In the first bank, 64 registers act as shift registers. In every clock cycle, registers in the shift registers are advanced by one register. So, after 64 clock cycles, once we get inputs in all 64 registers of the first bank, it is parallel loaded in to the second bank. The registers of the second bank will act as the input of the second $R4^3$ block.

The control of the architecture is achieved by a six-bit up-counter for each $R4^3$ block. For input memory, six bits are required for write address and read address. For radix-4 block, the mode signal is generated using a two-bit up-counter. The same write address and read address are provided to all the 64 RAMs. The signals are synchronized with the previous stage delay.

The FFT architecture uses a continuous flow of data. Computations are constantly performed in FFT. The FFT takes a single input per clock cycle.

4 Implementation Results and Analysis

Xilinx Virtex 7 (XC7V2000T) device is used for FPGA implementation. FPGA results are presented in Tables 1 and 2 for 64 × 64 FFT with and without pipelining, respectively. Table 3 shows a comparison of 2D FFT architectures.

From Table 3, it is evident that our architecture takes fewer cycles and area (slices) with improved frequency. The operating frequency is 194.024 and 293.936 MHz for non-pipelined and pipelined architecture, respectively. The output is produced in

Table 1 Resources used for 64 × 64 FFT without pipelining

Resources	Used/available	Percentage
Slices	4,742/305,400	1.55
Flip-flops	11,482/2,443, 200	0.47
Block RAM	64/2,584	2.47

Table 2 Resources used for 64 × 64 FFT with pipelining

Resources	Used/Available	Percentage
Slices	7,879/305,400	2.58
Flip-flops	11,404/2,443, 200	0.466
Block RAM	64/2,584	2.47

Table 3 Different FFT architecture comparison

	Size	FPGA	Freq.(MHz)	Slices	Clk cycles	Throughput	Cont. flow
This work (without pipelining)	64 × 64	XC7V2000T	194.024	4,742	4096	194.024	Yes
This work (with pipelining)	64 × 64	XC7V2000T	293.936	7,879	4096	293.936	Yes
[6]	64 × 64	XC7V2000T	156.254	5,692	4096	156.254	Yes
[18]	64 × 64	XC4VSX35	100	6,472	4516	90.6	No

an ordered way without using supplementary hardware. The memory is reduced considerably in our architecture. The FFT architecture uses a continuous flow of data.

4096 cycles are consumed by memory block. To read data from memory, one clock cycle is needed. So, total latency requirement is 4106 clock cycles but effectively for 64 × 64 FFT, 4096 clock cycles are needed.

5 Conclusion

In this article, we have designed an efficient 64 × 64 point 2D FFT architecture with the radix-4 block structure. The output of first radix-4^3 block is provided as input to the next radix-4^3 block. To obtain output in normal order, the six-bit counter is used. There is a significant improvement in terms of resources used and throughput when compared with the existing architectures for FFT. Eight bits are used to represent the real part and eight bits for the imaginary part. Verilog HDL is used to implement the 2D FFT architecture in RTL. As the twiddle factors are generated and used directly rather than storing in memory, it improves computation time and area. The designed architecture takes 4096 clock cycles to perform 64 × 64 2D FFT. FPGA implementation of 2D FFT is done with an operating frequency of 194.024 and 293.936 MHz for without pipelined architecture and with pipelined architecture, respectively.

References

1. Proakis JG (2001) Digital signal processing: principles algorithms and applications. Pearson Education India

2. Puschel M, Moura JM, Johnson JR, Padua D, Veloso MM, Singer BW, Xiong J, Franchetti F, Gacic A, Voronenko Y et al (2005) Spiral: code generation for dsp transforms. Proc IEEE 93(2): 232–275
3. Frigo M, Johnson SG (1998) Fftw: An adaptive software architecture for the fft. In: Proceedings of the 1998 IEEE international conference on acoustics, speech and signal processing, ICASSP'98 (Cat. No. 98CH36181). IEEE , vol 3, pp 1381–1384
4. Kee H, Bhattacharyya SS, Petersen N, Kornerup J (2009) Resource efficient acceleration of 2-dimensional fast fourier transform computations on fpgas. In: 2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC). IEEE, 2009, pp 1–8
5. Babionitakis K, Chouliaras VA, Manolopoulos K, Nakos K, Reisis D, Vlassopoulos N (2010) Fully systolic fft architecture for giga-sample applications. J Signal Proces Syst 58(3):281–299
6. Kala S, Mathew J, Jose BR, Nalesh S (2018) Radix-43 based twodimensional fft architecture with efficient data reordering scheme. IET Comput Digital Tech 13(2):78–86
7. Kala S, Nalesh S, Maity A, Nandy S, Narayan R (2013) High throughput, low latency, memory optimized 64k point fft architecture using novel radix-4 butterfly unit. In: 2013 IEEE International Symposium on Circuits and Systems (ISCAS2013). IEEE, 2013, pp 3034–3037
8. He S, Torkelson M (1998) Design and implementation of a 1024- point pipeline fft processor. In: Proceedings of the IEEE 1998 Custom Integrated Circuits Conference (Cat. No. 98CH36143). IEEE, pp 131–134
9. Oh J-Y, Lim M-S (2005) New radix-2 to the 4th power pipeline fft processor. IEICE Trans Electron 88(8):1740–1746
10. Yin X-B, Yu F, Ma Z-G (2016) Resource-efficient pipelined architectures for radix-2 real-valued fft with real datapaths. IEEE Trans Circuits Syst II Express Briefs 63(8):803–807
11. Xia K, Wu B, Zhou X, Xiong T (2016) A generalized conflict-free address scheme for arbitrary 2k-point memory-based fft processors. In: 2016 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2016, pp 2126–2129
12. Xing Q-J, Ma Z-G, Xu Y-K (2017) A novel conflict-free parallel memory access scheme for fft processors. IEEE Trans Circuits Syst II Express Briefs 64(11):1347–1351
13. Chen R, Prasanna VK (2014) Energy optimizations for fpga-based 2-d fft architecture. In: 2014 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, pp 1–6
14. Lenart T, Gustafsson M, Öwall V (2008) A hardware acceleration platform for digital holographic imaging. J Signal Proces Syst 52(3):297–311
15. Yu C-L, Kim J-S, Deng L, Kestur S, Narayanan V, Chakrabarti C (2011) Fpga architecture for 2d discrete fourier transfo rm based on 2d decomposition for large-sized data. J Signal Proces Syst 64(1):109–122
16. Yang K-J, Tsai S-H, Chuang GC (2012) Mdc fft/ifft processor with variable length for mimo-ofdm systems. IEEE Trans Very Large Scale Integr (VLSI) Syst 21(4):720–731
17. Kristensen F, Nilsson P, Olsson A (2003) Flexible baseband transmitter for ofdm. In: Proceedings of IASTED Conference Circuits Signals System, 2003, pp 356–361
18. Rodríguez-Ramos J, Castello EM, Conde CD, Valido MR, Marichal-Hernández J (2008) 2d-fft implementation on fpga for wavefront phase recovery from the cafadis camera. In: Adaptive optics systems, vol 7015. International society for optics and photonics, 2008, p 70153

Design and Simulation of Fourth-Order Delta-Sigma Modulator-MASH Architecture

Sujata Sanjay Kotabagi and P. Subbanna Bhat

Abstract The delta-sigma modulator is used as an oversampling ADC, which processes the signal input and the quantization noise by two distinct transfer functions, thus, separating the signal and noise into two distinct frequency bands. The resulting output of the modulator is digitally processed to separate the signal from noise. The performance of the ADC can be improved by increasing the order of the modulator, which imposes constraints on the system stability and on the input signal range. This paper examines the issues related to the design of a stable fourth-order MultistAge noise SHaping architecture(MASH) ADC, using two second-order delta-sigma modulators in cascade. The architecture achieves second-order Signal Transfer Function (STF) with fourth-order noise shaping . The modulator employs 32 MHz—clock, to obtain 8-bit sample sequence at 128 kHz sample rate. The resulting output is processed by a digital low pass filter to separate the signal from noise. The FIR filter is implemented using frequency sampling technique.

Keywords Oversampling · Noise shaping · Quantization noise · Linear model · Stability · Filter

1 Introduction

Analog to digital converters (ADC) are of two types: Nyquist rate ADC and oversampling ADC. In the former, the minimum sampling frequency is defined to be twice the bandwidth of the analog input ($f_s = 2f_b$, where f_b is the input signal bandwidth) (Fig. 1).

S. S. Kotabagi (✉)
School of Electronics, KLE Technological University, Hubli, India
e-mail: sujatask@bvb.edu

P. Subbanna Bhat
School of Electronics and Communication, KLE Technological University, Hubli, India
e-mail: p.subbannabhat@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_54

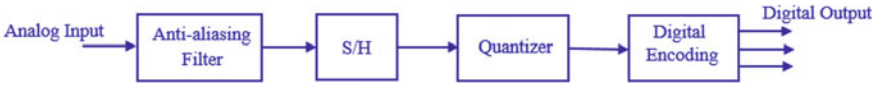


Fig. 1 Nyquist rate ADC

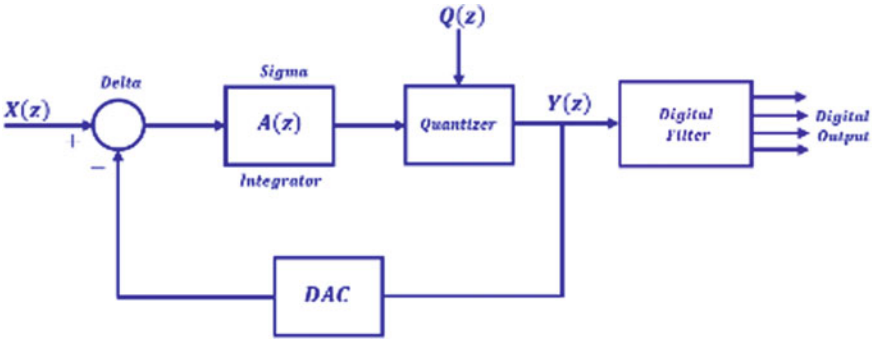


Fig. 2 Oversampling ADC

The oversampling ADC (Fig. 2) samples the analog signal at a rate much higher than the Nyquist rate, which is exploited to reduce the quantization noise in the signal bandwidth [1]. The delta-sigma modulator is a closed loop system, which processes the signal and quantization noise differently using Signal Transfer Function (STF) and Noise Transfer Function (NTF), which results in confining the signal and noise into two mutually exclusive frequency bands and this is known as noise shaping. The output of the modulator is in digital form, containing both the signal and quantization noise, which is filtered using Digital Signal Processing (DSP) to obtain the digital encoding of the signal input [1].The performance of delta-sigma ADC can be improved in three ways:

(i) By increasing the sampling rate. Since the (quantization) noise is distributed uniformly across sampling frequency range, with the increase in the sampling rate, noise content within the signal bandwidth is reduced.
(ii) By enhancing the resolution of the quantizer block within the loop. This will reduce the quantization noise $Q(z)$ injected at the output node of the loop. But, more complex quantizer could cause greater delay, which would adversely affect the stability of the loop.
(iii) By enhancing the order of the delta-sigma modulator, causing higher-order STF and NTF. Increase in the order of the modulator would adversely affect the stability of the modulator system.

This paper examines the issues related to the design of a stable fourth-order multi-stage noise shaping architecture (MASH) ADC using two second-order delta-sigma modulators in cascade. The architecture achieves second-order STF with fourth-order NTF.

2 First-Order Delta-Sigma Modulator

The basic architecture of a first-order delta-sigma modulator is shown in Fig. 3. The loop contains an integrator, followed by 1-bit or multi-bit quantizer, and an appropriate DAC which converts the digital output of the quantizer back to analog signal. The integrator in the loop is often implemented using switched capacitor components, which provides the advantage of combining the mixer and integrator in one unit. However, since the switched capacitor integrator is a discrete time system, the transfer function of the integrator is defined in the Z-domain.

The output $Y(z)$ contains two components—one of them is a modified version of the input signal and the other is the quantization noise confined to the high frequency band.

The quantizer inside the loop is a comparator (1-bit ADC) which is a nonlinear unit. For the purpose of analysis, the comparator is modeled as a linear gain amplifier with the quantization noise injected at the output node as shown in Fig. 3 [2]. Note that the input to the comparator is an analog signal from the integrator, the output is a logic signal shuttling between '0' and '1'. The linear gain is obtained as the ratio of average output of the comparator, to the average input (analog) signal to the comparator. It has been observed through simulations that the effective gain of this linear model changes with analog input swing (to the comparator), which in turn depends on the signal input to the modulator. That is, the modulator loop gain fluctuates with the input swing. This would impair the modulator stability, as well as the signal integrity of the modulator output, especially in higher-order systems.

Now consider the linear model of the first-order modulator, where G_1 and G_C are the gains of the integrator and comparator, respectively. If $G_{F1} = G_1 G_C$ is the forward gain of the modulator, then the output is given by

$$Y(z) = \frac{G_1 G_C z^{-1}}{1 + (G_1 G_C - 1)z^{-1}} \cdot X(z) + \frac{1 - z^{-1}}{1 + (G_1 G_C - 1)z^{-1}} \cdot Q(z) \qquad (1)$$

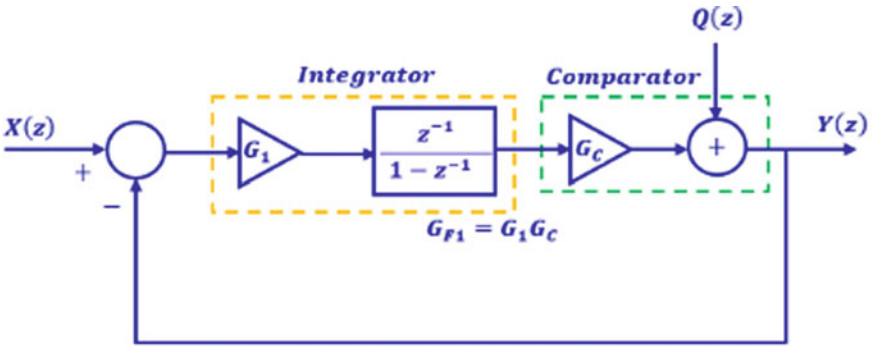


Fig. 3 First-order delta-sigma modulator—linear model

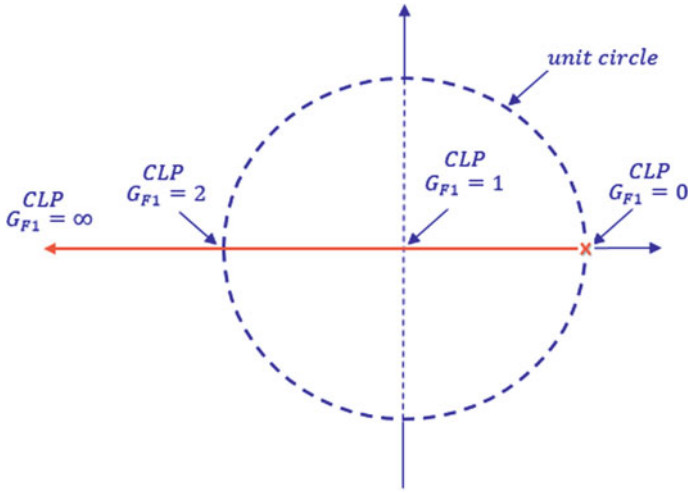


Fig. 4 Root loci (CLP) of first-order modulator

$$= \frac{G_{F1}z^{-1}}{1 + (G_{F1} - 1)z^{-1}} \cdot X(z) + \frac{1 - z^{-1}}{1 + (G_{F1} - 1)z^{-1}} \cdot Q(z) \tag{2}$$

From the control system perspective, for variable gain factor G_{F1}, the root loci of the above first-order modulator is shown in Fig. 4. Here, we see that the system is stable only for the range [2]

$$0 \leq G_{F1} \leq 2 \tag{3}$$

For a chosen value of $G_{F1} = G_1 G_C = 1$, Eq. (2) reduces to

$$Y(z) = z^{-1} \cdot X(z) + \left(1 - z^{-1}\right) \cdot Q(z) \tag{4}$$

The output of the modulator $Y(z)$ contains two components, one of them is the delayed version of the input signal and the other containing the quantization noise processed through a first-order high pass filter.

3 Second-Order Delta-Sigma Modulator

A second-order modulator [2] and its simplified block diagram is shown in Fig. 5a, b.

The modulator consists of two loops, with the 1-bit quantizer placed in the second loop.

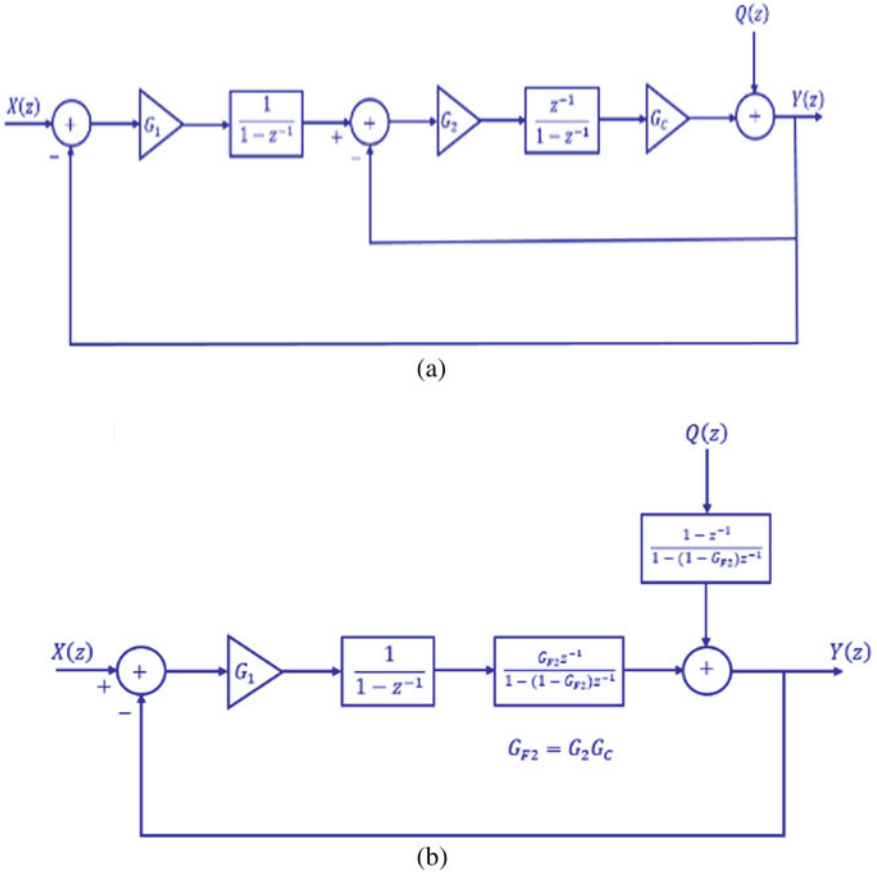


Fig. 5 **a** Second-order modulator with single delay, **b** simplified block diagram

Note that each of the two switched capacitor integrators contribute one-half unit delay ($z^{-0.5}$); the two together contributing one unit delay, which is shown with the second integrator.

The modulator output is given by

$$Y(z) = \frac{G_1 G_{F2}}{(z-1)(z-1+G_{F2}) + G_1 G_{F2}} X(z) + \frac{(z-1)^2}{(z-1)(z-1+G_{F2}) + G_1 G_{F2}} Q(z) \tag{5}$$

where $G_{F2} = G_2 G_C$ is the forward gain of the second loop.

The characteristic equation is given by

$$z^2 + z(G_{F2} - 2) + 1 - G_{F2} + G_1 G_{F2} = 0 \tag{6}$$

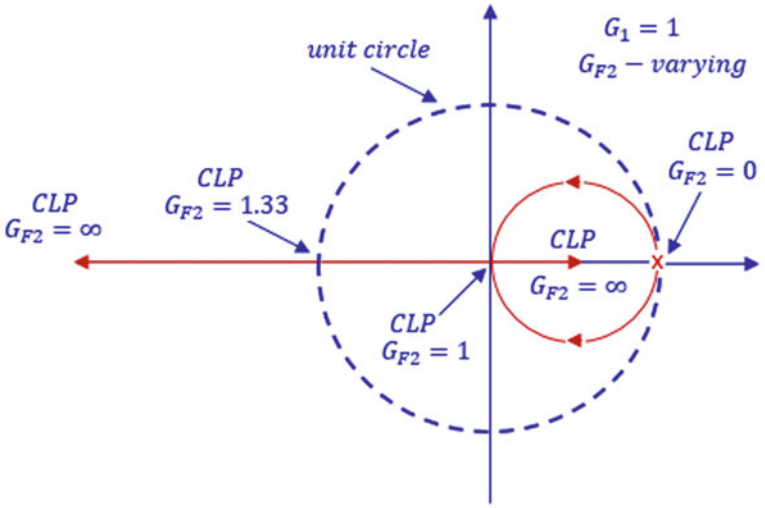


Fig. 6 Root loci of second-order modulator using single delay; $G_1 = 1$

In this paper, the architecture of Fig. 5a is used for implementation of the basic block of second-order modulator. The root loci of the second-order modulator with $G_1 = 1$; $G_2 = 1$; $0 < G_{F2} < \infty$ is shown in Fig. 6.

The system is stable for the range $0 < G_{F2} < 1.33$

For a chosen value of $G_{F2} = G_2 G_C = 1$, Eq. (5) reduces to

$$Y(z) = z^{-1}X(z) + \left(1 - z^{-1}\right)^2 Q(z) \tag{7}$$

Note that while the STF has reduced to a unit delay function (all pass filter), the NTF is a second-order high pass filter.

4 Second-Order Modulator Implementation

The second-order modulator has essentially the following circuit blocks—two Discrete Analog Integrators (DAI), a fully differential clocked comparator and two single bit DACs, whose output feeds back to the mixer units (Fig. 7).

The modulator uses a system clock of 32 MHz. The DAIs are implemented using fully differential Op-Amps and switched capacitor components. The circuit operates with a 2-phase clock (ϕ_1, ϕ_2) of 32 MHz frequency, with each phase having 40% duty ratio. The two successive integrators are operated with alternate sequence of clock phases (ϕ_2, $\phi_1$and ϕ_1, ϕ_2) to minimize signal propagation delay. Each integrator causes half delay ($z^{-0.5}$); the two integrators together cause one unit delay (z^{-1}); this combined unit delay is shown with the second integrator.

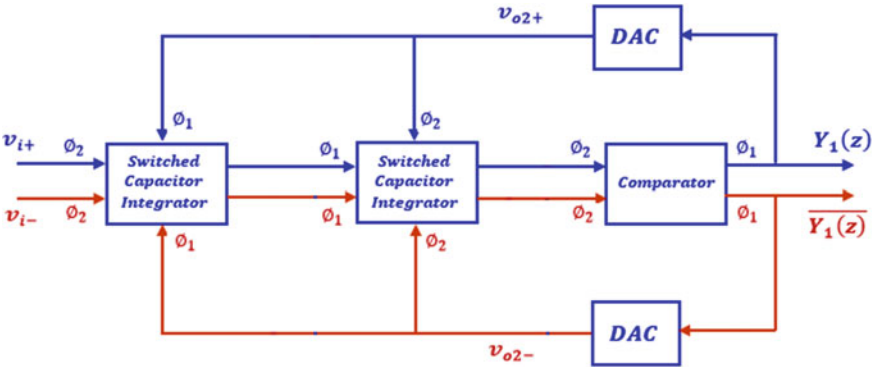


Fig. 7 Second-order modulator block schematic

4.1 Discrete Analog Integrator

At the base of the DAI, is a fully differential Op-Amp with common mode feedback (Fig. 8) [3]. The OPAMP is built for single ended power supply ($V_{DD} = 3.3$ V, UGB = 32.77 MHz; Phase Margin = 68.94°; Gain Margin = 12.48 dB at 119.4 MHz) (Fig. 9).

The fully differential discrete analog integrator (DAI) [2] is built using OPAMP, and switched capacitor components as shown in Fig. 10. The integrator includes the mixer unit and provides a fully differential output which feeds the comparator (1-bit quantizer).

The feedback signal is 'mixed' with the input signal during the ϕ_1 phase. The resulting transfer function is given by

$$V_{o+}(z) = \frac{C_1}{C_2} \frac{V_1(z)z^{-0.5} - V_2(z)}{1 - z^{-1}} \tag{8}$$

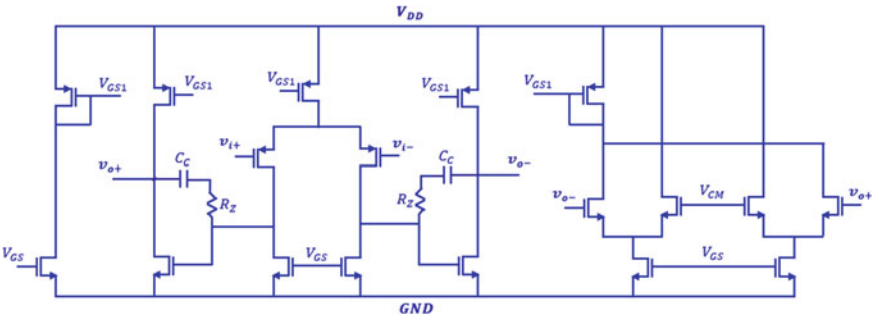


Fig. 8 OPAmp with common mode feedback

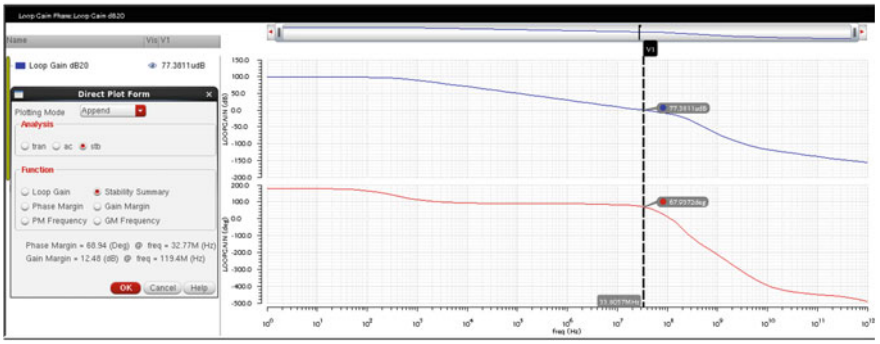


Fig. 9 OPAMP frequency and phase response characteristics

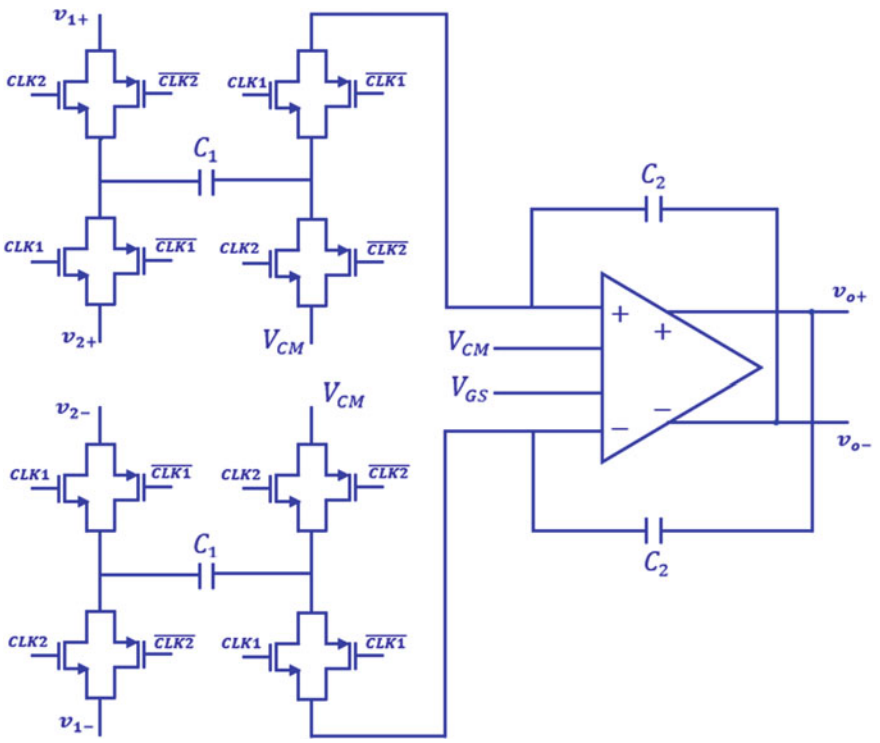


Fig. 10 Discrete analog integrator

The time delay of one-half clock period is indicated by ($z^{-0.5}$) in the transfer function of Eq. (8). The successive integrators in the loop employ the same clock ϕ_1 and ϕ_2, but in reverse order, thus, optimizing the total loop delay to one unit delay as shown with the second integrator block (Fig. 5).

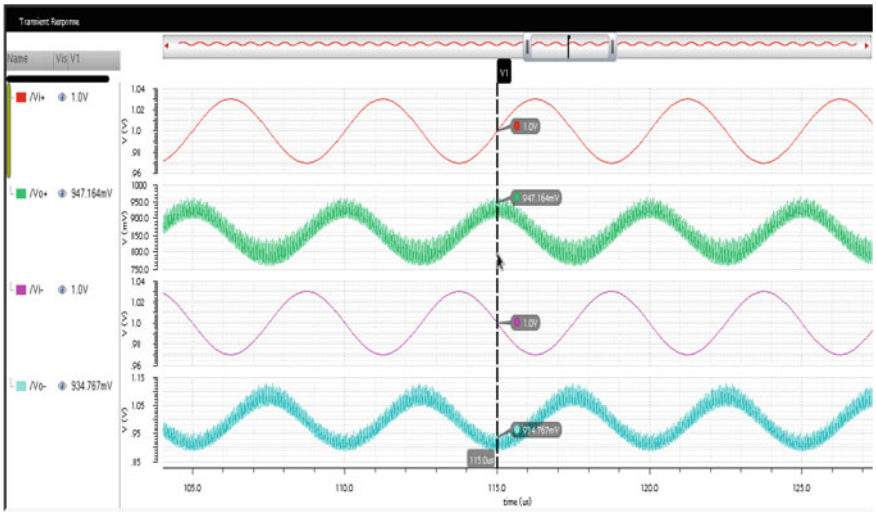


Fig. 11 Input–output waveforms of DAI ($f = 200$ kHz; Clock = 32 MHz)

The integrator gains G_1 and G_2 are decided by the ratio of the capacitors, thus,

$$G_1 = \frac{C_1}{C_2} \text{ and } G_2 = \frac{C_3}{C_4} \tag{9}$$

The response of the integrator of Fig. 10 for a typical sinusoidal signal ($v_{i1} = -v_{i2} = \text{A sin}\omega t$; $v_{0-} = -v_{0+} = A\omega \cos\omega t$; for $G_1 = 0.1$; $A = 30$ mV; $\omega = 2\pi$ (200 kHz) is shown in Fig. 11

4.2 Dynamic Comparator

A fully differential, dynamic, charge distribution, clocked comparator (Fig. 12) with dual outputs is used as single bit quantizer [4]. The comparator draws power only during the ϕ_1 phase of the clock period; is turned off during ϕ_2 (does not draw power from the supply); however, during ϕ_2 phase, the integrator output is written into the comparator. During ϕ_1, it compares two pairs of signals (v_{i+} with $V_{\text{REF}+}$) (v_{i-} with $V_{\text{REF}-}$) simultaneously. The two decisions are expected to overlap, and the decision point is given in (10):

$$(v_{i+} - V_{\text{REF}+}) = (V_{\text{REF}-} - v_{i-}) \geq 0 \tag{10}$$

Superposing the two decision points of (10), we can get a single point, thus:

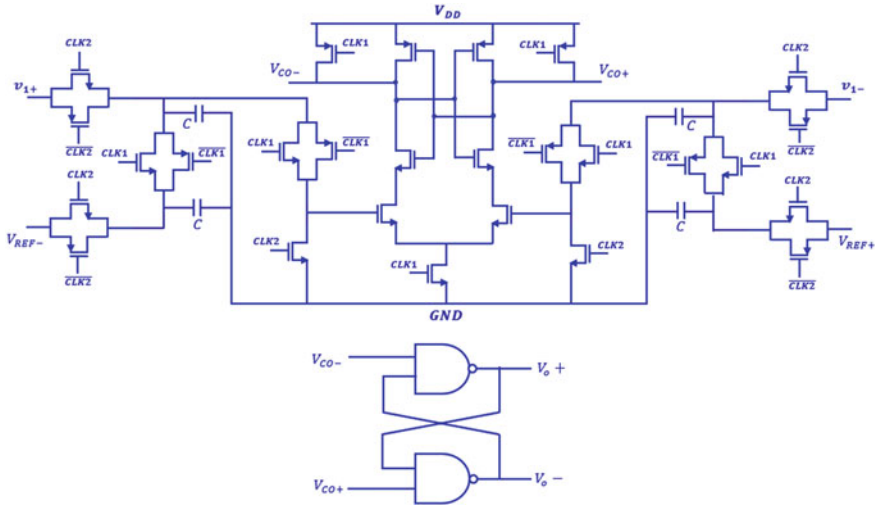


Fig. 12 Charge distribution clocked comparator

$$(v_{i+} + V_{\text{REF}-}) - (v_{i-} + V_{\text{REF}+}) \geq 0 \tag{11}$$

Equation (11) is accomplished by aggregating the voltage signals ($v_{i+} + V_{\text{REF}-}$) using a pair of charge sharing capacitors at one input of the comparator, and ($v_{i-} + V_{\text{REF}+}$) at the other input (Fig. 12). During the ϕ_2 phase of the system clock, the comparator is off, but the individual capacitors are charged by the analog input signals as well as the reference voltages; and during the ϕ_1 phase, the logic output of the comparator (V_{0+}, V_{0-}) are transferred to the output latch. The comparator response for a 200 kHz sinusoidal signal ($C = 50$ fF, $V_{\text{REF}+} = V_{\text{REF}-} = 1$ V, $V_{\text{DD}} = 3.3$ V) using a two-phase clock of 32 MHz as shown in Fig. 13.

4.3 Digital to Analog Converter

The output of the comparator (logic signal) is converted into appropriate analog signal levels using a 1-bit DAC.

A 1-bit DAC is built using two CMOS inverters, with their power supply voltage (2.0 V) derived from the bandgap reference, instead of the normal V_{DD} (Fig. 14). Thus, the output of the DAC is made independent of the power supply fluctuations.

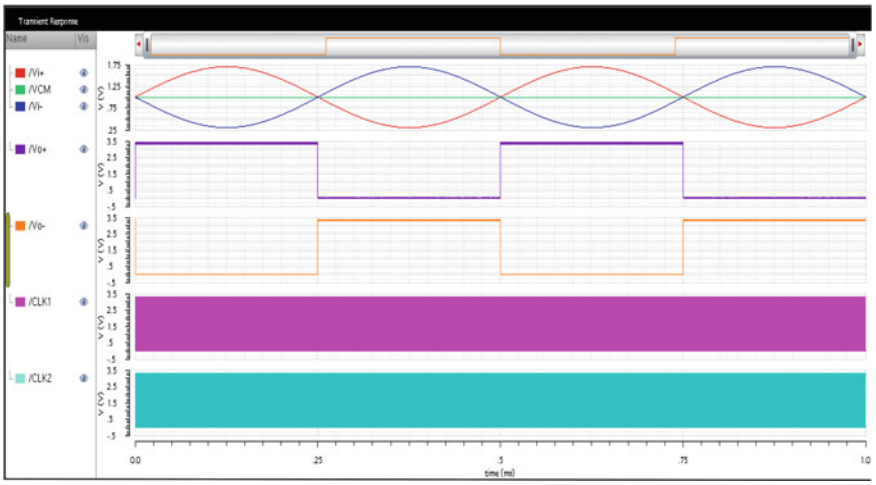


Fig. 13 Input–output waveforms of clocked comparator ($f = 200$ kHz; Clock $= 32$ MHz)

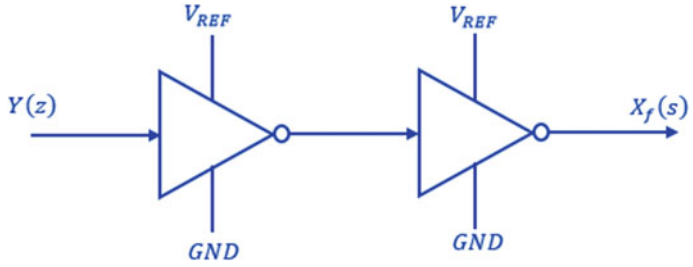


Fig. 14. 1-bit DAC

4.4 Second-Order Modulator Operation

Figure 15 shows the cadence schematic for the second-order modulator.

Tables 1 and 2 indicate the range of input signal (for which the integrator remains in linear mode), and the variation of the comparator gain with the input signal for set values of integrator gains (G_1, G_2)

Columns 1 and 2 of Table 1 indicate the limits of the input signal range for which both the integrators are in the linear range. Column 5 list the limits of the comparator gain for the extreme values of the input signals (listed in columns 1 and 2). Column 6 of Table 1 indicates the overall forward gain of the modulator (for the input signal range defined in columns 1 and 2). The system is stable for $0 < G_1 G_{F2} < 1.33$ (Fig. 7)

From the above simulation results, it is clear that the comparator gain is dynamic and is dependent on the input voltage fluctuations. Column 6 indicates that the system is stable.

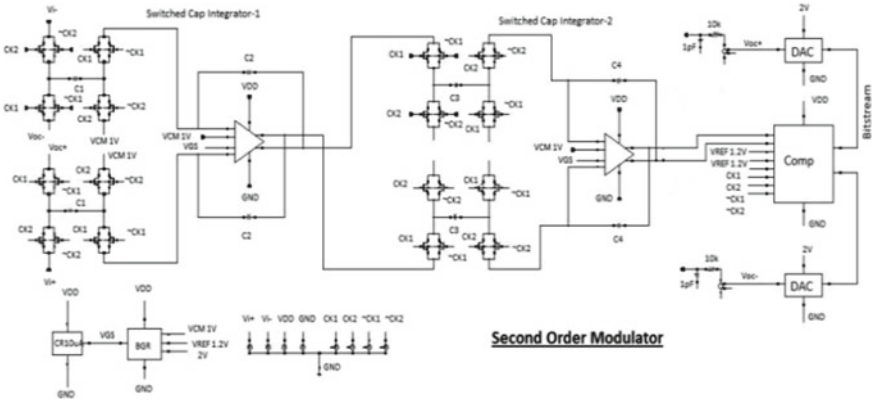


Fig. 15 Second-order modulator cadence schematic

Table 1 Comparator gain variation with input signal range ($G_2 = 1.0$)

$V_i +$ (volts)	V_i-(volts)	G_1	G_2	G_C	$G_1\ G_{F2}$
1.8	0.2	0.1	1	1.38	0.138
1.8	0.2	0.2	1	1.33	0.266
1.7	0.3	0.3	1	1.39	0.417
1.6	0.4	0.4	1	1.36	0.543
1.5	0.5	0.5	1	1.36	0.679
1.3	0.7	0.6	1	1.25	0.749
1.3	0.7	0.7	1	1.24	0.869
1.25	0.75	0.8	1	1.18	0.941
1.2	0.8	0.9	1	1.12	1.008
1.25	0.75	1	1	1.160	1.160

Table 2 Comparator gain variation with input signal range ($G_1 = 1.0$)

$V_i +$ (volts)	V_i-(volts)	G_1	G_2	G_C	$G_1\ G_{F2}$
1.2	0.8	1	0.1	1.141	0.114
1.3	0.7	1	0.2	1.200	0.240
1.35	0.65	1	0.3	1.235	0.370
1.35	0.65	1	0.4	1.235	0.494
1.3	0.7	1	0.5	1.196	0.598
1.3	0.7	1	0.6	1.193	0.716
1.3	0.7	1	0.7	1.190	0.833
1.3	0.7	1	0.8	1.195	0.956
1.25	0.75	1	0.9	1.157	1.042
1.25	0.75	1	1	1.160	1.160

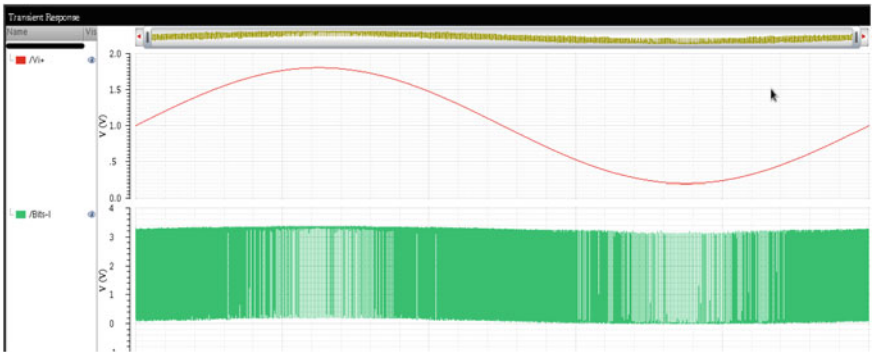


Fig. 16 Simulation results of second-order modulator for sine wave input (stable condition)

Thus, the above system has two limitations:

(i) Since the linear gain of the comparator depends on the signal strength, the pole locations shift with the signal, affecting both the gain and phase angle of the modulator transfer function.
(ii) For the chosen values of G_1 and G_2, the input signal is confined to a very small range (for which the integrator operates in the linear mode).

Figure 16 shows the simulation results (quantizer output) of the modulator for sinusoidal input ($G_1 = 0.1$; $G_2 = 0.1$; $G_C = 0.99$; $G_1.G_2.G_C = 0.0099$; $V_i = \pm 0.7$ sin ωt; $\omega = 2\pi(4\text{ KHz})$).

5 Fourth-Order Delta-Sigma Modulator

Further, increase in the order of the modulator to improve noise shaping poses serious challenges to system stability. However, the proposed architecture of a fourth-order MASH architecture modulator uses two identical blocks of second-order modulator in cascade as shown in Figs. 17 and 18 [2]. Here, the stability issues are confined

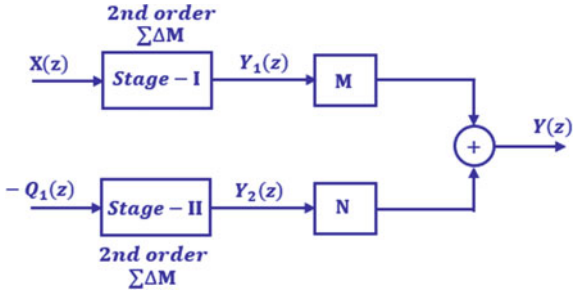


Fig. 17 Block diagram of fourth-order (2–2) cascaded modulator

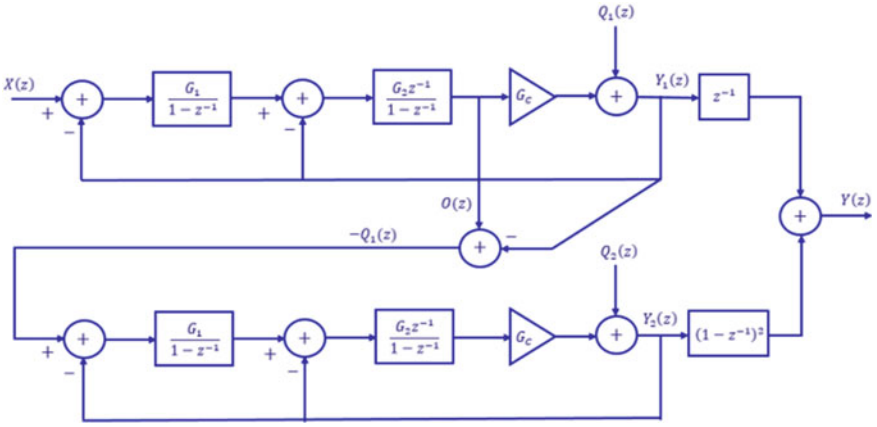


Fig. 18 Detailed fourth-order (2–2) cascaded modulator

to the second-order modulator, while the noise shaping transfer function (NTF) is improved to the fourth-order level.

The desired output of the first stage (second-order modulator) is given by

$$Y_1(z) = z^{-1}X(z) + \left(1 - z^{-1}\right)^2 Q_1(z) \tag{12}$$

From the output of the first stage, the negated version of the quantization error $[-Q_1(z)]$ is extracted using a mixer, which is fed to the second stage to obtain $Y_2(z)$.

$$Y_2(z) = -z^{-1}Q_1(z) + \left(1 - z^{-1}\right)^2 Q_2(z) \tag{13}$$

The output of both the stages is digitally combined as shown in Fig. 18 to produce the desired final output $Y(z)$ containing a second-order STF and fourth-order NTF:

$$Y(z) = MY_1(z) + NY_2(z) = z^{-2}X(z) + \left(1 - z^{-1}\right)^4 Q_2(z) \tag{14}$$

That is,

$$Y(z) = M[z^{-1}X(z) + \left(1 - z^{-1}\right)^2 Q_1(z)] + N\left[-z^{-1}Q_1(z) + \left(1 - z^{-1}\right)^2 Q_2(z)\right] \tag{15}$$

Comparing terms in (15),

$$M = z^{-1};\ N = (1 - z^{-1})^2$$

Because of the noise cancelation, the output reduces to:

$$Y(z) = \frac{G_1 G_{F2}}{z^2 - z(2 - G_{F2} - G_1 G_{F2}) + (1 - G_{F2})} X(z) + \frac{G_1 G_{F2}(z-1)^4}{z^2\left[z^2 - z(2 - G_{F2} - G_1 G_{F2}) + (1 - G_{F2})\right]} Q_2(z) \quad (16)$$

For values $G_1 = 1$, $G_{F2} = 1$, the output reduces to (14)

As in the case of second-order modulator, for a different set of gain factors $G_1 = 0.5$, $G_{F2} = 1$, the closed loop pole locations would change, affecting the signal integrity of the output.

$$Y(z) = \frac{0.5}{z^2 - 0.5z} X(z) + \frac{0.5(z-1)^4}{z^2\left[z^2 - 0.5z\right]} Q_2(z) = \frac{0.5z^{-2}}{1 - 0.5z^{-1}} X(z) + \frac{0.5\left(1 - z^{-1}\right)^4}{\left[1 - 0.5z^{-1}\right]} Q_2(z) \quad (17)$$

Thus, the design of the modulator is essentially centered on the second-order modulator [2].

6 Results

Figure 19 shows the fourth-order MASH architecture input and output signals along with the bit-stream outputs of both stages. The two stages are identical with the same gain factor settings $G_1 = 0.1$; $G_2 = 0.1$; $G_C \approx 1.0$; $G_1 G_2 G_C \approx 0.01$. The system is simulated with an input signal $V_i = \pm 0.7 \sin \omega t$; $\omega = 2\pi \times 2$ KHz. Bit stream_1 and Bit stream_2 along with the reconstructed signal after the DSP low pass filter (Figs. 20 and 21).

Figure 22 compares the responses of the second-order modulator to the fourth-order modulator for the same input signal and modulator gain factors (G_1 =

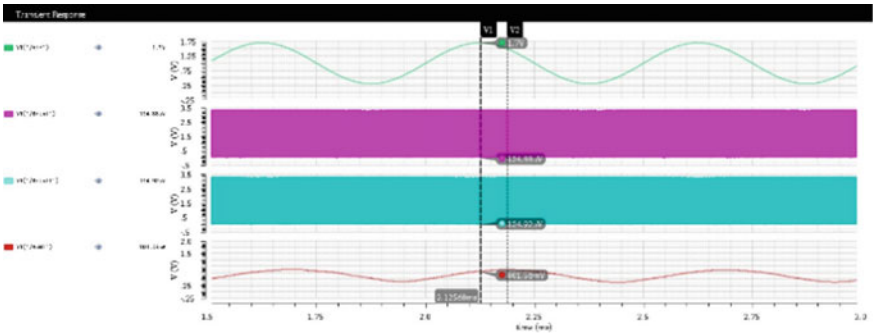


Fig. 19 Input to the fourth-order ADC, bit stream_1, bit stream_2, DAC output after DSP filter

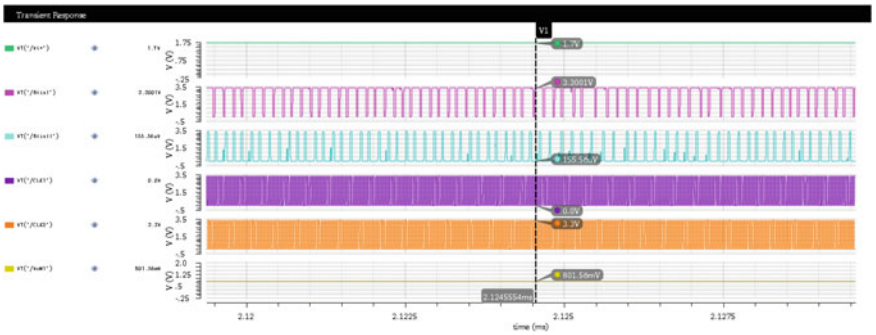


Fig. 20 Zoomed version_1 of Fig. 19 demonstrating the modulator bit streams

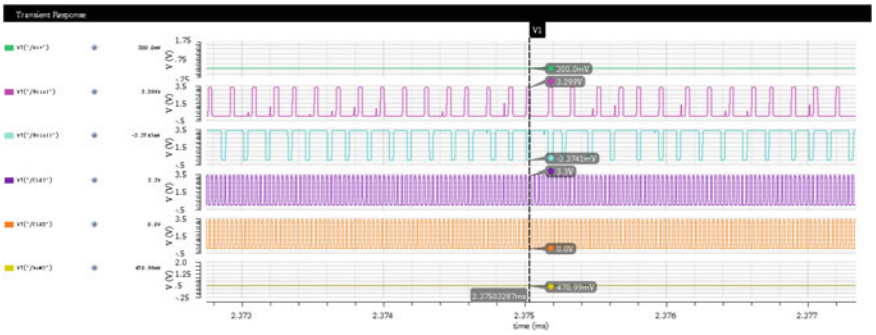


Fig. 21 Zoomed version_2 of Fig. 19 demonstrating the modulator bit streams

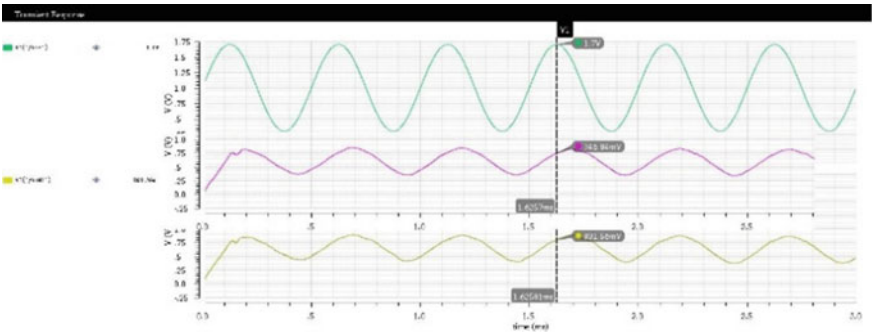


Fig. 22 Comparison of responses of second-order and fourth-order ADCs

0.1; $G_2 = 0.1$; $G_C \approx 1.0$; $V_i = \pm 0.7 \sin 2\pi \times 2000t$). The following points may be noted:

- For a sinusoidal input, the (reconstructed analog) output of the ADC indicates a marginally distorted sinewave, which is attributed to the shifting of closed loop poles of the modulator.
- The instantaneous value of the reconstructed analog signals for the (second order and fourth order) modulators are different; this is attributed to the difference in the order of NTF in the two cases. The input signal reaches the output port through the STF as well as NTF. While the STF component is the same in both the cases, the NTF components are different.

7 Conclusion

This paper examines the stability issues related to the second-order and fourth-order (MASH architecture) delta-sigma modulators and presents a strategy for its implementation. The work includes the following steps:

- Design and simulation of second-order modulator for fully differential input signal.
- Identifying the range of loop gain factors for linear and stable operation of the (second order) modulator.
- Implementation of fourth-order MASH architecture by cascading two second-order modulators.

It has been shown that the second-order modulator is stable for the limited range of loop gain of $0 \leq G_1 G_{F2} \leq 1.33$. And the system is linear for limited range of the input signal.

An optimum pole placement (from the perspective of signal integrity) can be achieved with $G_1 = G_2 = 1$, but this choice restricts the input signal to a very small range (approximately $\pm 0.25V$, for a system power supply of 3.3 V.). A wider range of input signal ($\pm 0.5V$) can be achieved with good stability margins with a setting of $G_1 = 0.5, G_2 = 1.0$.

[*The schematics of this paper are simulated in UMC 180 nm technology on Cadence Virtuoso 6.1.5 platform with a supply voltage of 3.3 V,* $V_{REF} = 1.2$ *V,* $V_{CM} = 1$ *V,* $I_{REF} = 10\ \mu A$ *and clock frequency of 32 MHz*].

References

1. Jacob Baker R, CMOS circuit design, layout and simulation. IEEE Press
2. Jacob Baker R, CMOS mixed signal circuit design. IEEE Press
3. Parihar RS, Gupta A, Design of a fully differential two-stage CMOS Op-Amp for high gain, high bandwidth applications. IEEE J Microchip Technol Inc 40(1):32–38
4. Sumanen L, Waltari M, Hakkarainen V, Halonen K, CMOS Dynamic Comparators for Pipeline ADC. Helsinki University of Technology, Electronic Circuit Design Laboratory P.O.Box 3000, FIN-02015 HUT, Finland
5. Pavan S, Schreier R, Themes G, Understanding delta-sigma converters, 2nd edn. IEEE Press, Wiley-Interscience.

Multiplexer-based High-speed S-box Architecture for PRESENT Cipher in FPGA

Piyush Modi, Pulkit Singh, and Bibhudendra Acharya

Abstract The Internet of Things (IoT) is the technology that will revolutionize our world. It is a system of interrelated devices linked to the Internet. It comprises large number of resource-constrained devices, machines and sensors. Actualization of IoT requires large amount of devices like sensors and actuators which requires large power but with bounded power supply and area. The aim of lightweight cryptography is to provide security solution for resource-constrained devices. The conventional cryptography algorithms are slow, less responsive and require more energy and storage space as compared to lightweight algorithm. Therefore, researchers are more oriented toward lightweight-based algorithms. This chapter discusses hardware implementation of PRESENT cipher on a field-programmable gate array (FPGA) platform. In this work, our goal is to improve its performance in terms of speed and minimize its area as much as possible. In our proposed design, we have implemented a combinational S-box through multiplexers and some logic gates. In this chapter, various platforms for implementation such as Spartan-6, Virtex-5 and Virtex-4 are used, and as a result, it achieves high throughput per slice for PRESENT cipher as well as highest maximum frequency.

Keywords Internet of things (IoT) · PRESENT cipher · Advanced encryption standard (AES) · Field-programmable gate array (FPGA)

P. Modi · P. Singh · B. Acharya (✉)
Department of Electronics and Communication Engineering, National Institute of Technology Raipur, Raipur, Chhattisgarh, India
e-mail: bacharya.etc@nitrr.ac.in

P. Modi
e-mail: modipiyush16@gmail.com

P. Singh
e-mail: pksingh.phd2015.elx@nitrr.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_55

1 Introduction

The ever-changing design and development of Internet of Things (IoT) enable devices creating a tough challenge for the amalgamation of different working technologies under one umbrella providing solution to different applications. Many applications demand integration of scientific knowledge from sensing of data, sending it through some communication channel, processing the received data, storing the data on the cloud, and artificial intelligence. Here, the data initially is sensed by the sensor and is processed by signal conditioning system at the physical layer; then, the data is transmitted through the network layer over the Internet so that it can be accessed by the concerned person and different data mining and big data application process the data and provide the output to be notified/informed to the concern person or have a database to take decision in future. The data being communicated through the Internet is prone to many security issues, application spoofing, cyber-attacks, and personal privacy. The big data application raises the concern regarding the privacy and security of the devices enabled under IoT [1, 2].

Security features such as data integrity, authentication and data confidentiality are the main concerns for any cryptography mechanism. Providing security against any attack is the aim of any cryptographic algorithm. Security attacks may be active or passive and can violate security services. Power consumption, processing speed, and sizes of the computational memories are the factors required for lightweight cryptography implementation. Cryptography branches like symmetric key and public key cryptographies take large processing power and rise power and energy consumption. On the other hand, IoT devices are considered as constrained low-cost devices with limited processing power, bounded memory size, and restricted power/energy budget. There is always requirement of optimizing cryptographic algorithms in hardware for cost, area, throughput, power, and energy consumption. However, it is hard to achieve different features like cost, throughput, and power/energy consumption at the same time [3].

AES is not acceptable for exceedingly limited environments such as healthcare devices, despite it is an outstanding and favoured for many block cipher applications. In 2001, National Institute of Standards and Technology established Advanced Encryption Standard (AES), also known as Rijndael. The aim of AES was to provide encryption of electronic data to secure US government documents. In AES, S-box creates nonlinearity and it also influences AES security and confidentiality. Many block ciphers use S-box [4]. S-box in AES is known as Substitute Bytes Transformation (SubBytes) and Inverse Substitute Bytes Transformation (InvSubBytes). The Substitute Bytes Transformation is used for encryption, and it does byte-by-byte substitution of the block. AES S-box takes 256-bit input. All 256-bit input is divided into 16 states in the form of 4 * 4 matrix. Each state consists of 8 bit or 1 byte. Each individual byte of the state is changed into a new byte [5]. The least significant 4 bits of a byte is used to represent as row value, and the most significant 4 bits of a byte is used to represent as column value. These row values and column values are used

as indexes into the S-box which is in the form of 16 * 16 table. S-box gives a unique 8-bit output for a particular row and column value.

1.1 Related Work

Various new block cipher algorithms are described as lightweight block ciphers. As compared to conventional cryptography, lightweight cryptography algorithms are easy, straightforward, and fast especially in hardware implementation [6]. An S-box is one of the parts of encryption and decryption of any lightweight algorithm. It is only nonlinear component in the model.

Lightweight algorithms differ from each other on the basis of block size, key size, and their construction. S-box is the major section of any cryptography algorithm. During encryption, S-box holds the responsibility of creating dubiety in the plaintext.

As a result, it concludes that the robustness of any cipher lies in S-box. S-box is responsible for a strong encryption and decryption system. A typical S-box also contains permutation box to increase confusion. There are many works done in order to make efficient S-box and find the S-box confusion strength [7]. Speed of cryptographic algorithm can also be increased by using parallel processing design [8].

An efficient cryptographic S-box has been the topic of research for many researchers and scholars. S-box used in AES is 8-bit S-box, and all 8-bits are in the form of 16 * 16 table, and it is used with big applications [9, 10]. Lightweight block ciphers such as PRESENT [11], mCrypton [12], KLEIN [13], LBLOCK [14], and HUMMING [15] are the examples of 4-bit S-box ciphers.

1.2 Structure of the Chapter

In this chapter, we have proposed multiplexer-based substitution box architecture for PRESENT cipher having 64-bit datapath architecture. The contributions of this chapter are as follows:

- Proposed a hardware structure with decreased number of slices which makes our architecture as an area efficient.
- Also, maximized throughput which makes it a good performance architecture with combinational logic.

The rest of the chapter is organized in the following way. Part II briefly defines the algorithm used to implement PRESENT cipher. In Sect. 3, the proposed architecture for substitution box is elaborated. Part IV consists of design methodology for the implementation. Section 5 provides experimental result information and discusses the simulation outcomes and comparison. Part VI concludes the work.

2 Algorithm Overview

PRESENT cipher [11] is an ultra-lightweight block cipher used for limited resource environments. It has 64-bits block size with two variants. These variants have key sizes of 80 bits and 128 bits, respectively. The round function of PRESENT cipher is based on substitution–permutation network (SPN). It means that each round contains substitution box (S-box) and permutation box (P-box) as shown in Fig. 1. It has two parts, namely datapath and key scheduling.

2.1 Datapath

The hardware implementation of PRESENT cipher is shown in Fig. 1. Initially, 64-bit input plaintext adds with 64 bits of master key. Resultant signal is then passed through sixteen S-boxes. S-box has 4×4 input–output bits, which is most nonlinear components of lightweight algorithms. Furthermore, bit-wise permutation is used. The final ciphertext is producing after 31 iterations.

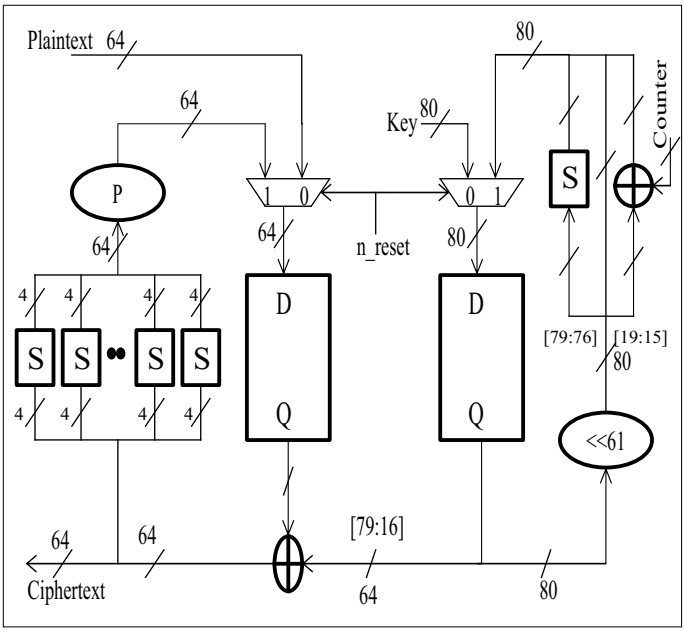


Fig. 1 Datapath and key scheduling algorithms of PRESENT cipher

2.2 Key Scheduling

There are two variants for key scheduling. One is for 80-bits, whereas other is for 128-bits. Only, 64 bits of master key adds with 64 bits of plaintext. Key scheduling algorithm includes bit shifting, S-box, and adding round constant opeartions. A 5-bit round counter and a 4-bit S-box are used for subkey generation.

The round counter is increasing its value by one for the next round. In addition, 61-bit left shifting operation is also used for providing additional confusion as shown in Fig. 1. Total, 31 subkeys are generated by this algorithm.

3 Proposed Architecture

S-box takes most of the cell area in any architecture. The most straightforward approach of implementation of S-box is LUT-based S-box, but it takes large area. Taking idea from the work [3], we have proposed complete combinational form of the PRESENT S-box having multiplexers and OR gates. Our proposed architecture takes comparatively less area with much speed performance.

First of all, we make truth table consisting of inputs and outputs of PRESENT S-box; then, Boolean expressions are used with Karnaugh map to form simplified Sum of Products (SoP) expressions for all output bits of the S-box [15]. It denotes the input bits as {S[3], S[2], S[1], S[0]} in this chapter, with S[3] being the most significant input bit. There are output bits as {Sout[3], Sout[2], Sout[1], Sout[0]}. The SoP expressions are given in Table 1.

Observing Table 1, it can see that realizing these Boolean expressions require large hardware size. Therefore, factorization is done to reduce the hardware space. The factorized expressions are shown in Table 2.

Based on the Table 1, the factorization in Table 2 is same for paper [16].

As seen from the S-box output expressions, the S-box can be easily implemented with the help of logic gates such as AND and OR gates. Apart from logic gates, it can also be constructed using multiplexers or decoders. A multiplexer or data selector is

Table 1 Output of PRESENT S-box in sum of product form

S-Box output	Expression
Sout[3]	(S[3]'S[2]S[1]) + (S[3]S[2]'S[1]) + (S[3]S[2]'S[0]) + (S[3]'S[1]S[0]) + (S[3]'S[1]'S[0])
Sout[2]	(S[3]'S[2]'S[1]') + (S[2]'S[1]S[0]') + (S[3]S[2]S[1]') + (S[2]'S[1]'S[0]) + (S[3]'S[2]S[1]S[0])
Sout[1]	(S[3]S[2]S[0]) + (S[3]S[2]'S[1]') + (S[3]'S[1]S[0]') + (S[3]'S[2]'S[1]) + (S[3]S[2]'S[0]')
Sout[0]	(S[3]'S[1]S[0]) + (S[3]S[1]S[0]') + (S[3]S[2]'S[0]') + (S[3]'S[2]'S[0]) + (S[3]'S[2]S[1]'S[0]') + (S[3]S[2]S[1]'S[0])

Table 2 PRESENT S-box output expression and factors used

Output bits	Output expression
Sout[3]	(f1)+(f2)+(S[0]f5)+(S[1]f5)+(S[2]f6)
Sout[2]	(S[2]'f7)+(f3)+(S[0]f8)+(S[3]'f8)+(S[2]f2)
Sout[1]	(S[0]f9)+(S[1]'f5)+(S[0]'f6)+(S[2]'f6)+(f4)
Sout[0]	(f2)+(S[3]f7)+(f4)+(S[3]'S[2]'S[0]) +(S[2]f1)+(S[0]f3)

Factor	Expression
f1	S[3]'S[1]'S[0]'
f2	S[0]f6
f3	S[1]'f9
f4	S[0]'f5
f5	S[3]S[2]'
f6	S[3]'S[1]
f7	S[1]S[0]'
f8	S[2]'S[1]'
f9	S[3]S[2]

the digital circuit which takes several inputs as analog or digital inputs and transmits it to single output line. Multiplexers also have select lines which are used to select one input at a time and transmit it to output. In this proposed work, twenty six 2 * 1 multiplexer and four OR gates are used to construct our substitution box as shown in Fig. 2. Table 3 is mentioning multiplexer numbers with inputs, selection key, and output.

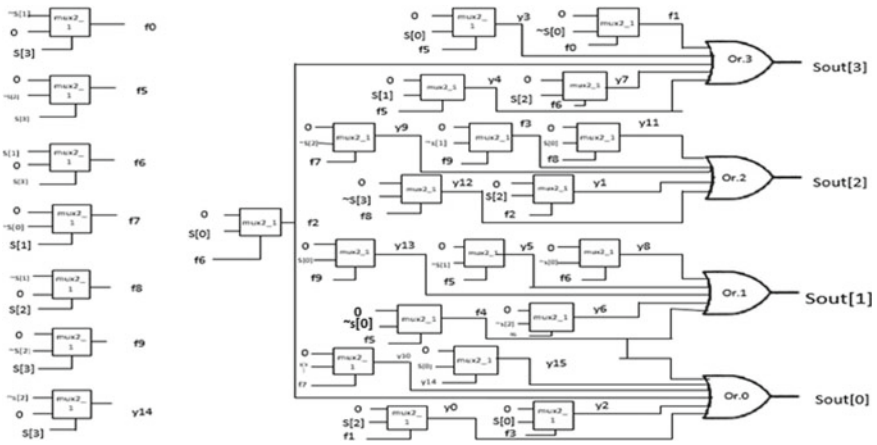


Fig. 2 Proposed Multiplexer-based S-box

Table 3 Different inputs and outputs of multiplexers used

MUX No	Input 1	Input 2	Selection key	Output
1	~ S[1]	0	S[3]	f0
2	0	~ S[0]	f0	f1
3	0	S[0]	f6	f2
4	0	~ S[1]	f9	f3
5	0	~ S[0]	f5	f4
6	0	~ S[2]	S[3]	f5
7	S[1]	0	S[3]	f6
8	0	~ S[0]	S[1]	f7
9	~ S[1]	0	S[2]	f8
10	0	~ S[2]	S[3]	f9
11	0	S[2]	f1	y0
12	0	S[2]	f2	y1
13	0	S[0]	f3	y2
14	0	S[0]	f5	y3
15	0	S[1]	f5	y4
16	0	~ S[1]	f5	y5
17	0	~ S[2]	f6	y6
18	0	S[2]	f6	y7
19	0	~ S[0]	f6	y8
20	0	~ S[2]	f7	y9
21	0	S[3]	f7	y10
22	0	S[3]	f8	y11
23	0	~ S[3]	f8	y12
24	0	S[0]	f9	y13
25	~ S[2]	0	S[3]	y14
26	0	S[0]	y14	y15

From f0 to f9 and y0 to y15 are 26 outputs of the multiplexers. Moreover, outputs of some multiplexers are also used as selection keys for another multiplexers.

All four OR gates take inputs from outputs of some of the multiplexers, for example, OR gate-3 takes inputs f1, f2, y3, y4, y7 whereas OR gate-2 takes inputs f3, y1, y9, y11, y12. In addition, OR gate-1 takes inputs f4, y5, y6, y8, y13 and OR gate-0 takes inputs f2, f4, y0, y2, y10, y15.

The output of OR gates is the output of S-box. The advantage of using multiplexers as combinational S-box is that it is easy to construct multiplexers and it also takes many common inputs. By using multiplexers, the proposed architecture reduces the integrated circuit packages required by any particular circuit design. As a result, this reduces the overall cost.

4 Methodology

4.1 Environment

The proposed PRESENT architecture is synthesized in *Xilinx* using ISE design suite 14.6, and simulation is done using *ISim* simulator. Spartan-6, Virtex-4 & Virtex-5 families are used for comparisons. Each family holds different number of LUTs, flip-flops and slices as well as different maximum frequencies and throughput values.

4.2 Measurement Parameters

The architecture implemented in FPGA is evaluated on the basis of various parameters such as area and speed. It evaluates performance behavior with the help of throughput and latency. Parameters like slices, LUTs, and flip-flops are used to define area of any structure. Latency is defined as number of executed clock cycles required for getting the first output. In cryptography, latency is described as the time period between the instants of initiating the plaintext input into a cipher and getting the corresponding ciphertext output.

Area size of any structure is responsible for the power consumption. As much as area increases, it will increase power consumption and vice versa. Throughput is a function of operating frequency. For evaluating throughput as performance parameter, it is defined by Eq. (1).

$$\text{Throughput} = \text{Block\_size } X \text{ Operating Frequency/Latency} \tag{1}$$

5 Results and Discussion

The proposed PRESENT cipher is implemented on the *Xilinx* Spartan-6, Virtex-4, and Virtex-5 FPGA platforms. This architecture achieves higher throughput as compared to all other designs for a particular device in Table 4. It gets throughput of 611.2 Mbps for Spartan-6 device, 1263.27 Mbps for Virtex-5, and 1077.05 Mbps for Virtex-4 device. The proposed design also obtains higher throughput per slice (in *Mbps/slices*) with values 9.26, 19.73, 5.55 for Spartan-5, Virtex-5, Virtex-4 devices, respectively. It shows highest maximum frequencies of 296.05, 611.9, and 521.7 MHz for Spartan-5, Virtex-5, and Virtex-4, repesctively. A total latency of 31 clock cycles are achieved for each device which is remarkably least as compared to other designs in Table 4. Despite of speed or performance parameters, the proposed designs got a respectable low number of slices of value 64 for Virtex-5 device.

Table 4 FPGA implementation results of different

Designs	Latency (Cycles)	Slice	LUTs	Flip-flops	FMax (MHz)	Throughput (Mbps)	Throughput/Slice (Mbps/Slice)
Spartan-6							
[17]	55	58	202	200	160.21	186.42	3.21
[17]	303	44	157	203	164.23	34.69	0.78
[18]	132	48	147	73	206.40	100.07	2.08
Proposed	**31**	66	239	149	**296.05**	**611.2**	**9.26**
Virtex-5							
[17]	55	88	283	200	271.67	316.12	3.59
[17]	303	72	237	203	289.69	61.19	0.84
[18]	132	75	182	73	321.96	156.10	2.08
Iterative [19]	47	87	285	200	250.89	341.64	1.19
Serial [19]	295	70	237	203	245.76	53.32	0.76
Boolean [16]	295	62	222	201	236.57	51.32	0.82
Proposed	**31**	**64**	213	149	**611.90**	**1263.27**	**19.73**
Virtex-4							
[17]	55	192	382	200	284.33	330.86	1.72
[17]	303	131	258	203	288.52	60.94	0.46
[18]	132	151	279	73	223.51	108.37	0.71
Proposed	**31**	194	318	149	**521.70**	**1077.05**	**5.55**

6 Conclusion

In this chapter, a high-speed architecture of the lightweight block cipher PRESENT on FPGA is proposed. It manages to have a respectable area using multiplexers-based S-box. As compared to related works, it can be observed that this design architecture achieved high performance in terms of speed or throughput, throughput per slice and maximum frequency for each device. Despite having high-speed performance for each device, it can be noted that this design also got respectable less area requirement valued at 64 slices for virtex-5 device as compared to other related works. Our proposed PRSENT cipher is optimum for high-speed applications.

References

1. Chui M, Löffler M, Roberts R (2010) The internet of things. McKinsey Q. 2:70–79
2. Barki A, Bouabdallah A, Gharout S, Traore J (2016) M2M security: challenges and solutions. IEEE Commun Surv Tutorials 18(2):1241–1254

3. Keoh SL, Kumar SS, Tschofenig H (2014) Securing the internet of things: a standardization perspective. IEEE Internet Things J. 1(3):265–275
4. Daemen J, Rijmen V, Leuven KU (1999) AES Proposal : Rijndael. Complexity, pp 1–45
5. Tu C (2011) Design of an improved method of Rijndael S-box. Commun Comput Inf Sci 231 CCIS(PART 1):46–51
6. Paar C, Pelzl J (2010) SHA-3 and The Hash Function Keccak. Underst Cryptogr—A Textb Students Pract
7. Hussain I, Shah T, Mahmood H, Gondal MA, Bhatti UY (2011) Some analysis of S-box based on residue of prime number. Proc Pakistan Acad Sci 48(2):111–115
8. Singh P, Acharya B, Chaurasiya RK (2019) High throughput architecture for KLEIN block Cipher in FPGA. In: 2019 9th Annual Information Technology, Electromechanical and Microelectronics Conference, pp 64–69
9. Daemen J, Rijmen V (2002) The design of Rijndael. New York, p 255
10. Nyberg K (1991) Perfect nonlinear S-boxes. Lecture Notes in Computer Science (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 547 LNCS, pp 378–386
11. Bogdanov A et al (2007) PRESENT : an ultra-lightweight block cipher. Springer, Berlin, pp 450–466
12. Lim CH, Korkishko T (2006) "mCrypton—a lightweight block cipher for security of low-cost RFID tags and Sensors. Inf. Secur. Appl. Springer, LNCS, vol 3786, pp 243–258
13. Gong Z, Nikova S, Law YW (2012) KLEIN: a new family of lightweight block ciphers. Lecture Notes in Computer Science (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 7055 LNCS, pp 1–18
14. Wu W, Academy C (2011) LBlock : a lightweight block Cipher LBlock : a lightweight block Cipher ⋆," February 2015, pp 327–344
15. Knudsen LR, Robshaw MJB (2011) The block cipher companion. A brief mindfulness intervention on acute pain experience: an examination of individual difference, vol 1, pp 1689–1699
16. Tay JJ, Wong MLD, Wong MM, Zhang C, Hijazin I (2015) Compact FPGA implementation of PRESENT with Boolean S-Box. Proceedings of 6th Asia Symposium Quality and Electronic Design. ASQED 2015, pp 144–148
17. Hanley N, O'Neill M (2012) Hardware comparison of the ISO/IEC 29192-2 block ciphers. In: Proceedings of 2012 IEEE Computer Society Annual Symposium VLSI, ISVLSI 2012, pp 57–62
18. Lara-Nino CA, Morales-Sandoval M, Diaz-Perez A (2016) Novel FPGA-based low-cost hardware architecture for the PRESENT Block Cipher. In: Proceedings of 19th Euromicro conference on digital system design. DSD 2016, pp 646–650
19. Kavun EB, Yalcin T (2011) RAM-based ultra-lightweight FPGA implementation of PRESENT. In: Proceedings of 2011 international conference reconfigurable computing and FPGAs, ReConFig 2011, pp 280–285

Effective Delay and Energy Conservation Strategies for Query-Driven Wireless Sensor Networks

Itu Snigdh and Nilesh Kumar Sahu

Abstract A sensor node is usually coupled with more than one sensor. This imposes more burdens on individual nodes in data gathering functions. Also, the energy consumption increases in order to power a greater number of transducers as well as in carrying out data communication and aggregation functions for different sensed parameters. This article proposes multi-tree and multiple-tree algorithms for improving network parameters in the usual data collection WSN application scenarios. A multi-tree simply suggests a strategy where single parameter-type sensors are arranged in a tree-based topology so as to reduce the energy consumption as well as the end-to-end delay in data communication. The multiple-tree exemplifies a more effective way of query result propagation in data centric applications where delay requirements are not very stringent.

Keywords Wireless sensor networks · Energy consumption · Delay · Query processing · Tree-based algorithms

1 Introduction

A wireless sensor network (WSN) has opened the challenge for distributed and cooperative computing and communication. According to Sohraby et al. [1], voluminous unstructured and heterogeneous data needs to be communicated and collected at the central information processing station for major WSN applications. A WSN comprising of homogeneous sensor nodes implies that every sensor node has equivalent capabilities. These usually are measured as their computational or communication capabilities. Seldom do we see in literatures that sensors are grouped according to the environmental parameters they monitor. A typical sensor network is equipped with nodes using many transducers to capture different but allied environmental

I. Snigdh (✉) · N. K. Sahu
Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Mesra, Jharkhand 835215, India
e-mail: itusnigdh@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_56

parameters, for example, temperature and humidity, vibration, and stress–strain. These sensors are responsible for measuring the parameters and forwarding it toward the base station in form of data packets.

These data packets may contain similar or a muddle of parameters. Researchers discuss many algorithms and strategies for routing these packets in the most energy efficient manner. As energy in a WSN is the prime concern, data aggregation also finds its place in literature given by Sun [2] and Arici et al. [3]. Let us assume a simple scenario where each sensor needs to monitor only one parameter and forwards it further. Also, further assuming that the single type of data gets routed via different nodes scattered in the WSN. We can now imagine that the rate of aggregation or the degree of convergence of data packets at the intermediary nodes would drastically increase if the parameters monitored are similar. Thus, all the packets could be merged effectively and the burden of data forwarding due to repeated transmissions would greatly reduce.

Further, these packets would be initiated only on the event of query initiated at the sink. Hence, it would result in lesser collision or contest for the channel thereby reducing the end-to-end delay. The data and queries in a query-driven WSN are forwarded between the base station and the location where the target phenomenon has occurred or is observed as cited by Georgios et al. [4]. The first approach is a single hop approach where each sensor node directly communicated with the base station or the sink. The major limitation of this approach is severe energy depletion of the farther nodes; this eventually causes profound limitation in the lifetime of the network. We know that radio signals require a lot of power. Unlike messages running through wires, they decay in an accelerated manner. Also, as sound and radio decay according to the inverse square law, on doubling the distance; we require four times the amount of power. This shortcoming of the direct communication approach is overcome by multi-hop packet transmissions over short communication ranges. This approach saves energy considerably and reduces the communication interference among the nodes competing for channel access.

Energy cost optimization has been suitably catered to in spanning tree protocols stated by Daniele et al. [5], Wang et al. [6], and Johansson et al. [7], for short range communications. In addition, shortest path traversals guarantee lifetime enhancement and congestion management has been suggested by Gnawali et al. [8] and Chakraborty et al. [9]. Effective utilization of limited resource is done by topology construction and control methods using cluster heads and middle heads also. However, the cluster head election and frequent rotation of roles turn out to be computationally and energy intensive. Focusing on data collection systems, the requirement is to balance the traffic distribution thereby reducing the hotspots against route failures. Techniques to reduce the individual workload of the nodes require distributing optimally the number of messages to be processed as suggested by Hossein and Moghavvemi [10] or use data aggregation before forwarding messages given by Joohwan et al. [11]. Similar data gathering illustrated by Yingchi et al. [12],Gong and Jiang [13],are based on balanced tree topologies that work for minimization of data collision and overall network energy consumption.

A simple strategy that could be adopted is to dedicate each sensor to sense exactly only one parameter and thus reduce the workload burden. Analyzing this structure becomes easier if we assume that the sensors communicate only when a query requiring their parameter is initiated. It also requires that either sensors connect to the other 'same type' sensors or the network balances equivalent proportions of 'different sensors' in the region of interest. The architecture in discussion can be visualized as a backbone-based architecture that can be realized in two ways.

A WSN is established where nodes of one level communicate to a higher-level node auxiliary linked with the sink. This structure comes analogous to the cluster-based strategy presented in Jin et al. [14], Younis and Fahmy [15] with the constraint that the higher-level nodes are fixed and the communication path is nonvolatile. This constraint further motivates us to simplify our scheme and present energy consumption analysis for tree-based backbone networks.

The proposed multi-tree-based architecture was given for heterogeneous sensor nodes by Snigdh and Saurabh [16], in which one node connected with three different types of nodes for tree generation. During query processing, node having three child sensor nodes have more data to transfer to its parent node and that node must be active for all type of query. So, its energy depreciation is more as compared to leaf nodes and it will become a bottleneck to architecture.

The other alternative backbone structure is a multiple-tree-based architecture where nodes of one type associate together and form a sub-tree to the sink. Hence, for each different class of nodes, we have a sink node responsible for communicating with the base station. For analyzing the performance of such a strategy, we consider two important parameters, i.e., the end-to-end delay and the residual energy. We have two strategies to generate a backbone network comprising of heterogeneous nodes applied to a structural health monitoring applications. The process of query generation and its response forwarding can be well depicted in Fig. 1.

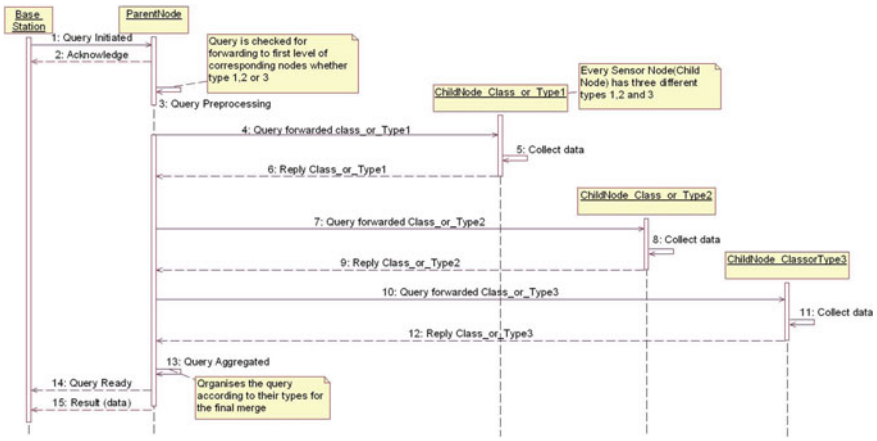


Fig. 1 Query processing in multi-tree strategy

2 System Model

The multi-tree and multiple-tree algorithm is given for receiving more than one type of information about a given area using heterogeneous sensors. This strategy uses query routing tree for data gathering and transmission. The algorithm starts by tree generation, which is initiated by sink. Sink broadcasts a message in its one hop range to get the Euclidian distance of neighboring sensor nodes and then connect to one sensor of each type, which is nearest to it. Connected nodes will again broadcast a similar message in their one hop range and connect to each type of node. The process repeats for all connected sensors and establishes the network topology. After constructing the topology using multi-tree or the multiple-tree algorithms, data is collected by the heterogeneous sensors.

In the multi-tree strategy, the sink node multicasts query given by the user to all the sensor nodes in the connected tree as shown in Fig. 2.

The query can be of temperature type, accelerometer/vibration type, or stress/strain type. On receiving the query, the first-level sensor nodes simply forward it to their child nodes. If the query is of a particular type, only the 'same' type sensor nodes respond to their parents. Other two types of sensor nodes do not in communicating their collected data however they forward any data received from child nodes to their corresponding parent if the destination ID is the sink. Further, the non-leaf nodes append multiple data received before forwarding it to the sink. Figure 3a shows the processing of query for temperature-type query. In this figure, after receiving the temperature-type query, sensor node '*T*' senses the data and transmits it to their parent. Other two type of sensors, i.e., sensors '*A*' and '*S*' only forward the data to their parent. Similarly, Fig. 3b, c shows the query processing for accelerometer/vibration-type and stress/strain-type query. The multiple-tree in contrast to the multi-tree algorithm in connects to only similar type of nodes.

For query processing, the sink multicasts query of desired type at regular intervals. The sensors connected to sink receive query and transmits it to next-level sensors in the tree if query parameters match with their node type. This reduces the unnecessary forwarding of data if the node is not concerned with the parameter required in the query. Therefore, query propagates in the only tree having node type same as the

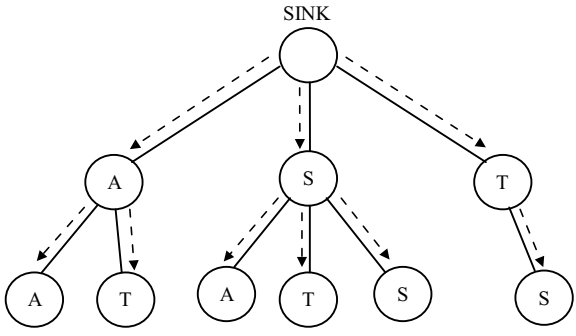


Fig. 2 Multi-tree architecture

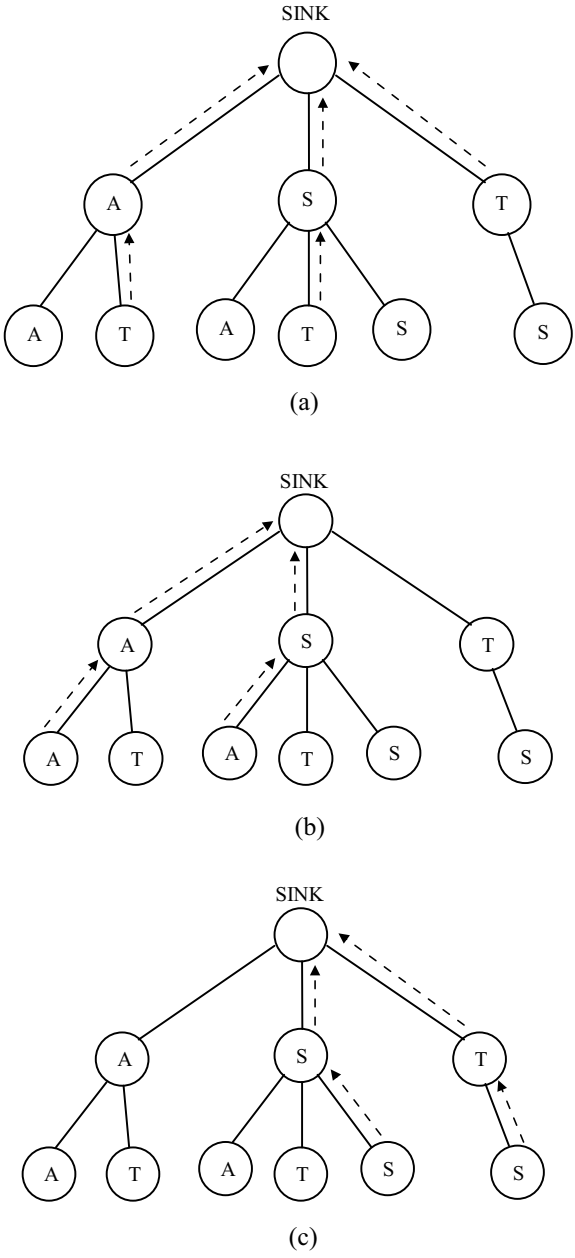


Fig. 3 Data processing **a** Query type Q_T, **b** Query type Q_A, **c** Query type Q_s

query type. After query propagated through all the nodes in tree, they respond by sending gathered data to the sink. Other types of sensors simply sleep with periodic listening to the channel for any message. Hence, we ensure that the sensors of each type respond to the query of their interest only.

2.1 Multi-tree Algorithm

/***Deployment of nodes***/

1. Random Deployment of heterogeneous node.
2. Begin
3. Node_List[0] = Sink;
4. K=1;
5. For(i=0 to N-1)
6. {
7. For(j=i to N-1)
8. Dist[j] = Calculate $ED_{i,j}$;
9. Sort($ED_{i,j}$);
10. Connect Node_List[i] and Array_T[i];
11. Node_List[k]=Array_T[i];
12. K++;
13. Repeat STEP 5 to STEP 10 for Array_A[];
14. Repeat STEP 5 to STEP 10 for Array_S[];
15. }
16. End

2.2 Multiple Tree Algorithm

/***Tree Generation***/
/*…Initialization…*/

$V = V_A \cup V_S \cup V_T$;//total node, three type of node A, S and T.
V_a= {}; //set which store connected node of A type sensor node
V_s = {}; //set which store connected node of S type sensor node
V_t= {}; //set which store connected node of T
type sensor node

1. begin
2. Sink finds one minimum distance node from each V_A, $V_S$and V_T and store it into V_a, V_s, V_t respectively
3. Until V_a != V_A
   V_a finds minimum distance node from V_A-V_a, connect to it and store into V_a.
4. Until V_s != V_S
   V_s finds minimum distance node from V_S- V_s,connect to itand store into V_s.
5. Until V_t != V_T
   V_t finds minimum distance node from V_T-V_t connect to it and store into V_t.
6. end

/* Query Processing */

1. Sink generates query
2. Sink multicasts the query to member nodes
3. Tree which have node responsible for query generated will propagate query to next level node.
4. node will forward result to respective parent node until it reaches to sink node
5. Compute the energy consumption using the energy consumption function based on the radio energy model as referred to by Wang and Yang (2007).

```
For(Query_T)
 {
   {
      If (Node_List[i] in  Intermediate_node_T[])
Then
 Econsumed = Eqr + mEqt + m.Edr + (m+1)Efusion + Edt ;
If (Node_List[i] in Intermediate_node_S[] or Intermediate_node_A[])Then
 Econsumed = Eqr + mEqt + mEdr + mEfusion + Edt ;

If(Node_List[i] in Leaf_node_T[]) Then
 Econsumed = Eqr + Edt ;
 If(Node_List[i] in Leaf_node_A[] or Leaf_node_S[]) Then
 Econsumed = Eqr ;
                }
 For(Query_A)
                Repeat Step 5 to Step 14.
 For(Query_S)
                Repeat Step 5 to Step 14.
 }
 End
```

2.3 Assumptions

The following assumptions have been made into account, while implementing the proposed multi-tree algorithm:

1. Sensor nodes are randomly dispersed in a field.
2. Sensor nodes and sink are stationary, and sink is located the centre of the monitoring field.

3. Sensor nodes are heterogeneous and location aware.
4. Sensor nodes have limited and irreplaceable battery power.
5. Sensor nodes have the function of data fusion and can aggregate N data packets into a packet to minimize the total energy usage.
6. The tree formed under a constraint that each sensor will connect to only same type of sensor node within their communication range.
7. The simulation parameters considered are:

    a. Each sensor powered by 2 J,
    b. Transmission packet size (k) is 1000 bit,
    c. Electronic energy Eelec is 50 nJ/bit,
    d. Node classes: Temperature, pressure, humidity type.

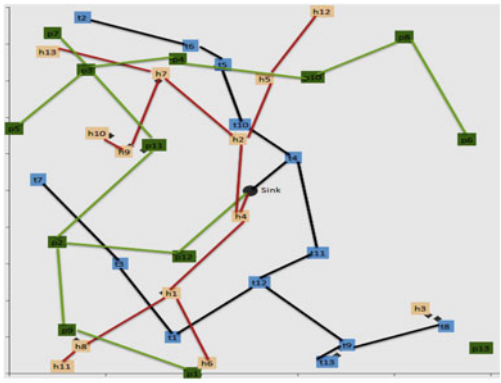
3 Results and Discussions

The proposed multiple-tree-based architecture, multi-tree-based architecture and minimum spanning tree-based architectures are compared for analysis. The client–server architecture is implemented in JAVA to evaluate the query processing parameters that are mainly the residual energy of each node and the delay time for each query response. Figure 4a–c are the snapshots of the tree topologies constructed while executing the algorithms on JAVA interface. These figures depict the communication path followed while communication backbone is generated in all the three respective trees. An MST is implemented by the standard Prim's algorithm given by Jarník [18] simply to compare the shortest path routes that would be taken by any routing algorithm for message delivery.

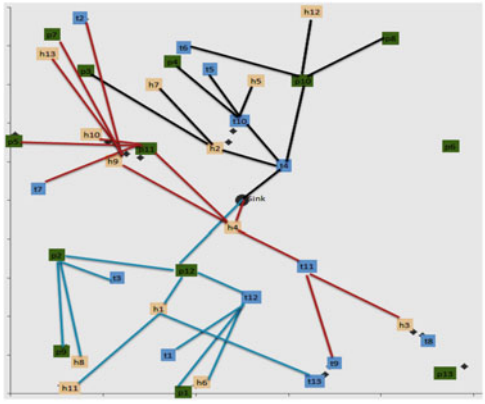
As the main constraint in a backbone-based network is single point of failure. For any ad hoc network, more work needs to be done by the nodes nearer to the sink or base station. Thus, these nodes become the 'Hot Spots' whose failure affects the entire network operation drastically. Hence, the first analysis focuses on the energy consumed by the first-level nodes or simply the nodes within one hop distance connected directly to the sink. The second-level nodes are the intermediate-level nodes that are not leaf nodes. Hence, these nodes also carry the additional task of data forwarding and aggregation.

Figure 5a shows the energy consumption of the level 1 nodes of a multi-tree, multiple-tree, and a corresponding MST. Level 1 nodes are the nodes within the one hop communication range of the sink and hence have to perform the maximum data forwarding. It is observed that the multiple-tree has the best performance as compared to the other two strategies considered. Figure 5b shows the energy consumed by the non-leaf nodes. It is also seen that multiple-tree is still the best as compared to multi-tree or MST-based topologies.

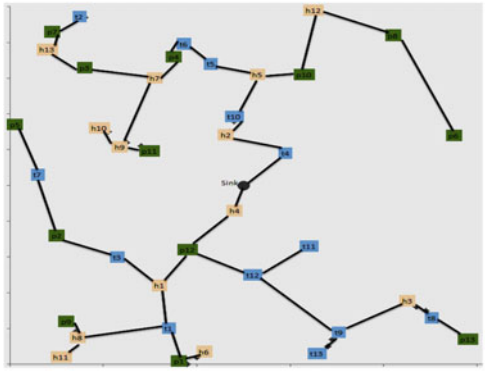
Figure 6a–c shows the average energy consumption of all the nodes iterated over 6,00,000 random queries. The proposed algorithms are compared for different communication ranges namely, 25, 30, and 35. The idea behind taking different communication ranges is to alter the topology as more number of nodes qualifies to



(a)

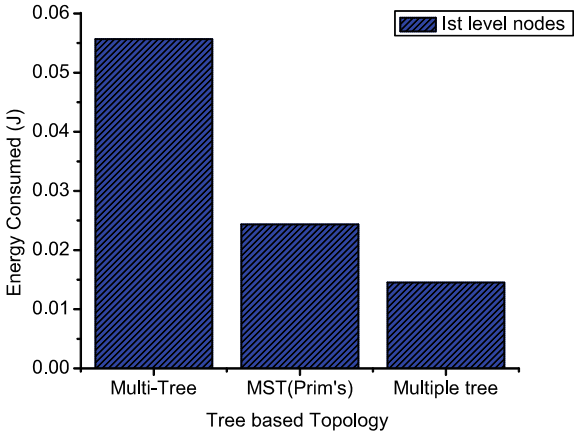


(b)

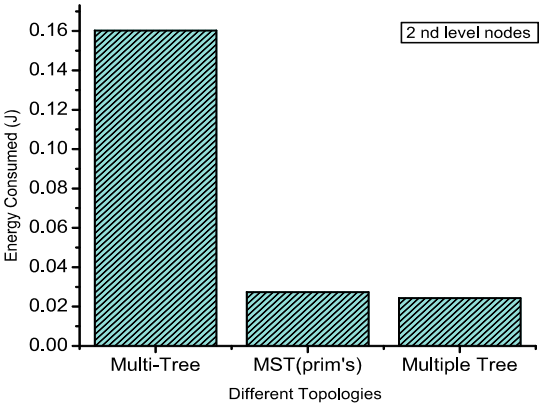


(c)

Fig. 4 Formation of route topologies. **a** multiple-tree, **b** multi-tree, **c** Prim's MST



(a)

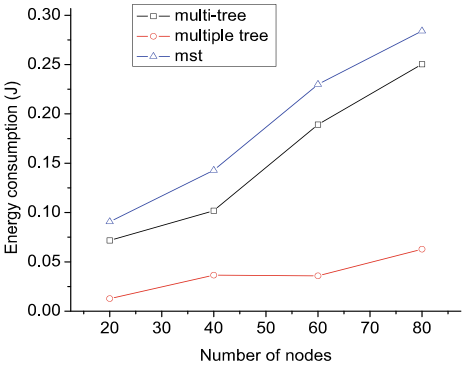


(b)

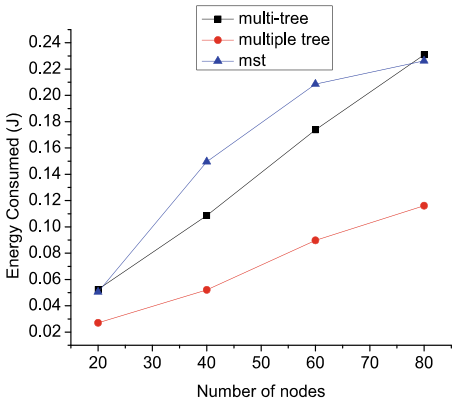
Fig. 5 Energy consumption. **a** level 1 nodes, **b** non-leaf nodes

serve as non-leaf nodes and the connectivity is better analyzed. It is observed that even while we increase the density of nodes in the region of interest and communication range, the energy consumption on an average is significantly low for multiple-tree topology. Even the multi-tree performs better than a simple MST-based topology on an average.

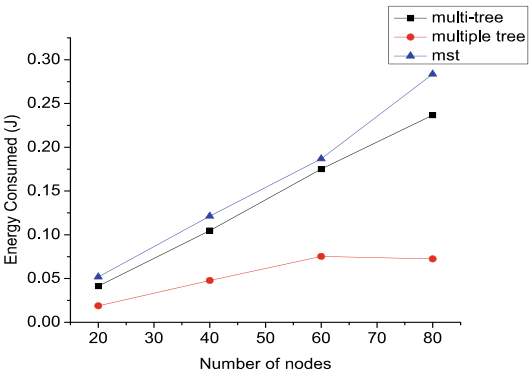
For analyzing the delay incurred in processing 6,00,000 random queries, only the multi-tree and multiple-tree topologies have been depicted in Fig. 7. Figure 7a–c considers five different random deployments and varied communication-based topologies. We confirm that if a WSN application requires stringent delay constraints in query response, multi-tree is better than the multiple-tree topology. Thus, multi-tree may be chosen over multiple-tree topologies if the delay is more important than energy conservation.



(a)

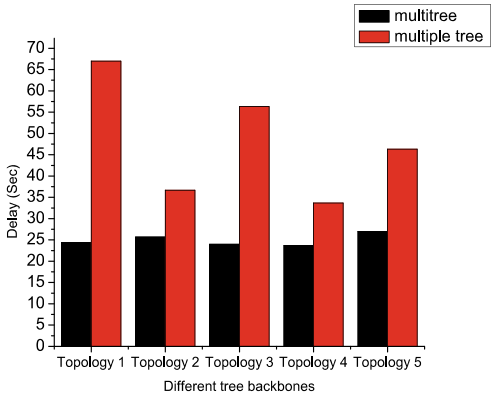


(b)

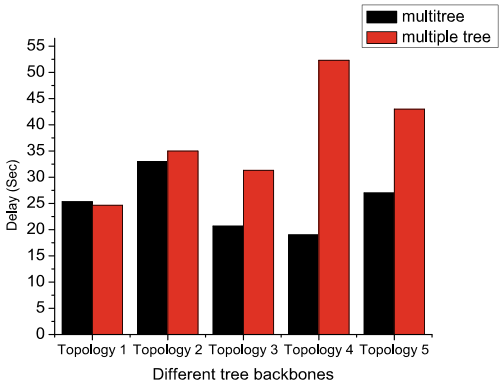


(c)

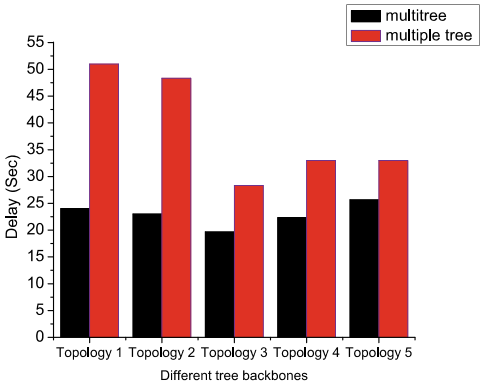
Fig. 6 Average energy consumption of nodes. **a** communication range: 25, **b** communication range: 30, **c** communication range: 35



(a)



(b)



(c)

Fig. 7 Average delay in query response.
a communication range: 25,
b communication range: 30,
c communication range: 35

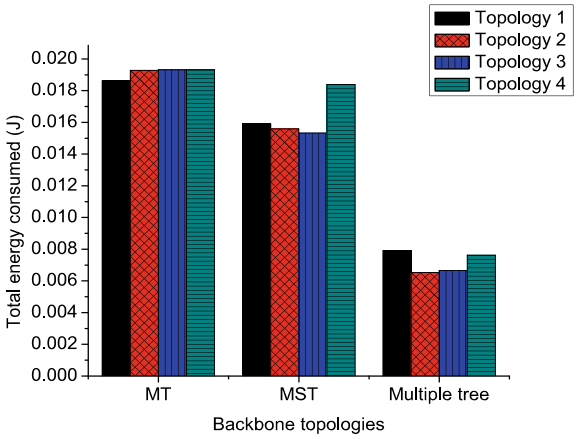


Fig. 8 Total energy consumed

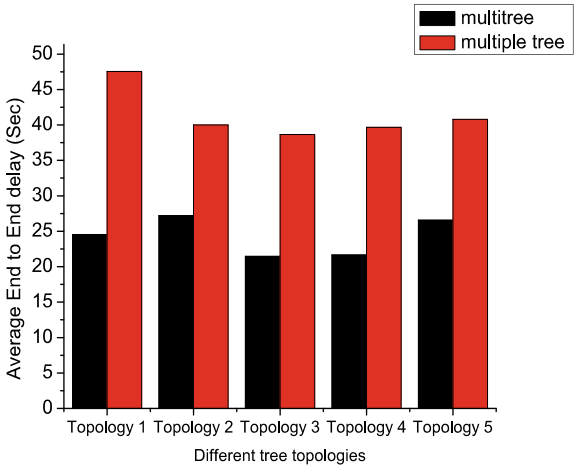


Fig. 9 Average end-to-end delay

Lastly, we present the total energy consumed for all nodes under different random deployment scenarios in Fig. 8. Figure 9 outlays the average end-to-end delay in query response over the different topologies. The lifetime of the network as to when it registers the first-node dead statistics is depicted in Fig. 10. The total number of queries serviced predicts the minimum time when the tree-based network would require restructuring for further operations.

4 Conclusion

From the simulation results, we validate our claim as to the reduction in the energy consumption and the delay in query response. We analyze the performability of

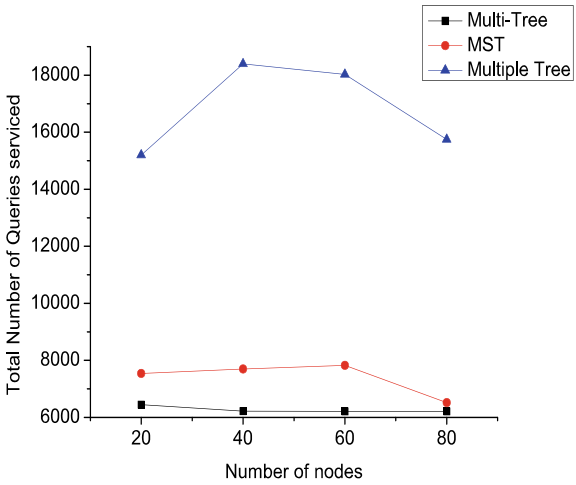


Fig. 10 Total queries serviced

a sensor network in terms of data communicated in structural health monitoring applications which is usually redundant. The energy comparison depicts that the multiple-tree strategy shows improvement over a multi-tree strategy. Also, there is a significant energy saving under the same network scenario in comparison to the MST backbone-based communication architecture. In contrast, the delay is better for multi-tree-based communication as compared to multiple-tree strategy. Hence, multi-tree-based topology can be used for real-time sensor network applications that require faster communication. On the contrary, multiple trees may be adopted where energy constraints are stringent. Also, the following results depict the prospects of using multiple sensors for reduced data latency. It can further be addressed to improve the reliability of networks. Also, the sleep cycle scheduling can be implemented to conserve energy in a more effective and organized way.

Acknowledgements This paper is based on the preliminary results of M. Tech students S. Saurabh and S. Surani.

References

1. Sohraby K, Minoli D, Znati T (2007) Wireless sensor networks: technology, protocols and applications. Wiley, Hoboken
2. Sun J-Z (2008) QoS aware query processing algorithm for wireless sensor networks. J Comput 3(11):32–41
3. Arici T, Gedik B, Altunbasak Y, Liu L. PINCO: a pipelined in-network compression scheme for data collection in wireless sensor networks. In: 12th international conference on computer communications and networks, pp 539–544
4. Georgios C, Zeinalipour-Yazti D, Gunopulos D (2010) Minimum-hot-spot query trees for wireless sensor networks. In: Proceedings of the ninth ACM international workshop on data engineering for wireless and mobile access. ACM, pp 33–40

5. Daniele M, Ortolani M, Re GL (2007) A network protocol to enhance robustness in tree-based WSNs using data aggregation. International conference on mobile adhoc and sensor systems, (MASS 2007), IEEE, pp 1–4
6. Wang J, Jee-Hyong K (2012) ViTAMin: a virtual backbone tree algorithm for minimal energy consumption in wireless sensor network routing. IEEE international conference on information networking (ICOIN), pp 144–149
7. Johansson T, Osipov E, Carr- Motyckova L (2008) Interference aware construction of multi- and convergecast trees in wireless sensor networks. Next generation teletraffic and wired/wireless advanced networking. Lect Notes Comput Sci 5174:72–87
8. Gnawali O, Rodrigo F, Kyle J, David M, Philip L (2009) Collection tree protocol. In: Proceedings of the 7th ACM conference on embedded networked sensor systems, 2009, pp 1–14
9. Chakraborty S, Nandi S, Karmakar S (2011) A tree-based local repairing approach for increasing lifetime of query driven WSN. 14th international conference on computational science and engineering (CSE), IEEE 2011, pp 475–482
10. Fariborzi H, Moghavvemi M (2009) EAMTR: Energy aware multi-tree routing for wireless sensor networks. IET Commun 3(5):733–739
11. Joohwan K, Lin X, Shroff NB, Sinha P (2010) Minimizing delay and maximizing lifetime for wireless sensor networks with anycast. IEEE/ACM Trans Netw (TON) 18(2):515–528
12. Yingchi M, Xiaofang L, Yi L (2010) Workload based query routing tree algorithm in wireless sensor networks. In: International conference on computational intelligence and software engineering (CiSE), pp 1–4, doi: 10.1109/ICISE.2010.5691166
13. Gong B, Jiang T (2011) A tree-based routing protocol in wireless sensor networks. In: International conference on electrical and control engineering (ICECE), pp 5729–5732
14. Jin Y, Wang L, Kim Y, Yang X (2008) EEMC: an energy efficient multi-level clustering algorithm for large-scale wireless sensor networks. Sci Direct Comput Netw 52:542–562
15. Younis O, Fahmy S (2004) HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mob Comput 3(4):366–379
16. Snigdh I, Saurabh S (2014) A multi tree based approach for performance analysis in hierarchical wireless sensor networks. Int J Digit Inf Wirel Commun (IJDIWC) 4(4):438–446
17. Wang Q, Yang W (2007) Energy consumption model for power management in wireless sensor networks. In: 4th Annual IEEE communications society conference on sensor, mesh and ad hoc communications and network (SECON) 2007
18. Jarník V (1930) O jistém problému minimálním [About a certain minimal problem]. Práce Moravské Přírodovědecké Společnosti (in Czech) 6:57–63

Segregation of Speech and Music Signals for Aiding the Hearing Impaired

Ankit Ranjan, Shubhankar Sinha, Anuranjana, and Deepti Mehrotra

Abstract Separation of voice and music is an existing as well as an interesting problem but can also be a challenging task; at the same time, it is of immense importance for various other researches such as audio content analysis. This study is inherent investigation of the implementation of sound segregation in hearing aid improvement. The prime idea of the study is to help and improve the knowledge in hearing disabled children. Therefore, we have studied the issues dealt with physically impaired child and compared to the responses of normal individuals. In this study, the focus is on implementing the four algorithms collectively for efficient as well as detailed sound production. The algorithms used in the study are REPET, adaptive REPET, SVS-RPCA, MLRR.

Keywords Source separation · Audio · Quality assessment · Objective measure

1 Introduction

Human beings have an advanced hearing perception of sounds. A human being can listen and differentiate between various instrumental notes, tones, voices, and

The original version of this chapter was revised: The author name "Ankita Ranjan" has been corrected as "Ankit Ranjan". The correction to this chapter is available at https://doi.org/10.1007/978-981-15-7486-3_71

A. Ranjan (✉) · S. Sinha · Anuranjana · D. Mehrotra
Amity University Uttar Pradesh, Noida, Uttar Pradesh, India
e-mail: ankitranjan621@gmail.com

S. Sinha
e-mail: shubhankar2002@gmail.com

Anuranjana
e-mail: aranjana@amity.edu

D. Mehrotra
e-mail: mehdeepti@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_57

noise in a complex auditory scene. One of the most important skills is the ability to decide what is considered relevant in the audio signal and what is not and then focuses on the relevant part while disregarding the rest [1, 2]. In general, de-noising problem is the problem of separating different sound sources, where the "signal" is the important part of the audio stream, and the noise is everything else. The individuals having suffered from hearing loss/profound hearing face problem of discriminating the various sounds that they come across. Various hearing aids and cochlear implant have become an easy solution for these kinds of individuals. The child or individual starts receiving digitized signals and hence starts perceiving them. The clarity of the sound depends on how fine the hardware is and how well it is tuned with the child implanted with cochlear [3, 4].

Music–voice separation refers to the problem of trying to separate vocals from the instruments in a song in order to produce acapella track containing only vocal's and an instrument track containing only instruments. A number of algorithms are proposed in the literature for designing the sound filters for sound segregation. In this paper, various approaches are reviewed and among them, four most recommended approaches are discussed in detail [5].

To evaluate the performance of the algorithm, there are many toolboxes present in the MATLAB, like BSS_Eval (Blind Source Separation) for audio source separation and PEASS (Perceptual Toolbox) toolbox.

BSS_Eval [6] is a MATLAB tool to check the execution of (Blind) source partition calculations inside the assessment structure where the first source sign is accessible as input signals. The measures depend on the decomposition of each estimated source signals into various parts, comparing to the objective source. The PEASS toolbox provides a set of perceptually motivated objective measures for the evaluation of audio source separation. Similar to BSS_Eval, the distortion signal is decomposed into three components: overall perceptual score (OPS), target perceptual score (TPS), interference perceptual score (IPS).

Noise in the sound sample can be segregated using various filters like Kalman filter, Butterworth filter, and RLS filter. To increase the efficiency of the project, we adopt the Butter worth filter for the noise removal because the loading time of Butterworth filter is less as compared to other filters. The Butterworth filter [7] is a signal processing filter which has a flat frequency response as conceivable in the passband. The spectral subtraction strategy utilized here is a straightforward viable technique for noise diminishment. In this technique, a normal signal spectrum and average noise range [8] are assessed in parts of the recording and subtracted from each other, with the goal that sound-to-noise ratio (SNR) [5] is moved forward. It is accepted that the signal is twisted by a wideband stationary, added substance commotion, the clamor gauge is the same amid the investigation, and the reclamation stage is the same in the first and reestablished flag.

2 Related Work

The issue of REPET [1] is basically for redundancy which is center standard in music. Numerous melodic pieces are specifying by a basic rehashing structure over which shifting components are superimposed. This is particularly valid for the pop melodies where an artist regularly overlays fluctuating vocals on a rehashing backup. A novel and straightforward approach is for isolating the rehashing "frontal area" in a mixture. The fundamental thought to distinguish the intermittently rehashing sections mode got from them and concentrate the rehashing designs by means of time-recurrence concealing. Adaptive REPET [2] is intended to work on full melody, by applying REPET to neighborhood time windows of tune, with a given cover timescale for the windows.

Specifically, while REPET expect that instruments contain some segment which rehashes for the whole melody. Adaptive REPET expect that this rehashing bit may change in different parts of the melody (e.g., the verse and the tune may have distinctive rehashing foundation). SVS_RPCA [3] accept that the monotonous way of the instrument components relegates them to a low-rank subspace, while the voice is sparse and has more variety. It actualizes the robust principal component analysis strategy of a matrix factorization calculation for disintegration into a low-rank matrix. MLRR [4] endeavors to enhance SVS_RPCA calculation by rather utilizing deterioration into the low-rank matrix and one sparse matrix. The two low-rank matrixes speak to the subspace of sound relating to the voice and instruments, and the sparse matrix contains deviations from these subspaces.

Nonnegative matrix factorization (NFM) [9] is mainly used for the de-noising of speech signals in the non-stationary signals. Non-stationary noise is mainly created due to cellular phone and other telecommunication devices. Standard approach for segregating these types of signals is spectral subtraction and wiener filtering. They mainly restricted to stationary or quasi-stationary noise in practice.

Reverberation cancelation [10] is mainly used in adaptive filtering and noise cancelation that mainly focuses on removing the reverberation from the sound sample. It is desirable to develop a strategy for post-processing environment. If the system removes all the reverberation, but at the same time, it also destroys the character of the recorded instruments.

Speech enhancement [11] is mainly focused on the improvement of speech quality and intelligibility using various algorithms and techniques. Sound samples always contain the background noise. Background noise can be decreased by using spectral subtraction, cepstral mean subtraction, and Kalman filtering. Among all of them spectral subtraction is the oldest one of the first algorithms proposed for removal of background sound.

Bayesian filter [12] basically estimates a dynamic system's state from a noisy observation. Stochastic filtering technique is briefly reviewed with the emphasis with the help of nonlinear and non-Gaussian filtering. Under linear quadratic Gaussian circumstances, the desired Kalman filter can be created within the Bayesian framework. Bayesian filter mainly focused on the particle filtering.

Gaussian prior [13] provides the kernel support for many applications that fit the essential part of the Gaussian. It is basically a square support; i.e., in the case of 2D filtering, it can be decomposed in the form of 1D filtering for rows and columns. When the radius of the filter is small, then it is the fastest way to calculate the filtering result is direct 1D convolutions.

The difference of the geographic data can be analyzed by using the ANOVA technique. Spatial data basically contains different kinds of spatial information. Geographic information system provides the method that helps in the extracting the desired information (Table 1).

The above-mentioned methods have been analyzed thoroughly and extensively before deciding upon the current suite of techniques, namely REPET, adaptive REPET, SVS-RPCA, and MLRR. These methods have been chosen for the sole purpose of reducing the overall complexity of the software to such an extent that it becomes easier to both understand and execute properly for the end-user. The algorithms also have an individually low time complexity on their own, so any processing on their behalf is completed seamlessly in the background while putting minimal stress on modern hardware. After compilation, this aids in the successful execution of the entire suite even while running other programs simultaneously despite the heavy nature of MATLAB.

3 Methodology

The methods used for segregation and separation process of sound and music are discussed below.

REPET: It is implied as a "foreground/background" separation for well-known music. It takes interest in the sort of reiteration that is observed in the music. Specifically, numerous famous melodies contain instrumental components that repeat occasionally for the whole of the tune, similar to a rehashing harmony movement or like a circular drum track. REPET scans for these rehashing components that endeavors isolating rehashing "background" materials from the "foreground" materials, that is non-rehashing and by and large comprise of the vocal lines, and other non-rehashing instrumental lines which emphasize and give the tune [1].

Procedure for REPET.

1. Apply short-time Fourier transform (STFT)
2. Inverse STFT
3. Autocorrelation function using fast Fourier transform according to Wiener–Khinchin theorem
4. Beat spectrum using autocorrelation function and the calculating repeating period from the beat spectrum
5. Repeating mask from the magnitude spectrogram and the repeating period.

Table 1 Analysis of different algorithms for sound segregation

Algorithim	Features	Advantages	Disadvantages
REPET	It analyses the repeating sound structures over a defined time frame To avoid these pitfalls, it shortens the interrupt to accommodate smaller instances	It carefully takes the repeating structure of the song It also runs faster as compared to the rest of the music separation algorithm Mainly focus on analysis the repeating sound structures over a defined time frame and then match them with other similar instances but for a larger sample	REPET does not differentiate between various voice and music like melodious, non-repeating instrumental, and vocal line. This might become undesirable for voice segregation More noise is generated due to the gaps in processing
Adaptive REPET	There is no defined time frame for detection, it works for the background sound This is due to the algorithm overcompensating for not being able to identify the modulation by reducing the overall time threshold	Smaller instances are recognized more easily, leading to clearer signal Background sound is clearer compared to REPET	The adaptive REPET does not work for gradual or volume changes. If there is frequent changes in instrumental frequency, during the song, then also this algorithm is not suitable for such types of music Users provide input to help determine the repeating section(s). It becomes difficult when large number of songs repetition part is considered for automatic processing
SVS-RPCA	It classifies the singer's voice tone into three clear distinctions—soprano, alto, and tenor	SVS-RPCA is a separation algorithm which requires only original song from the user as input, makes automatic processing easier SVS-RPCA algorithm is useful for instances of pitch variation	There is no focus on background sound here, so this method can be used only if we wish to discard the background music as "noise"

(continued)

Table 1 (continued)

Algorithim	Features	Advantages	Disadvantages
MLRR	By using MLRR, we can separate the instrumental sound	Better features than SVS-RPCA.Blind separation is not there, more computation time an online dictionary algorithm required	The vocals suffer degradation as it isn't processed properly by MLRR
NMF	Nonnegative matrix factorization uses training data to identify speech patterns and separate them from the background sound	Since it has a variety of data classes to mine and chooses a suitable pattern, it provides good foreground separation	The background sound, however, suffers as it is given no emphasis whatsoever by the algorithm and cannot be worked with as a result
Reverberation filter	Uses a modified, multi-channel least mean square algorithm to suppress noisy channel creep	Achieves unprecedented word error reduction rate of 41.9%, hence being very useful in broadcasts	Mainly used in multiple speeches only channels, as it has little to no effect on musical noise
Spectral subtraction	Uses masking and gain-based extraction to siphon the background sound out of the channel specified by the user	Background sound is very operable and can be manipulated quite easily	Foreground sound suffers from quite a lot of artifacting in multi-channel application
Bayesian filtration	Creates a convergence mask on the source signal using angular spectrum-based GCC-PHAT metrics	Due to a convergence algorithmic approach, results are more accurate	Time complexity is quite large compared to alternatives
Gaussian prior	Uses Laplace distribution, independent vector analysis and an online expectation–maximization algorithm	Independent vector analysis realizes the best SIR ratios by far	Interference ratio is the only component being addressed here
ANOVA	Genre-based variation analysis using time–frequency masking	Repetition is the core feature of this approach and works well with slower tempo tracks	Fast tempo tracks with little repletion such as country music suffers from this algorithm

SVS-RPCA: It executes a calculation from Candès, called robust principal component examination. The algorithm uses the repetitive nature of instrumental components, which consign to a low-ranked space, also the voices are sparse and have more variety [2]. Consequently, the principal component analysis (PCA) technique, a matrix factorization calculation for disintegration into low-rank and sparse lattices, was done by Huang et al. [5].

Procedure for SVS-RPCA.

1. Input dataset
2. RPCA masking initialization and sparse pattern recognition
3. Low-rank representation
4. Voice estimation using SDR BSS_EVAL toolbox
5. Final output obtained.

MLRR endeavors to enhance the SVS-RPCA calculation by rather utilizing a deterioration into two parts: one is two low-rank matrices and the other sparse matrix. The two low-rank matrices speak to subspaces of sounds relating to the voice and instrumentals, and the sparse matrices contain the deviations from these various subspaces.

Procedure for MLRR.

1. Input dataset
2. Lowest rank matrix formation and ALM optimization
3. Singular value thresholding
4. Matrix decomposition using online dictionary
5. Subspace structuring.

The various algorithms discussed above require specific parameters along with their specifications that are provided by the users. MLRR algorithm requires the user to provide an instrumental track for the separation process. The REPET algorithm specifies a time interval from the user which could approximate the location of specific repeating segments in the instrumental track [14]. Adaptive REPET almost follows the above procedure with the additional requirement of specification of parameters for a window procedure. This helps to search different instrumental segments as the song progresses in the track. The specific parameters required are a window length, step length, and the order of a median filter [1]. In this implementation, we have minimized user interaction, as not everyone is well-versed with the intricacies of using MATLAB, and have hence kept most of the specification behind wraps by taking default values baked into the code.

4 Data Collection and Implementation

The song samples used for calculations and subsequent testing have been acquired from various popular music sources and are general representations of five different genres: Pop, Rock, Melody, Hip Hop, and Acapella. Think of Acapella songs used here as studio recordings with a soft tune in the background, typically a piano or a

little bit of guitar. We have chosen 20 genre-accurate samples of 20secs playback time for each of the aforementioned 5 kinds of music, giving us a total sample size of 100 (Table 2).

The tool for carrying out the separation of voice and music channels has a very simple interface. There are buttons for opening files after processing, with an option to save them as well. You can play the sample before starting the process, after which a spectrogram is drawn up for the frequency (Hz) versus time (s) graph of the same.

Now, based on the similarity matrix, the sample's patterns are recognized from non-repeating foreground (vocals) and repeating background (instrumentals). This triggers the BSS_Eval toolbox to run in the background for the obtained background and foreground sounds and stored within the main directory. These values of SDR, SNR, and SIR are then aggregated to find a clear "winner," as in the highest positive value for each specifier. These values are used for further ANOVA testing as showcased later in this study.

The main objective of this tool is to simplify the separation of channels while keeping user interaction to the bare minimum and obtain optimal results. The diagrammatic representation of the tool used for the entry point of sound segregation a well-equipped interface has been designed as shown below. It is a simple interface which provides all the services offered by the system (Fig. 1).

Mean and SD of Data

ANOVA Test

All calculations for ANOVA testing have been performed at 0.10 error margin. The test cases for which ANOVA is successful are highlighted in green. These F and p values indicate a successful method for a specific genre of music, giving us the validation of consistent performance across the board for a specific genre against the method used for its separation. These values are calculated taking into consideration the mean and SD from the previous tables. The findings from the conducted testing phase justifies our use of the four-algorithm suite and makes our work justified.

Box Diagram

The box plots for each model are represented, and the models are ranked according to the order of median. The SDR range is larger, and the overall trend is like the vocals. The main effects (evidence for an effect) of model are seen for the Pop (f-ratio value is 4.24291, and the p-value is 0.019153), and we find that MLRR is best suitable for the Pop Song. In the same manner, we find the Box Diagram for different algorithms

Table 2 Average tempo (BPM) of the songs

Genre	BPM
Pop	70–80
Rock	110–120
Melody	60–70
Hip hop	130–140
Acapella	40–60

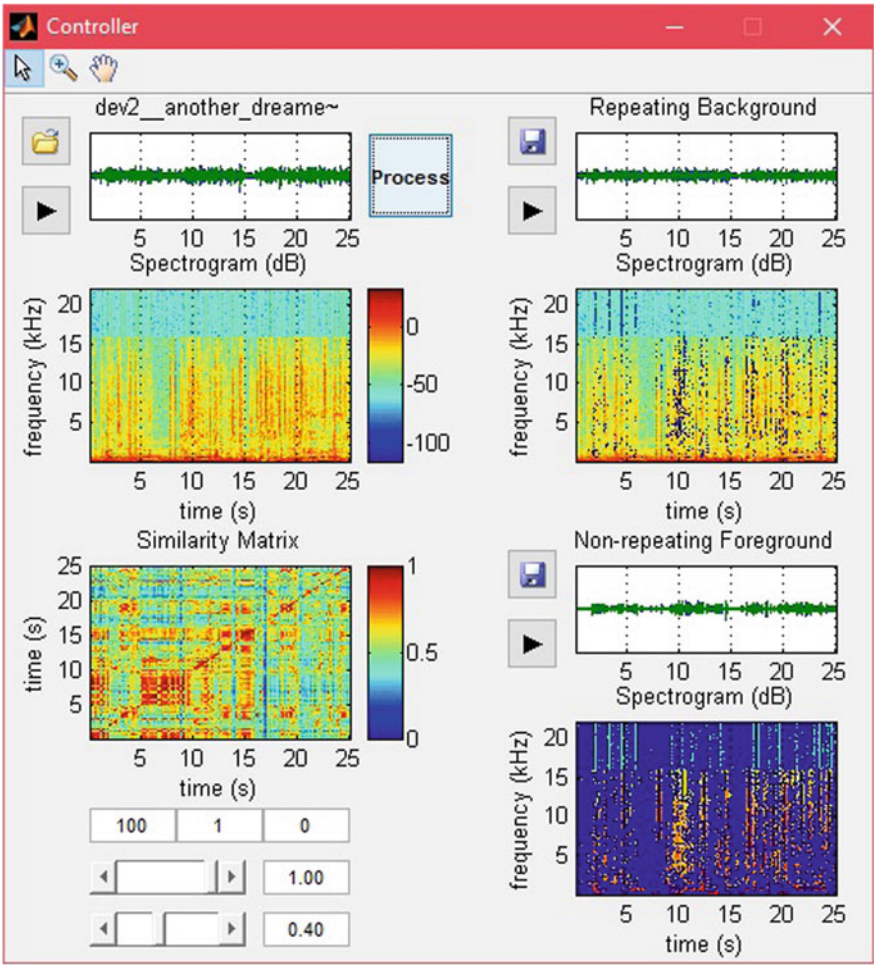


Fig. 1 Main GUI

and the result of the best-suited algorithm for different sound samples, as is listed in Table 6 (Fig. 2).

5 Conclusion

The work has been done with a comprehensive solution that will help the hearing impaired. The work emphasizes on how the segregation of music and speech for hearing-impaired children will be helpful. There are certain features like separating foreground and background sound in any kind of music sample which will help in

Table 6 ANOVA analysis of different sound samples

Genre	REPET		Adaptive REPET		SVS-RPCA		MLRR	
Value	F	P	F	P	F	p	F	p
Pop	0.17237	0.842104	0.78938	0.459025	1.82062	0.171216	4.24291	0.019153
Rock	1.95768	0.150564	1.33627	0.270931	3.30571	0.043821	1.16558	0.319061
Melody	0.94997	0.392897	0.08734	0.916486	4.13059	0.021124	1.77221	0.179192
Hip hop	1.46106	0.240546	8.75837	0.000482	0.23554	0.790911	0.30515	0.738211
Acapella	3.3784	0.041061	1.89621	0.159487	0.97821	0.382204	0.4758	0.623835

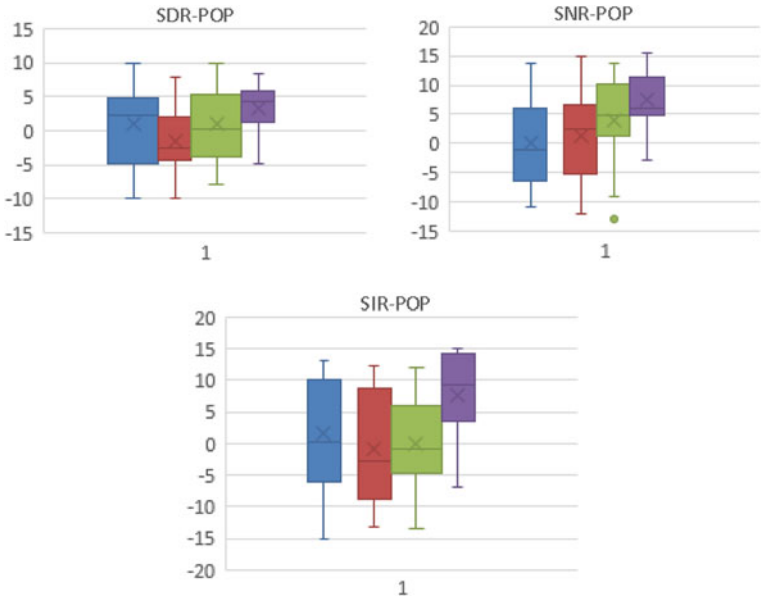


Fig. 2 Box diagram for different samples

providing rich music listening experience to the hearing-impaired listeners with lot more details and clarity at the same time. Another of its prime features includes noise and fluctuations removal technology that has been instilled in the software with the prime goal of providing the listeners with clean and ultimate sound quality without any kind of unwanted noise that tends to degrade the sound quality. The unique feature of separating the sounds of different musical instruments in any given track at any given point without losing any of its details. It can help the listeners with more detailed music listening experience.

References

1. Smith Z, Rafii, Pardo B (2013) Repeating pattern extraction technique (REPET): a simple method for music/voice separation. IEEE Trans Audio Speech Lang Process 21(1):73–84
2. Huang PT, Chen SD, Smaragdis P, Hasegawa-Johnson M (2012) Singing-voice separation from monaural recordings using robust principal component analysis. Proc ICASSP 57–60
3. Yang Y (2013) Low-rank representation of both singing voice and music accompaniment via learned dictionaries. In: International society for music ınformation retrieval conference (ISMIR)
4. Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? J ACM 58:11:111:37
5. Salamon J, Gómez E, Ellis DPW, Richard G (2014) Melody extraction from polyphonic music signals: approaches. Appl Challenges IEEE Signal Process Mag 31(2):118–134
6. Zhang Y-G, Zhang C-S (2005) Separation of voice and music by harmonic structure stability analysis. In: Proceedings IEEE ınternational conference multimedia expo, Amsterdam, The Netherlands, pp 562–565
7. Roman N, Wang D, Brown GJ (2003) Speech segregation based on sound localization. J Acoust. Soc Am 114:2236–2252
8. Kocinski J (2005) Blind source separation (BSS) of sound sources: forum acusticum, Budapest
9. Lee DD, Sebastian Seung H (2000) Algorithms for nonnegative matrix factorization. In: NIPS, pp 556–562
10. Loacher JPA, Burger J (2014) The intelligibilty of speech under reverberant conditions. Acustica 7:195–200
11. Magre SB, Janse PV, Deshmukh RR (2014) A review on feature extraction and noise reduction technique. Int J Adv Res Comput Sci Softw Eng 4(2):352–356. ISSN: 2277 128X
12. Ackerson GA, Fu KS (2014) On state estimation in switching environments. IEEE Trans Autom Control 15:10–17
13. Awning Y Research on image filtering method to combine mathematics morphology with adaptive median filter, Hefei University of Technology, Anhui
14. Sasou A, Tanaka K (2002) A waveform generation model based approach for segregation of monaural mixture sound, IEEE Explore, 2015

Designing of Low-Noise Amplifier and Voltage-Controlled Oscillator for Satellite Receiver in Ku Band

Vishnu Anugrahith Sateesh, Sanjay Kumar Surshetty, Vidushi Goel, Deepak Prasad, Vijay Nath, and Srikanta Pal

Abstract In this project, design of low-power Ku band CMOS voltage-controlled oscillator and low-noise amplifier for satellite communication is proposed. The proposed LNA provides low input impedance, low noise figure, and high gain which makes it suitable for use in satellite receivers to amplify low-power signals. This NMOS LC VCO has main advantage in low phase noise and low power. This VCO is designed on concept of negative resistance generated by MOSFETs impedance which reduce the phase noise. This circuit is designed using UMC 90 nm CADENCE ADE. For this circuit, we use 1.2 V power supply. Its average phase noise is −109.18 dBc/Hz at 1 MHz offset. The oscillator requires approx. 0.001127 mm^2 of chip area. It is highly useful for satellite communication.

Keywords Low-noise amplifier · Noise figure · NMOS LC VCO · MOSFETS · Phase noise · Power · Offset

1 Introduction

Satellites suffer from weight and size constraints, and therefore, large antennas cannot be employed. As the signal frequency increases, the size on the antenna is reduced. Therefore, satellite communication generally use the Ku band (12–18 GHz). The signal received by the receiver antenna is fed to the low-noise amplifier as shown in Fig. 1. The LNA amplifies the signal without degrading the signal-to-noise ratio, by introducing very less noise of its own to the amplified signal. Its high gain also reduces

V. A. Sateesh · S. K. Surshetty (✉) · V. Goel · D. Prasad · V. Nath · S. Pal
Department of ECE, Birla Institute of Technology, Mesra, Jharkhand, India
e-mail: sanjaysurishetty@gmail.com

V. A. Sateesh
e-mail: vishnuanugrahith@gmail.com

V. Goel
e-mail: vidushigoggle@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_58

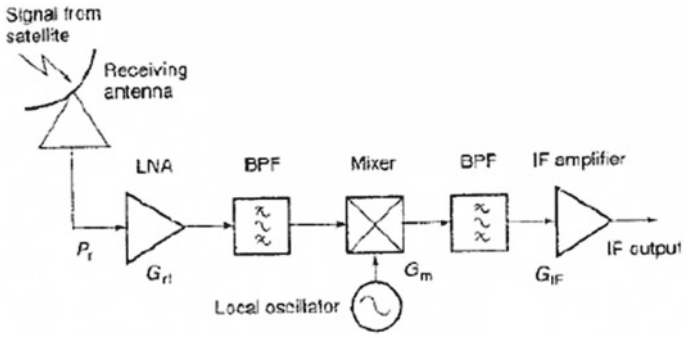


Fig. 1 A simplified block diagram of a satellite receiver

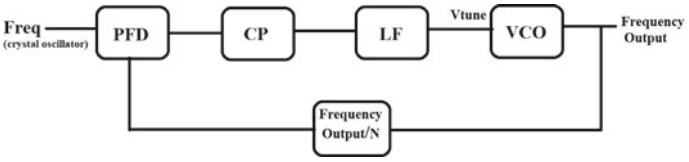


Fig. 2 Voltage-controlled oscillator

the noise contribution by the following stages of receiver. A voltage-controlled oscillator (VCO) is an electronic oscillator whose oscillation frequency is controlled by a voltage input. The applied input voltage determines the instantaneous oscillation frequency. A VCO is also an integral part of a phase-locked loop.

PFD Phase frequency detector.
CP Charge pump.
LC Loop filter.

From fixed crystal oscillator frequency as shown in Fig. 2, Vtune is generated which can be used to control output frequency of VCO. Different topologies have different phase noise performance and power dissipation. It can be designed using ring oscillator, cross-coupled LC CMOS VCO, NMOS VCO, etc. First, ring oscillator is designed using CMOS inverters. Since it has high phase noise, we designed LC VCO. The remainder of the research article is organized as follows. Section 2 discusses low-noise amplifier. Section 3 portrays the design of ring oscillator, proposed LC VCO. Section 4 is about phase noise performance of oscillators. Further VCO tuning range is discussed in Sect. 5, and simulation results are in Sect. 4 (Fig. 3).

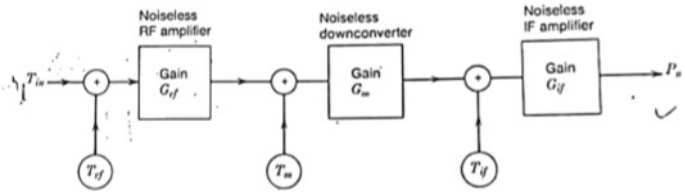


Fig. 3 Equivalent noise generators

2 Low-Noise Amplifier

A cascode common gate topology [1] with inductive biasing has been used. This method lowers the input impedance of the LNA. An unbiased cascode CG stage is shown in Fig. 4 [2]. The high gain of the LNA also minimizes the noise contribution from the mixer and intermediate amplifier [3] of the receiver.

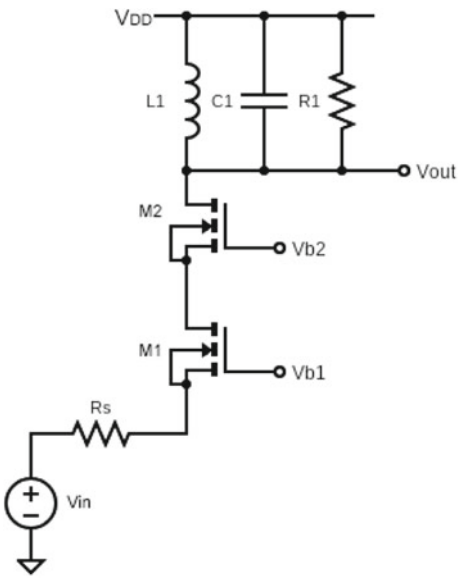
From Fig. 3, the noise power and the noise temperature are given by,

$$P_n = G_{IF} G_m G_{LNA} kB \left[T_{LNA} + T_{in} + \frac{T_m}{G_{LNA}} + \frac{T_{IF}}{G_m G_{LNA}} \right]$$

$$T_s = \left[T_{LNA} + T_{in} + \frac{T_m}{G_{LNA}} + \frac{T_{IF}}{G_m G_{LNA}} \right]$$

From the equations, it is observed that the noise temperatures of mixer and IF amplifier are divided by the LNA gain.

Fig. 4 Cascode CG stage



The input impedance is given by

$$R_{in} \approx \frac{1}{g_{m1}} + \frac{R_1}{g_{m1} r_{O1} g_{m2} r_{O2}} + \frac{1}{g_{m1} r_{O1} g_{m2}}$$

Since the product of the intrinsic gains g_{m1} and g_{m2} is large, the expression for input impedance reduces to

$$R_{in} \approx \frac{1}{g_{m1}}$$

This topology offers low input impedance, low noise figure [4, 5], low-power dissipation, and high gain in the frequency band of interest. The circuit has been designed for an input impedance of 50 Ω [6]. The addition of cascode device causes voltage headroom limitation due to stacking of two transistors, and also increases noise figure of the circuit due to the noise generated by M_2.

In Fig. 5, $C_X = C_{DB1} + C_{GD1} + C_{SB2}$. The above graph plots the ratio of noise voltage of M_2 to the output voltage, against frequency of operation. It can be seen from Fig. 6 that the noise contribution of M_2 is negligible below the zero frequency [7], given by

$$f_Z = \frac{1}{2 r_{O1} C_X}$$

When the zero frequency is lower than the transit frequency f_T of the MOSFET, the noise contribution of M_2 becomes significant. For the circuit to be properly biased,

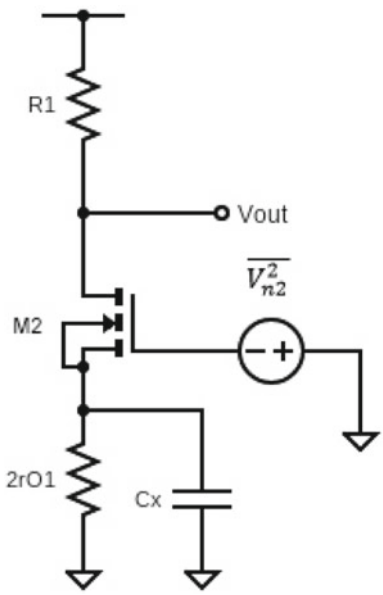


Fig. 5 Cascode transistor noise

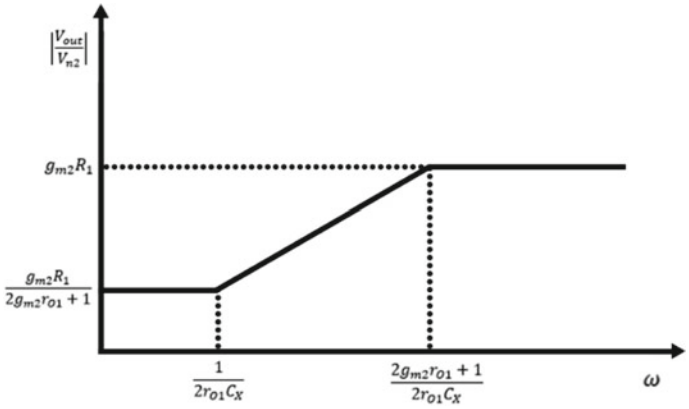


Fig. 6 Output noise contribution as a function of frequency

$V_{DD} > V_{GS2} + (V_{GS1} - V_{TH1})$, i.e., the cascode topology takes up an additional voltage headroom of $V_{GS1} - V_{TH1}$ [8]. This reduces the voltage available with the biasing resistor thereby increasing its noise contribution. In order to overcome this noise-headroom trade-off, we used an inductor for biasing in the designed circuit, thereby minimizing noise due to bias element, as shown in Fig. 7. This also cancels the input capacitance of the circuit and hence improves input matching.

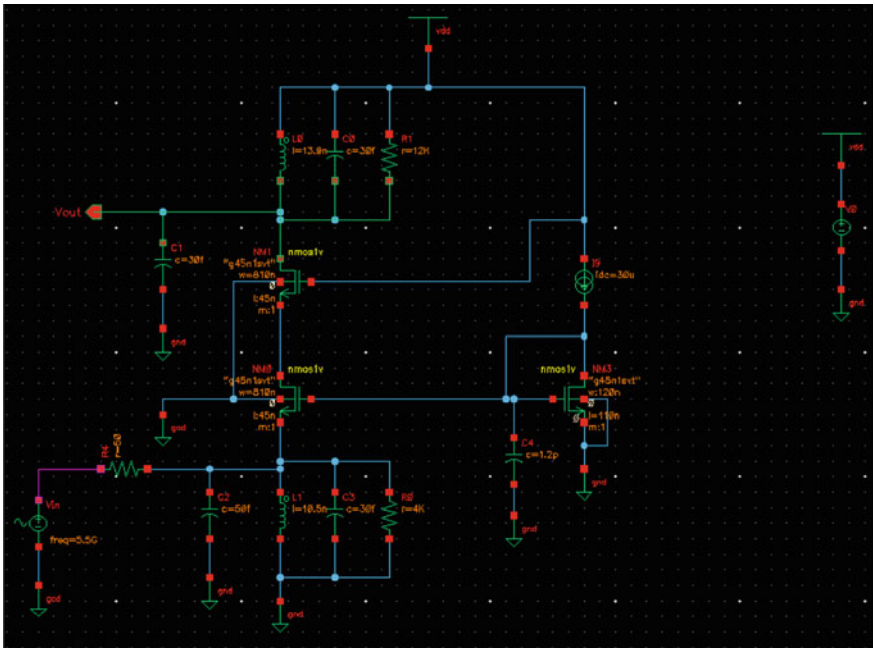


Fig. 7 Inductively biased cascode CG stage

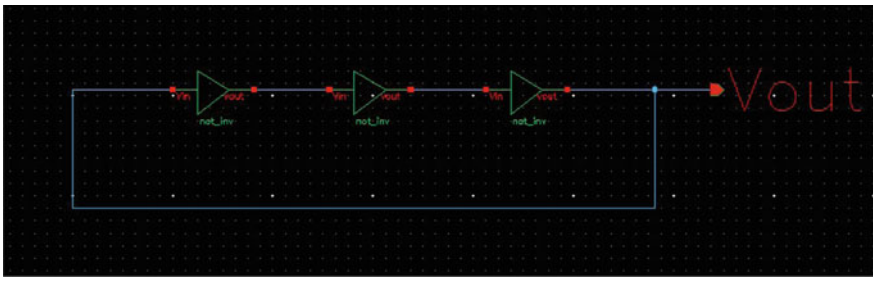


Fig. 8 Ring oscillator

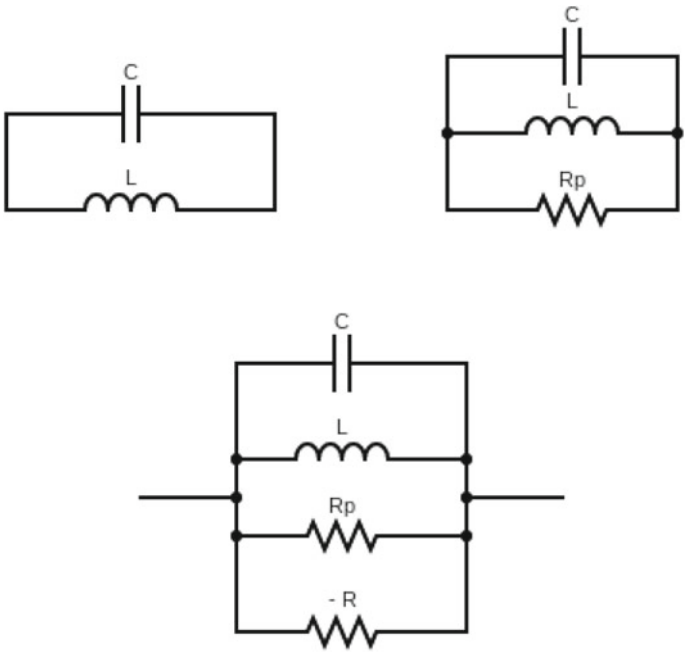


Fig. 9 LC tank circuit

The noise figure is given by

$$NF = 1 + \gamma + 4\frac{R_S}{R_1}$$

The noise figure is nearly 3db. It can also be observed that noise figure and input resistance decrease with increasing g_m. Therefore, a lower noise figure can be achieved if a small amount of impedance mismatch is allowed at the input.

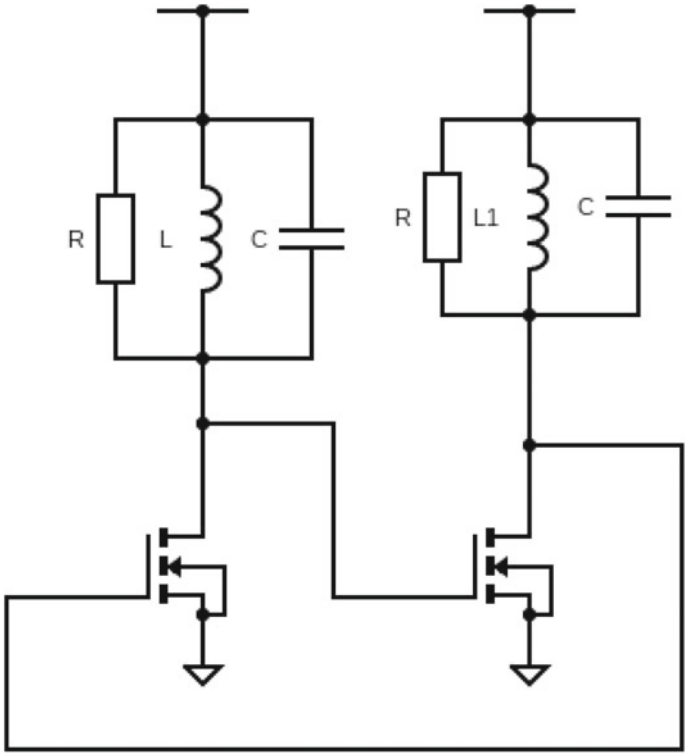


Fig. 10 LC voltage-controlled oscillator

3 Oscillator

A. *Ring oscillator*

A real ring oscillator only requires power to operate. Above a certain threshold voltage, oscillations begin spontaneously. To increase the frequency of oscillation, two methods are commonly used. Firstly, making the ring oscillator from a smaller number of inverters as shown in Fig. 8 results in a higher frequency of oscillation, with about the same power consumption. Secondly, the applied voltage may be increased. In circuits where this method can be applied, it reduces the propagation delay through the chain of stages, increasing both the frequency of the oscillation and the current consumed. The maximum permissible voltage applied to the circuits limits the speed of a given oscillator. Changing the supply voltage changes the delay through each inverter, with higher voltages typically decreasing the delay and increasing the oscillator frequency.

If 't' represents the time-delay for a single inverter and 'n' represents the number of inverters in the inverter chain, then the frequency of oscillation (f) is given by

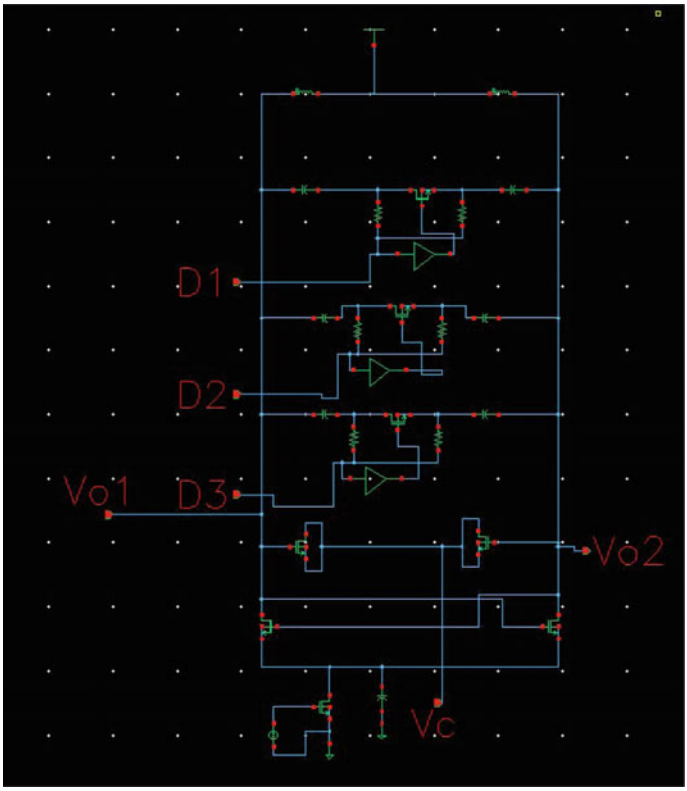


Fig. 11 Simulated LC voltage-controlled oscillator

Fig. 12 Ideal spectrum w_0

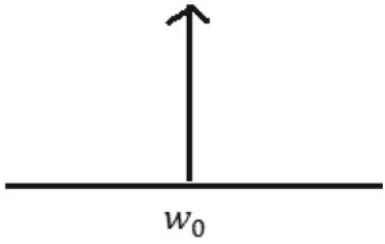
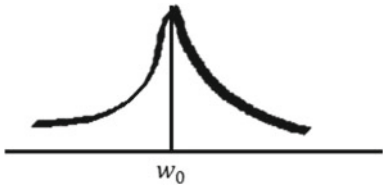


Fig. 13 Actual spectrum due to phase noise



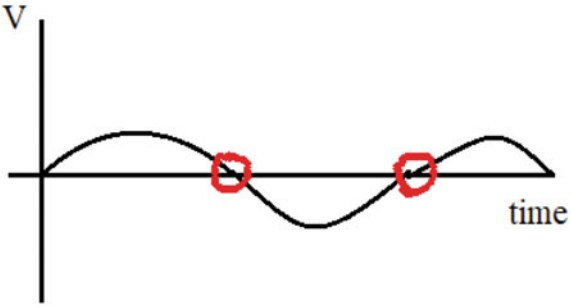


Fig. 14 Jitter

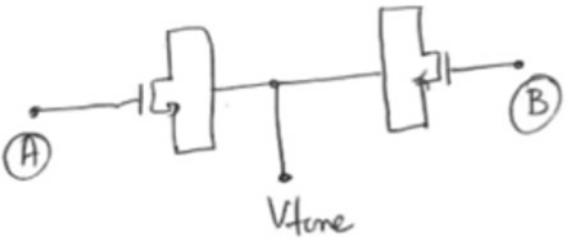


Fig. 15 NMOS varactor (continuous tuning)

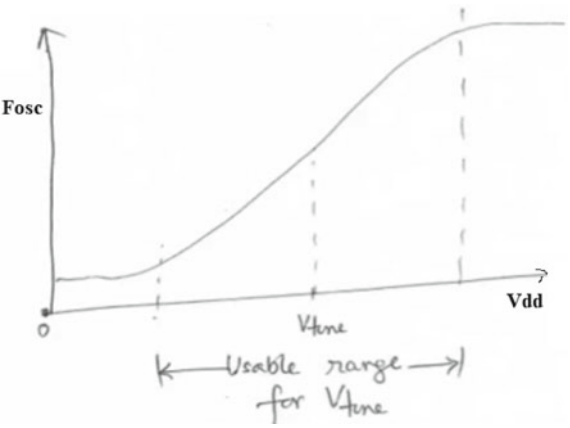


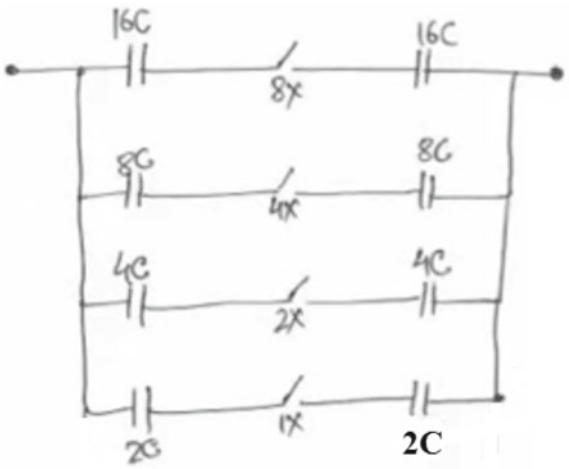
Fig. 16 Frequency variation of varactor

$$f = \frac{1}{2tn}$$

B. *LC voltage-controlled oscillator*

As we know that for oscillations, we require amplitude of system should be greater than 1 and phase of feedback system should be 180°.

We can see from Fig. 9, that in general for inductance there is resistance [9] in parallel and given as R_p. Now, this has to be nullified by negative resistance for



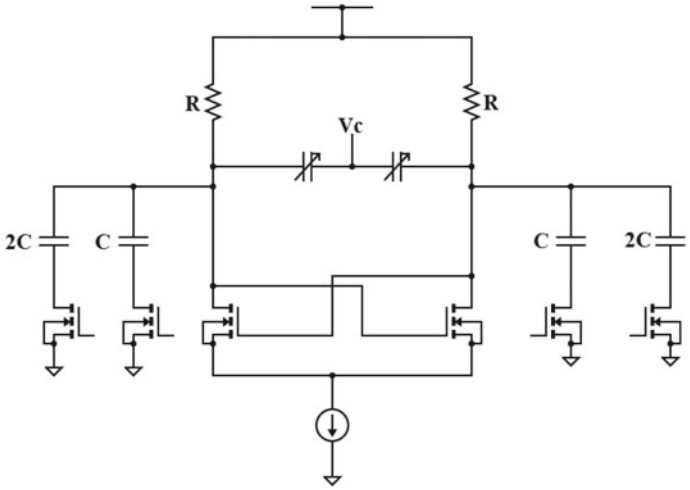


Fig. 18 LC NMOS VCO with discrete and varactor banks

getting sustained oscillations, and we use a topology which gives negative impedance $-R_p$. This can be achieved by using LC NMOS VCO as shown in Fig. 10. LC VCO [10, 11] have better phase noise performance compared to ring oscillators.

With single inverter topology using MOSFETS with RLC oscillators, we need two CS stages with R, L, and C where each stage produces 180° phase shift at resonance (Figs. 11, 12 and 13).

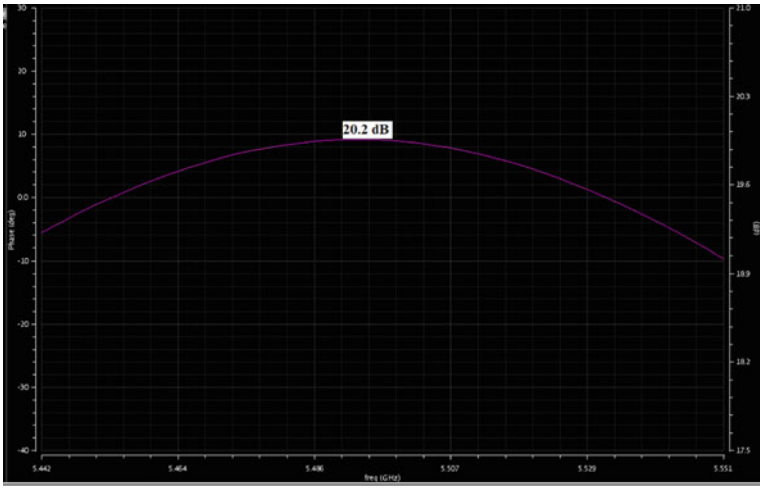


Fig. 19 Plot of gain versus frequency

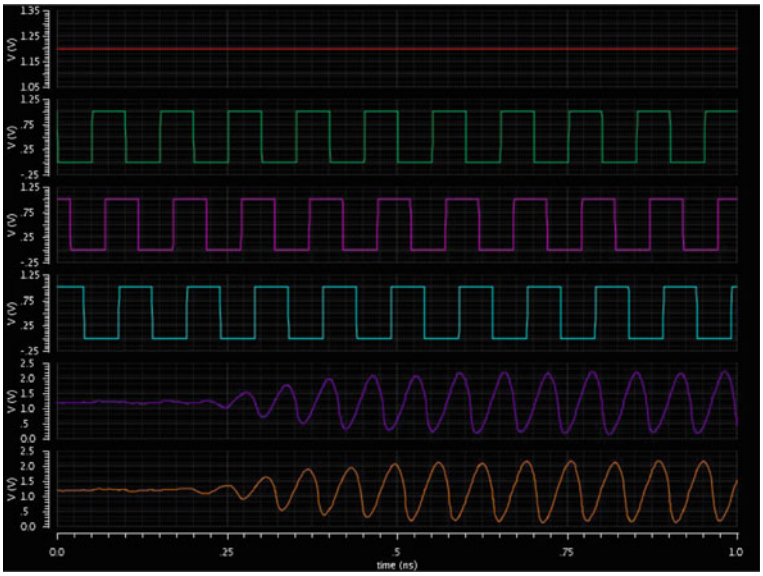


Fig. 20 Transient analysis of LC VCO

4 Phase Noise

Phase noise is represented in the frequency domain and consists of rapid, short-term, random fluctuations in the phase (frequency) of a waveform, caused by time domain instabilities (jitter as shown in Fig. 14). Jitter is a method of describing the stability

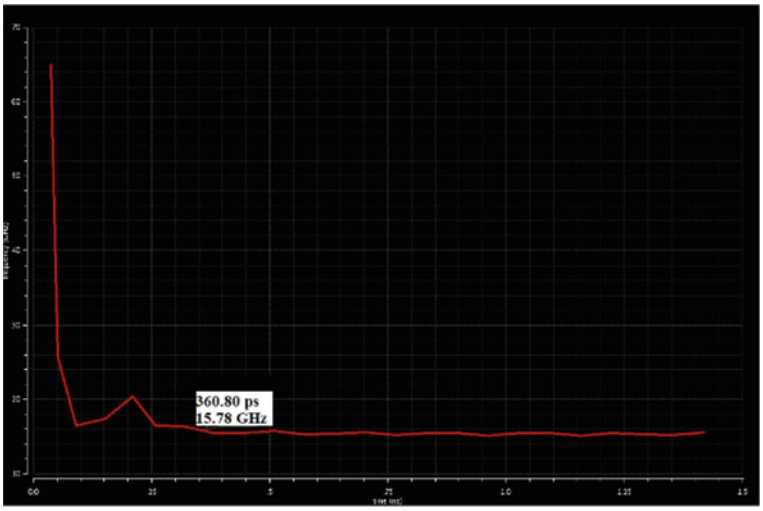


Fig. 21 Frequency of 15 GHz for LC VCO

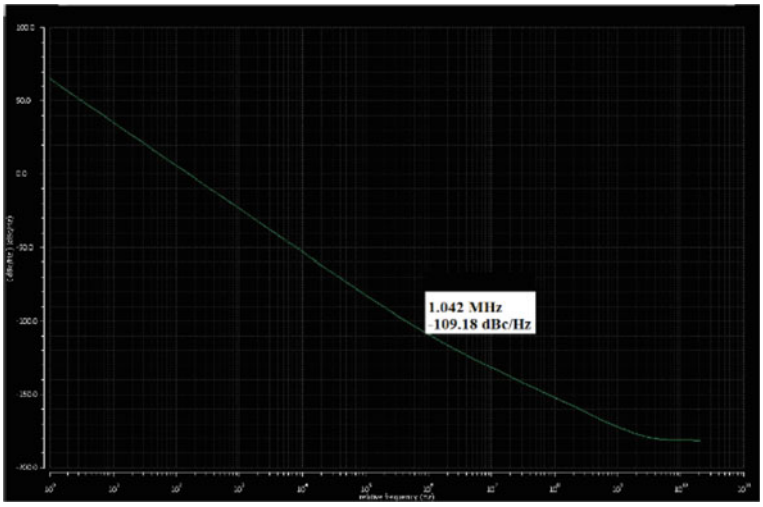


Fig. 22 Phase noise of LC VCO at 1 MHZ offset

of an oscillator in the time domain. It combines all the noise sources together and shows their effect with respect to time. Phase noise is caused from thermal noise and flicker noise. Change in voltage supply, noise from tail current sources, and noise from nmos causes phase noise. For supply voltage fluctuations, we use voltage regulator externally.

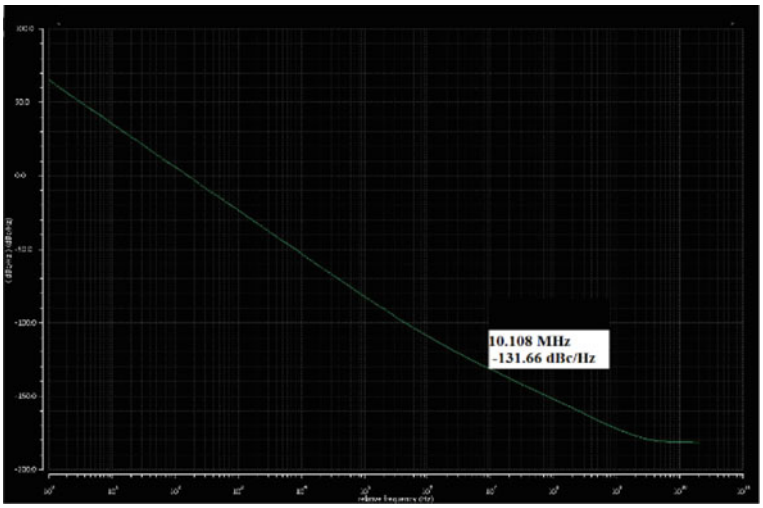


Fig. 23 Phase noise of LC VCO at 10 MHZ offset

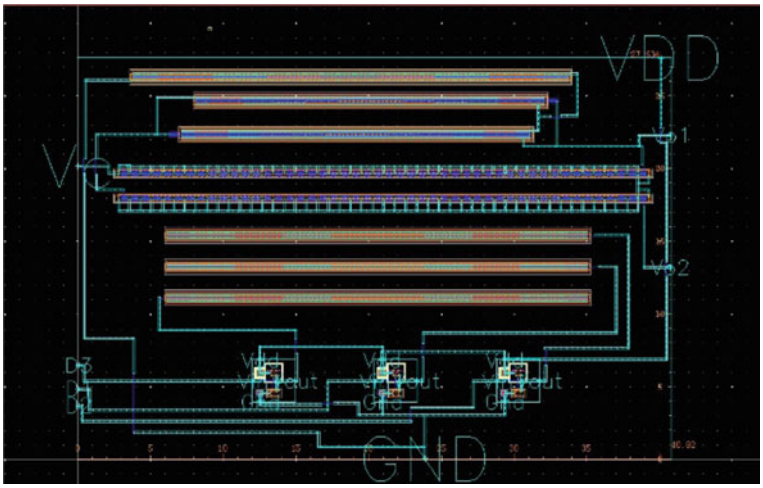


Fig. 24 Layout of LC VCO

We can measure phase noise by taking some offset. For example as shown in figure, noise power in 1 Hz bandwidth is at an offset delta f which is measured in dBc/Hz.

At low frequencies, the noise is dominated by flicker noise caused by current sources and up converted to $1/f^2$. After some offset as shown in figure, thermal noise is up converted to $1/f^3$.And next is $1/f$ noise from buffers. Phase noise is

dBc/Hz where dBc means phase relative to carrier. VCO noise at low frequencies is suppressed by loop filter, charge pump, and phase frequency detector at low frequencies used in PLL (Figs. 15 and 16).

5 Tuning Range

For tuning [12], we require discrete banks [13] of capacitors and varactor as shown in Fig. 17. Since the output voltage changes with PVT variation, we require analog control voltage. For reaching particular frequency, we use discrete tuning.

$$C_{\text{total}} = C_{\text{fixed}} + C_{\text{var}}$$

$C_{\max} = C$ (when switch is on in discrete bank).
$C_{\min} = Cgd + Cdb$ (when switch is off).

Phase noise [14] depends on Q, and tuning range [15] depends on $C_{\max}/C_{\min}$. As we can see $C_{\min}$ is directly proportional to widths and also Q is directly proportional to widths of switches. So, proper choice of widths of switches also should be taken care off.

Varactor is nothing but which has voltage variable capacitor. It has capacitance varying with voltage across drain and gate. NMOS varactor is shown in Fig. 15. LC NMOS VCO with discrete varactor banks and varactor is shown in Fig. 18.

6 Simulation and Results

The LNA has been designed for operation centered at a frequency of 5.5 GHz, bandwidth of 100 MHz, input impedance of 50 Ω. The plot of the amplifier gain vs frequency for the frequency band of interest is shown below. From the plot, we can observe that gain is highest (20.2 dB) at a frequency of 5.49 GHz. Also the gain remains within 1 dB of the maximum gain, in the frequency band of interest (Figs. 19, 20, 21, 22, 23, and 24).

7 Conclusion

From simulation results, we observe that low input impedance, low noise figure, and high gain (20.2 dB) provided by the proposed LNA. Therefore, it can be used in satellite receivers to amplify the low-power signal picked up by the receiver antenna without severely degrading its SNR. We have also seen methods of increasing tuning range of VCO and decreasing phase noise. Table 1 shows comparison of ring oscillator and LC VCO in terms of phase noise. Tables 2 and 3 compare our work with

Table 1 Comparison between three types of encoders

	3-inv ring oscillator	5-inv ring oscillator	LC NMOS VCO
Phase noise dBc/Hz (at 1 MHz)	−58.66	−62.15	−109.18
Phase noise dBc/Hz (at 10 MHz)	−87.31	−92.00	−131.66

Table 2 Comparison table for LNA

Characteristic	This work	[5]	[8]	[4]	[3]	[2]
Technology	45 nm	90 nm	180 nm	90 nm	–	0.13 μm
Supply voltage used	1.8 V	–	–	0.7 V	1.3 V	1.1 V
Central frequency of operation	5.5 GHz	5.5 GHz	2.65 GHz	1.5 GHz	1.5 GHz	1 GHz (max)
Gain reported	20.2 db	13.3 db	20.09	19.8 db	13.2 db	14.7 db

Table 3 Comparison table for VCO

Characteristic	This work	[16]	[17]	[18]
Technology	90 nm	180 nm	130 nm	130 nm
Supply voltage	1.2 V	1.8 V	1.2 V	0.4
Frequency (GHz)	10.5–16.6	59.1–65.8	5.24	14.1
Phase noise (dBc/Hz at 1 MHz offset)	−109.18	−82	−109.5	−100.6
Phase noise (dBc/Hz at 10 MHz offset)	−131.66	−113	−118.8	–

other papers. The tuning range is in the range 10.5–16.6 GHz. Phase noise is found to be −109.18 dBc/Hz at 1 MHz offset and −131.66 dBc/Hz at 10 MHz. This LC VCO utilizes active area of 0.001127 mm^2.

Acknowledgements The authors are heartily thankful to respond, Indian Space Research Organization (ISRO) Bangalore, India, for funding this research work. The sanction number is ISRO/RES/3/679/16-17. A piece of thanks also goes to Council of Scientific and Industrial Research (CSIR), New Delhi, India, for providing the senior research fellowship to the research scholar. The sanction number is 09/554(0042)/2018-EMR-I, Ack. No. 143183/2k17. The authors also want to thanks Vice-Chancellor, Dr. M. K. Mishra and Head of Department, Dr. S. Pal for their constant support and encouragement.

References

1. Azizan A, Murad SAZ, Ismail RC, Yasin MNM A review of LNA topologies for wireless applications. In: 2014 2nd international conference on electronic design
2. Idris MI, Yusop N, Chachuli SAM, Ismail MM, Arith F, Shafie N (2014) Design and analysis of low noise amplifier using cadence. J Theoret Appl Inf Technol 69(1)

3. Pattewar PC, Khanchandani KB Performance comparison of various low noise-high speed amplifier topologies for GPS applications. Int J Current Eng Technol E-ISSN 2277–4106, P-ISSN 2347–5161
4. Baig N, Noorbasha F (2019) CMOS low voltage LNA with improved noise figure. Int J Innov Technol Expl Eng (IJITEE) 8(6). ISSN: 2278-3075
5. Al-Sagar ZS, Salih MS, Jasim AK, Sameen AZ (2017) A comparison of low noise amplifiers design and techniques. J Eng Appl Sci 12(Special Issue 7):8077–8081
6. Das T (2013) Practical considerations for low noise amplifier design. RFLNAWP Rev. 0
7. Yadav N, Pandey A, Nath V (2016) Design of CMOS low noise amplifier for 1.57 GHz. In: 2016 international conference on microelectronics, computing and communications (MicroCom). https://doi.org/10.1109/microcom.2016.7522438
8. Lamba RK, Vithalani CH (2014) RF CMOS low noise amplifier design—a case study. I.J. Wireless and Microwave Technologies, vol 5, pp 14–24
9. Zou Q, Ma K, Yeo KS Design of a Ku-band low-phase-noise VCO using the dual LC tanks. IEEE Trans Circuits Syst II Express Briefs 59(5):262–266. https://doi.org/10.1109/tcsii.2012.2190855
10. Wang PK, Chou M-C, Chen Y-T, Chang Y-T, Chang D-C, Hsu SH A Ku-band low-phase-noise transformer coupled VCO for satellite communications. In: 2016 IEEE international symposium on radio-frequency integration technology (RFIT). https://doi.org/10.1109/rfit.2016.757814
11. Kim H, Ryu S, Chung Y, Choi J, Kim B (2006) A low phasenoise CMOS VCO with harmonic tuned LC tank. IEEE Trans Microw Theory Tech 54(7):2917–2923
12. Elbadry M, Kalia S, Harjani R (2014) A 52% tuning range QVCO with a reduced noise coupling scheme and a minimum FOMT of 196dBc/Hz. In: IEEE CICC, pp 1–4
13. Mirajkar P, Chand J, Aniruddhan S, Theertham SLow phase noise Ku-band VCO with optimal switched-capacitor bank design. IEEE Trans Very Large Scale Integr (VLSI) Syst 26(3):589–593. https://doi.org/10.1109/tvlsi.2017.2769709
14. Torres JA, Freire JC (2008) Ku band voltage controlled oscillator on SiGe. In: APCCAS 2008—2008 IEEE Asia Pacific conference on circuits and systems. https://doi.org/10.1109/apccas.2008.4746286
15. Zhang J, Zhang Y, Lü H, Zhang Y, Liu B, Zhang L, Xiang F (2015) A Ku-band wide-tuning-range high-output-power VCO in InGaP/GaAs HBT technology. J Semicond 36(6):065010. https://doi.org/10.1088/1674-4926/36/6/065010
16. Chang Y-H, Chiang Y-C (2014) A V-band push-push VCO with wide tuning range in 0.18 μm CMOS process. IEEE Microw Wirel Comp Lett 25(2):115–117 2015.
17. Yu X, El-Gouhary A, Neihart NM A transformer-based dual-coupled triple-mode CMOS LC-VCO. IEEE Trans Microw Theory Tech 62(9):2059–2070. https://doi.org/10.1109/tmtt.2014.2332498
18. Nguyen T, Lee J-W (2012) Ultralow-power ku-band dual-feedback armstrong VCO with a wide tuning range. IEEE Trans Circuits Syst II Exp Briefs 59(7):394–398

Language Diversity and Their Preservation in Jharkhand

Shubham Shrotriya and Saumendra Pattnaik

Abstract This chapter emphasizes the language diversity of Jharkhand and its importance for digital preservation. The chapter highlights the multilingual characteristics of the region enlisting several languages of Jharkhand spoken across the states among tribal and non-tribal groups. It highlights the importance of language preservation, its historical and cultural importance and reports the current status of undergoing projects for awareness and preservation of the endangered languages.

Keywords Language diversity · Digital preservation · Endangered languages · Jharkhand

1 Introduction

Jharkhand has rich diversity in language, flora and fauna, region, landscape, race, caste and so on. Apart from the huge racial diversity present in the state, the state has large linguistic diversity too. This multilingual nature of Jharkhand is a valuable resource which needs care and appreciation to survive and grow helping prosper the state's cultural and historical importance. This paper presents the bouquet of languages spoken in the state along with the technical advancement for its preservation.

Jharkhand was carved out of Bihar on November 15, 2000, because of poor development in the region in spite of having 40% of the India's total mineral deposit. Only 24% of the overall population lives in cities, and hence, it is primarily a rural state [1, 2]. The state is known to house some of the tribal kings of past known as Mundas Rajas. As per 1991 census, around 28% of the people are tribal, while 12% belong

S. Shrotriya (✉) · S. Pattnaik
Department of CSE, Siksha 'O' Anusandhan University, Odisha, India
e-mail: shubhamshrotriya16427@gmail.com

S. Pattnaik
e-mail: saumendrapattnaik@soa.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_59

Table 1 Multilinguism of Jharkhand

Indo-Aryan Languages	Bhojpuri, Hindi, Urdu and Bengali
Dravidian Languages	Korwa, Oraon and Sauria Paharia
Munda Languages	Kurukh, Santhali, Mundari, Bhumij, Paharia and Ho

Source Land and People of Jharkhand

to SC/ST categories. As per 2001 census, most of the people speak Hindi, followed by Santhali, Khorta, Bengali, Urdu and other tribal languages [3].

As per UNESCO's report on 'Atlas of the World's Languages in Danger (2010)', India has the largest number of endangered languages in the world [2]. Out of the total endangered languages across the planet, 197 languages belong to India. Out of these, majority of the languages are spread across North-east, Jharkhand, Orissa, Karnataka and the Himalayan belt [3]. According to UNICEF study on Language diversity in Jharkhand, 96% of the population communicates in tribal languages at their home [3]. There are close to 19 major mother tongues used by people of Jharkhand. The tribal language used as mother tongue by largest number of people is Santhali (33%), followed by Kurukh (9.5%), Mundari (7.6%), Sadri (6.7%), Ho (5.6%), Oraon (1.1%). Khadia, Pahadiya, Birhori and Pachhiyari are spoken by around 2% of the population [4]. As per UNESCO report, some of the tribal languages like Asur, Birhor, Malto, Kurux, Khariya, Mundari, Ho, Angika and Korwa are at a threat of being endangered [4].

2 Multilinguism of Jharkhand

Jharkhand is a land of multiculture and languages. It is home to many tribal and non-tribal languages. Table 1 shows the major languages of the region [5–7]. Details of many Jharkhand Languages are shown in [8, 9].

3 Mother Tongues in Jharkhand

Table 2 describes the mother tongue along with percentage of population speaking it as per 2011 census [8]. Santhali is the most spoken language of the place. Apart from these, 35% of the rural population speaks other tribal or regional languages like Khadiya, Ho, Mundari, Pachpargania and Nagpuri [10–12].

Table 2 People speaking different mother tongues

S. No	Language	Percentage of population
1	Santhali	33.1
2	Khortha	17.5
3	Kurukh	9.5
4	Other tribal languages	5.6
5	Regional languages	30.6
6	Hindi	3.7

4 Languages in Day-to-Day Life

This section briefs the use of languages used by people of Jharkhand in their day-to-day lifestyle [9]. Analysis shows that usually mother tongue is spoken just within the family while interaction with children, spouse or parents. Along with the mother tongue, some section of population also uses at least one other language at home. In such scenario too, the use of Hindi is limited [11, 12]. The second or third language used is a tribal or regional language. Many languages are having interrelation among them in terms of Phonemes.

4.1 Haat Bazaar

Most of the people while interacting with friends in the market used primarily their mother tongue. Many residents also use more than one language for communication. Hindi was mostly neglected among villages with very few people it for transaction.

4.2 Recreational Activity

Most of the people prefer regional and tribal languages for creating songs, music and other creative skills. People prefer using tribal and regional languages for singing, storytelling, story listening and listening to songs. Use of Hindi was more predominant while writing as compared to other tasks. Apart from Hindi, people preferred Santhali, Sadri, Kurukhand Ho languages for writing.

4.3 Interaction with Hindi Speakers

Most of the people find it easy to communicate in Hindi with Hindi speakers visiting to their villages as government officials, NGO members or businessmen. Although

people are not fluent in Hindi and make many grammatical mistakes while speaking, they have functional knowledge for Hindi.

4.4 Language Preference for Education

Around 70% of the people interviewed wanted mother tongue to be the medium of education for their kids, while rest 30% favored Hindi. This is due to fact that most of the people speak in their local language since birth.

4.5 Language Preference for Journals and Books

Most of the villagers prefer articles and journals on health and agricultural issues in their regional and tribal languages over Hindi language so that they can easily understand the concept of journals.

5 Impact of Language Diversity on Children

One of the studies suggested that most of the children faced learning challenges if mother tongue is different than that of language of instruction [3, 13]. Some of the common problems were reading, writing, comprehension and slow pace of learning.

Most of the students interact with friends in schools in their local language only. Teachers also felt that learning could be immensely improved if Hindi is not being imposed on them and they were taught in the language of their mother tongue. It is also reported that most of the students preferred their regional languages for communication and around 41% of the people used their mother tongue only to ask query to their teacher.

6 Languages in Communities

It is the common belief that Hindi is the mostly used language in Jharkhand has been strongly challenged with only 4% of the people using it as their first languages. Study shows that 96% of the people has mother tongue other than Hindi [3, 5]. Regional dominance of a language districts and talukas are also observed. For people speaking multi-languages, people preferred regional and tribal languages over Hindi for second and third languages [13–18]. People were found to have poor reading and writing skills in their tribal and regional languages due to the scarcity of enough reading

and writing material in that language. While interacting with people in marketplace, people like the dominance of local language in each part of India [19].

7 Linguistic Map of Jharkhand

This section describes the distribution of different languages in Jharkhand [3]. The entire region is distributed into three main zones:

- Santhal Pargana Division
- North Chotanagpur Division
- South Chotanagpur Division.

For each of the divisions, dominant language along with second most dominant languages is discussed.

7.1 Santhal Pargana Division

It consists of Pakur, Godda, Jamtara, Sahebganj and Dumka. Santhali emerged as the most spoken language of the region followed by Khortha and Bangla [20] especially around Jamtara district because of its closeness with West Bengal. Panchparganiya and Hindi are also used as second languages in some areas of Sahibganj and Dumka.

7.2 North Chotanagpur Division

It consists of Ramgarh, Bokaro, Dhanbad, Koderma, Giridih, Hazaribagh and Chatra districts. In this zone, Khortha emerged as the most dominant language along with Kurukh and Santhali. The second most dominant language of the region is Hindi. Areas like Ramgarh, Bokaro and Dhanbad have a heavy influx of people from other states leading to more use of Hindi. Khariya, Mundari, Birhori and Kurukh are other second dominant languages in Bagodar block of Giridih. Other major languages spoken in the area are Magahi, Khortha, Santhali and Hindi. Kurukh is the most dominant language in Palamu. Hindi is most widely spoken in blocks of Garhwa, while Sadri and Mundari are also spoken by many people.

7.3 South Chotanagpur Division

Mundari is the most spoken language in Khunti and some areas of Simdega. Sadri, Nagpuri and Kurukh are also major spoken languages in the region. Sadri and Nagpuri

are among most spoken languages around the capital. Panchparganiya language is preferred in Kudu block of Lohardaga. Other major languages of the region are Kurukh, Sadri and Khariya, Nagpuri and Hindi. Around the borders of Odisha and West Bengal, Ho is the most dominant language followed by Santhali, Gond and Odiya. Some people in the Kolhan division also prefer Gond, Odiya and Santhali.

8 Research Works

Languages of Jharkhand have been explored very less despite the huge diversity. The literature review shows very few works in this domain.

In one of the works, automata and digital corpus for Santhali language has been proposed by Akhtar et al. [22]. This chapter discusses Santali nouns in terms of gender, number and case. This chapter also shows parts of speech tags and lemmatized corpus in terms of noun, numeral, preposition, pronoun, verb, adjective, adverb, conjunction. A total of 590,314 tokens, 425,238 words and 63,199 sentences have been added in Santali corpus. A survey was conducted by M-TALL akhra of the Jharkhand Tribal Welfare Research Institute (JTWRI) with support of UNICEF titled "Language diversity of Jharkhand" [3]. The report highlighted the rich language resource of the state along with its impact on children education. In one of the works, "Ho Language Digit Recognition using MFCC and its Performance Evaluation for Different Classifier" by Prakash et al., an ASR for digit recognition was developed in the native Ho language, and its performance for online and offline recording modes was compared [21]. In one of another work, "Phonemes of Asuri" by Ganesh Baskaran, an attempt is made to explore the existing phonemes in Asuri language [22]. Asuri language is spoken around Chota Nagpur area which falls under the Munda family. Asuri language is mostly spoken by North Munda group. Asuri is at the threat of extinction with very few speakers. Most of the people speaking Asuri are bilingual and speak other major languages like Hindi, Mundari, Santali and Sadri which is probably the reason, the language is slowing getting endangered. In one of the works, "Khortha or Khotta: An Endangered Languages of India and the Urgency to Retain its Pure Variety" by Priya et.al., a study was made to analyze the dangers and challenges the languages of India are facing and the need to restore the language in its purest form [23]. They have found that because of mixing of locals with people speaking different languages like Hindi, Maithili and Bengali, there has been huge transformation in the native language. Different aspects of change of language like gender, age, education and area [24, 25] of living have been thoroughly investigated, and important observations were made. Female was found to use more pure form of the words than male as they are less in touch with the outside world. Similarly, old people spoke the language in much purest form as the young generation studies different languages and is in constant touch with smart phones and gadgets. Similar analogy was found with educated people and people living in urban areas. The study showed that the purest form of the language was found in semi-educated villagers. John Peterson et al. in their work "Language contact between Indo-Aryan and Munda

in eastern-central and South-Asia" (2010) present a linguistic convergence between Indo-Aryan languages of Nepal and Eastern India and Munda [12]. Jharkhand has been considered as the centroid of language convergence with Bangladesh to its east, Nepal and Bihar to its north and Orissa toward south. The study shows the large-scale lexicon burrowing from the Indo-Aryan languages into Munda.

9 Status of Tribal Language

This section discusses the status of tribal languages in the state and the measures taken by government for their preservation and development.

A department of tribal and regional languages is set in 1981 for research and post-graduate courses. It started initially with seven major tribal languages, i.e., Mundari, Santhali, Kurukh, Ho, Kharia, Kurmuli and Nagpuri, and later, two other language, Khortha and Pach-Pargana were added [3]. Government of Jharkhand has taken significant steps in protecting some of these tribal languages. Bi-lingualism exists in Jharkhand where people use more than one language in home and market. Threats of loss of vocabulary, language attrition and sharing of Indo-Aryan lexicons are commonly observed. The state has a weak language policy, and there is constant rift among people claiming their language as more superior than other and demanding more facilities in lieu of it. Regional languages of Jharkhand have to face stiff challenges from major languages and had to struggle hard to survive. Linguistic survey of India has enlisted 18 endangered languages, and there are initiatives to document them to preserve their vocabularies.

10 Conclusion

It is quite evident from the studies that Jharkhand contains bouquet of languages with as many as 32 mother tongues spoken across the state, but many are at the threat of endangerment due to many social and economic issues. Language endangerment influences both people speaking them and the language itself. Every language has cultural and historical importance associated with it, and a loss of language results in loss of cultural tradition tied with it. Traditional songs, poems and cultural values go away with the language. This also results in loss of social bonding among the people as their social values, traditions and culture are replaced by new ones. Losing a language can also result in loss of knowledge transcribed in that language in different fields like botany, medicine, arts, science, philosophy, etc. Jharkhand is seeing rapid transformation over past few years with setting up of many new industries bringing people from outside to their region. They bring with them their language and culture which results in mode mixing. Many young people from villages also leave their native home in search of jobs and education and move to new places. These factors also contribute to loss of language. Although it is inspiring to know that recently

these issues have been highlighted by many linguist and scientist of both India and abroad. Many NGOs along with world and national organizations like UNESCO and tribal and welfare department are putting their effort and money in preserving the wide culture and language diversity of the place. Government of India too has given national recognition to some of the languages of the region to promote the use of the tribal languages. It should be clearly understood that multilingualism is not a threat to the integrity of the nation but a prize to be proud of. Hence, the trend of diversity in languages must be preserved for recognition of different cultures of languages.

References

1. Traveljharkhand.com. (2018) History of Jharkhand, Jharkhand History. Available at https://www.traveljharkhand.com/jharkhandtourism/jharkhand-overview/jharkhand-history.html
2. UNESCO Atlas of the World's Languages in Danger (2018)
3. Hussain D (2013) A study on socio-linguistic pattern and its impact on children's learning in Jharkhand
4. Bhatt SC ed (2006) Land and people of Indian States and Union Territories: In: 36 Volumes. Rajasthan, vol 23. Gyan Publishing House
5. Bhatt SLC, Bhargava GK (2006) Language and literature; land and people of Indian states and Union Territories, vol 12. Jharkhand
6. Bell RT (1976) Sociolinguistics-goals approaches and problems. B.T. Batstord Ltd., London
7. Mallikarjun B (2002) Mother tongues of India according to the 1961 census,languages in India, vol 2.5
8. Das PK (2013) Ergativity in Khortha. https://www.academia.edu/23020698/Ergativity_in_Khortha_an_enigmatic_phenomenon
9. Minz D, Hansda DM (2010) Encyclopedia of schedule tribes of Jharkhand. Kalpaz Publications, Delhi
10. Singh KS (1994) The scheduled tribes. Oxford University Press, Delhi
11. Vidyarthi LP (1976) Tribal culture of India. Concept Publishing Company, New Delhi
12. Peterson J (2010) Language contact in Jharkhand. Linguistic convergence between Munda and Indo-Aryan in eastern central India. Himalayan Linguist 9.2 3(3):56–86, Dec 2016
13. Gnanasundram V (1982) Conditions for the code mixing behaviour. In: Third international conference on South Asian languages and linguistics
14. Holmes J (1992) An introduction to sociolinguistics. Longman, London
15. Hudson RA (1980) Sociolinguistics. Cambridge University Press, London
16. David CS (2008) Understanding mixed code and classroom, code-switching: myths and realities. New Horizon J 56:3
17. Kumar N (2015) Urban continuum in Jharkhand: beyond the state formation. IJRANSS 3(1)
18. Lal M (1992) Encyclopedia of Indian literature: Sasay to Zorgot, vol V. Sahiyta Akademi, New Delhi
19. Gajrani S (2004) History, religion and culture of Eastern India, vol 4. Gyan Publishing House, Delhi
20. Akhtar AK, Sahoo G, Kumar M (2017) Digital corpus of Santali language. In: 2017 international conference on advances in computing, communications and informatics (ICACCI). IEEE, pp 934–938
21. Jay Prakash RK, Lal AK, Chandra M Ho language digits recognition using MFCC and artificial neural network. In: National conference on recent trends in electronics and electrical engineering (NCRTEEE), Inderprastha Engineering College, Ghaziabad, India, February 16–17 2017

22. Baskaran SG (2015) Phonemes of Asuri. Language in India, p 59
23. Priya S, Singh R (2016) Khortha or Khotta: an endangered language of India and the urgency to retain its pure variety. Int J Human Cultural Stud (IJHCS) 3(3):190–201. ISSN 2356-5926
24. Palo HK, Chandra M, Mohanty MN (2017) Emotion recognition using MLP and GMM for Oriya language. Int J Comput Vis Robot 7(4):426–442. Scopus, ISSN: 1752-9131
25. Kuddus K, Chatterjee B (2014) Second language acquisition through technology: a need for underdeveloped regions like Jharkhand. Int Refer e-J Literacy Explor 2(2):52–259

Design of Energy Harvestor Using Piezoelectric Material

Atulya Arya, Shradha Shekhar, Avinash Priyam, and Vijay Nath

Abstract Piezoelectric materials can be used to convert oscillatory mechanical energy into electrical energy. This technology, together with innovative mechanical coupling designs, can form the basis for harvesting energy from mechanical motion. Piezoelectric energy can be harvested to convert walking motion from the human body into electrical power. Recently, four proof-of-concept Heel Strike Units were developed where each unit is essentially a small electric generator that utilizes piezoelectric elements to convert mechanical motion into electrical power in the form factor of the heel of a boot. The results of the testing and evaluation and the performance of this small electric generator are presented. The generator's conversion of mechanical motion into electrical power and the processes it goes through to produce useable power and commercial application of the Heel Strike electric generator are discussed.

1 Introduction

The piezoelectric effect converts mechanical strain into electrical voltage. The strain can come from a lot of different sources. Human motions, low-frequency seismic vibration and acoustic noise are few examples. The piezoelectric effect can be implemented for harvesting mechanical energy from walking. This energy can be converted to useful electrical energy that can be used to power various wearable electronic

A. Arya · S. Shekhar · A. Priyam (✉)
Department of EEE, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: priyam.avinash@gmail.com

A. Arya
e-mail: atulya.arya06@gmail.com

V. Nath
Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: vijaynath@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_60

devices such as sensors and Global Positioning System (GPS) receiver. Piezoelectric energy harvesting can also be used to power some consumer electronic device directly such as cellular phone, two-way communicator and pager.

Much of this electricity is still generated with dangerous and outdated technology centered on the burning of fossil fuel. The continued burning of fossil fuel will have catastrophic impacts on humanity. This danger can be mitigated through clean and innovative technique of energy harvesting.

Recently, four proof-of-concept Heel Strike Generators (see Fig. 1) were developed for converting the mechanical energy of walking into electrical energy. Each Heel Strike Generator utilizes four piezoelectric elements (each one being a lead zirconate titanate (PZT-5A) bimorph crystal stack) to convert mechanical motion into electrical power in the form factor of the heel of a boot, while the user walks, electrical power is generated. The goal of this research effort was to generate a 0.5 W of power at 1 Hz step rate since many electronic devices such as GPS receivers and communicator require power within this range to operate. The Heel Strike Generator relies on the piezoelectric effect to generate electric power. Piezoelectricity is the ability of some crystal to generate electric potential in response to an applied mechanical stress. When a crystal is under mechanical stress (e.g., by compression

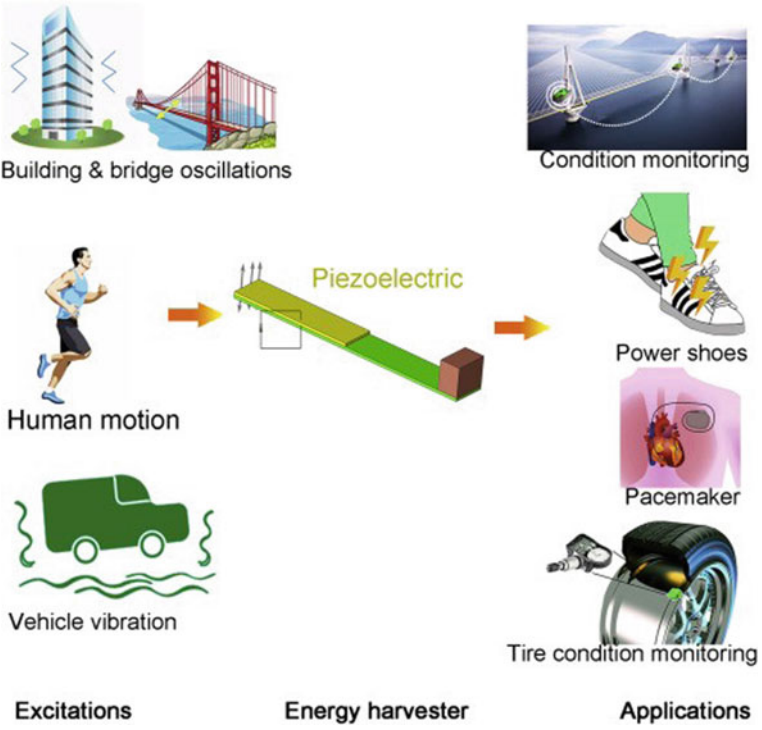


Fig. 1 Energy Harvester

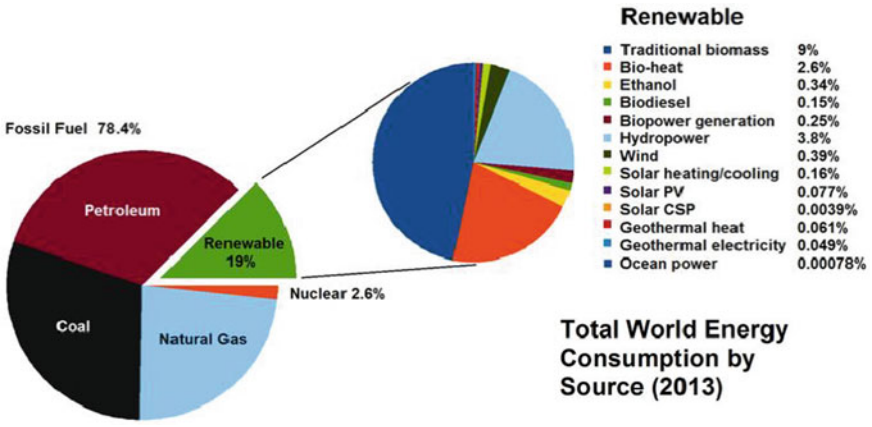


Fig. 2 Total world energy consumption (2017)

or expansion), the electrical charge of a dipoles becomes aligned, leading to the net electric polarization. This is responsible for electric potential across the crystal and provides convenient transducer effect between electrical and mechanical oscillations. If mechanical vibrations are applied to such crystals, they will respond with an electrical oscillation output which acts as a source of power. This effect can be exploited for harvesting energy from disturbance sources. For the electrical power to be useful, an additional electronic circuit would be required to rectify and regulate the power output in the most efficient way possible (Fig. 2).

2 Background

Energy production around the world has been steadily increasing over the last thirty years. In 1985, the world was using just under 10,000 TWh. In the year 2014, the world produced over 23,000 TWh of electricity. The graph shows the trends for electricity consumption around the world. The interesting thing about the graph is the relatively constant production of electricity in North America and Europe. However, the large increase in electricity production in the Asia Pacific is indicative of the developing juggernaut of China [1].

The method by which electricity is generated depends on the fuel being used. For most fossil fuel electricity generation including those with natural gas, coal and oil, the method is through burning the material and turning a steam turbine generator. But other forms of electricity generation also exist that do not involve burning fossil fuels. Some of these alternatives include solar, wind, hydroelectric, nuclear and geothermal/biofuel. Solar electricity generation utilizes the sun's light waves to generate an electric current. Wind and hydroelectric both use turbines to generate electricity through the wind or flowing water, respectively. Nuclear power

is similar to how fossil fuels generate electricity. Water is turned into steam to power a turbine, but the heating process for nuclear is free from CO_2 release. The same can be said for geothermal and biofuel generation technique that also utilize the steam turbine for electricity generation [1–3] (Figs. 3 and 4).

Smaller-scale electricity generation has also become a large industry. One of these methods is using a crankshaft to turn a magnetic through a coil of wire to induce a voltage. This same technique is used in many applications including a shake flashlight that utilizes a magnetic and a coil to generate a voltage. As the magnetic moves back and forth through the coil, an alternating current can be generated. Small-scale solar has also been increasingly used on the roofs of houses and business, but also in the realm of electricity generation for personal devices [3, 4].

Cal Poly has a history of energy harvesting projects. The first project is the Energy Harvesting from Exercise Machines which seeks to harness the energy generated by people using exercise machines and delivers that energy back to the electric grid [5, 6].

The block diagram shown in Fig. 5 describes the input and output of the proposed Faraday rechargeable battery system. The input to the kinetic energy harvesting

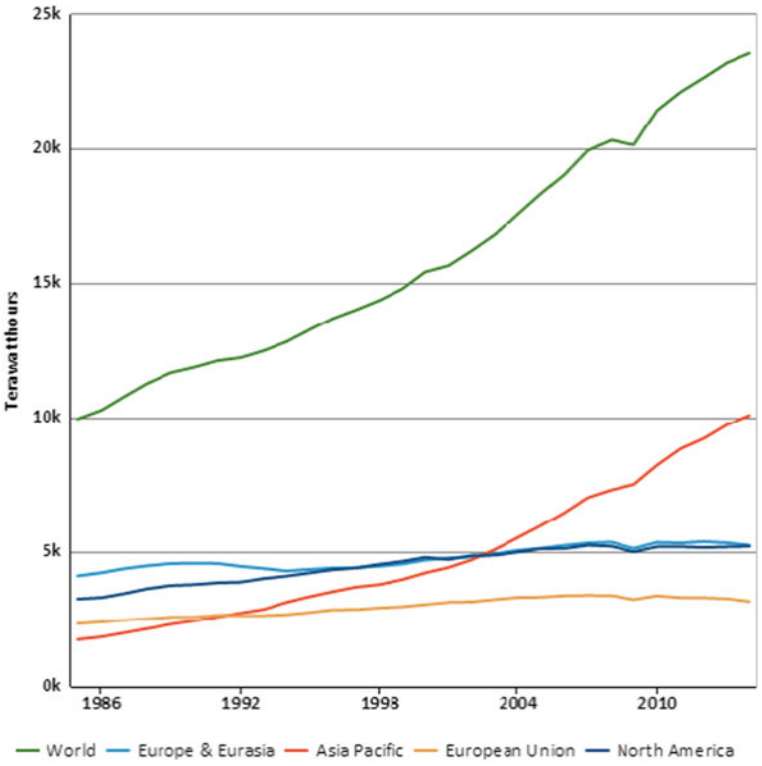


Fig. 3 Electricity consumption over the last 30 years

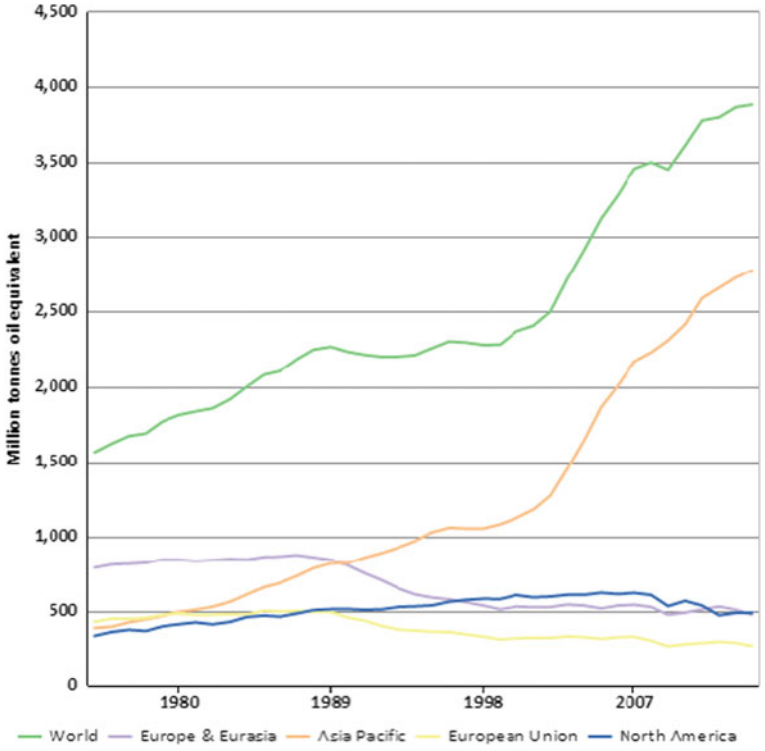


Fig. 4 World consumption of coal in Mtoe

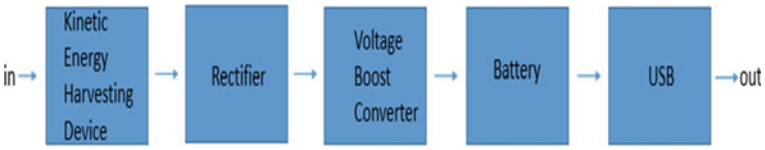


Fig. 5 High-level system block diagram

device will be created by the bike user physically pedaling. This will result in a fluctuating sinusoidal wave. We then need to pass that wave through a rectifier to make the signal a constant DC voltage. A constant DC voltage is needed to charge our battery and output from the USB. Since we want our output to work with a USB connection, we need the output voltage to be 5 V. This will require us to boost the output of our rectifier which can be achieved using a boost converter. That boosted voltage will then be stored in a battery for future use via the USB output connection [7].

The completed system will have two available inputs. The main input is the alternating current signal produced from the kinetic energy harvesting block. The kinetic

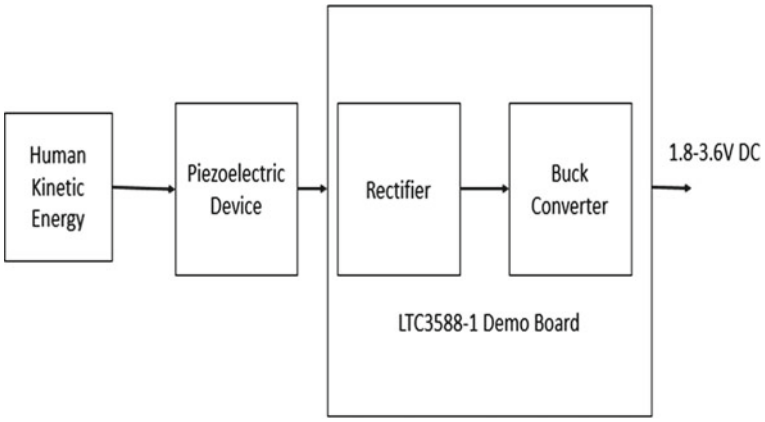


Fig. 6 Complete system diagram

energy will be generated by the user through pedaling a bicycle or accelerating in a car. This will provide the power from the rest of the circuit to operate. A secondary input can be found on the removable battery. This battery can be recharged through a micro-USB port when the battery is not present in the bicycle system. There are three outputs of the system. The first output is from the battery to the USB port where electronic devices can be charged. Another output powers the on-board lighting system.

Since the lights will primarily be used at night time, there will be a switch controlling whether the lights are on or not. This gives the user flexibility when using the system. Once the energy harvesting circuitry was designed and simulated, the next step was to design the piezoelectric device that will convert mechanical energy into electrical energy. The preliminary design used Faraday's law of induction. This law states that if a time-varying magnetic field is applied to a coil of wire, an electromotive force (EMF) will be produced across the coil. The faster the magnetic field changes through the coil, the higher the EMF that will produced. The EMF can also be increased by increasing the number of coils that the magnet passes through. This limited the design variables to two [8, 9] (Figs. 6, 7 and 8).

3 Description

The performance of the proof-of-concept Heel Strike Generator was tested and evaluated to determine its suitability for military and commercial applications. In the test and evaluation phase of development, various resistive loads were applied to the Heel Strike Generator and the power output was measured. Recommendations for further enhancement of the Heel Strike Generator are noted based on the electrical test and evaluation results.

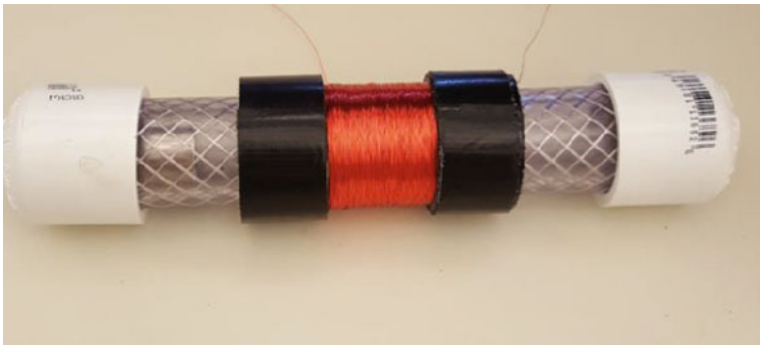


Fig. 7 Piezoelectric device

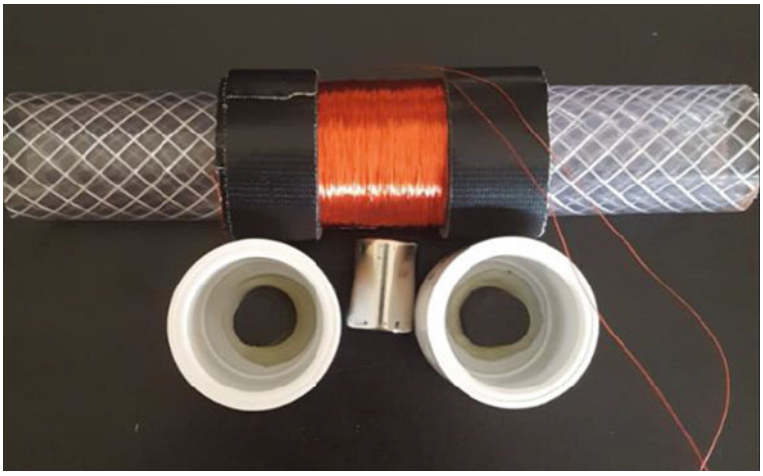


Fig. 8 Top view of Piezoelectric element

The Heel Strike System consists of two major pieces: the Heel Strike Generator and the power electronics circuit. The Heel Strike Generator is the device where it has a mass of 0.455 kg and has approximate dimensions of 8.89 cm (L) by 7.94 cm (W) by 4.29 cm (H). The principle components of the Heel Strike Generator are four PZT-5A dimorph crystal stacks, lead screw, bearing and rotary cam, whose operation are described elsewhere in Sect. 3. The power electronics circuit is 5.2 cm^2 with a height of 1.7 cm and has a mass of 10 g. Its purpose is to convert unusable power from the Heel Strike Generator to useable power. The power electronics circuit is connected to the Heel Strike Generator to form the Heel Strike System [10–12] (Figs. 9 and 10).

This circuit will store very small energy pulses over a relatively long time, in a low leakage storage capacitor and then periodically discharge that capacitor into a load. The current pulses from the piezoelectric phases are in AC waveform. A set

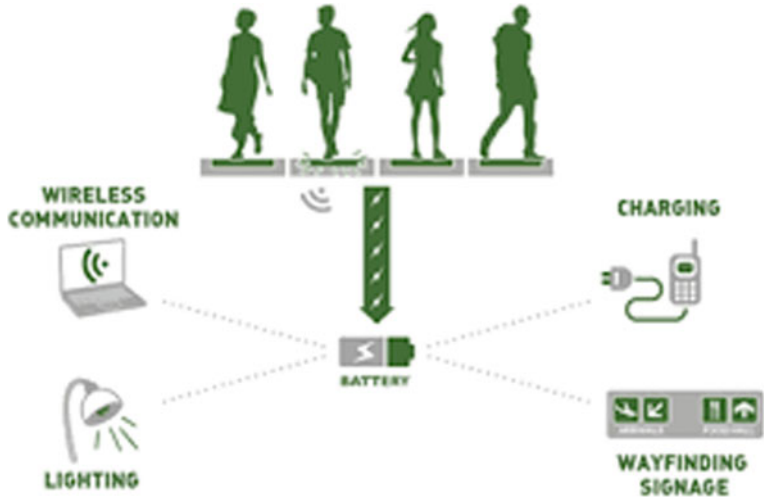


Fig. 9 General working

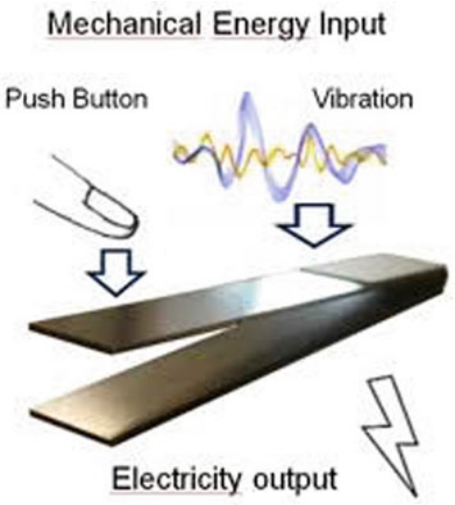


Fig. 10 Conversion of mechanical energy to electrical energy

of zener diodes in the circuit will rectify the AC waveform from each of the four phases and produce DC pulses of current that charge the storage capacitor in steps. Power is provided from the storage capacitor to an integrated DC–DC converter IC which starts operation. This switch is part of a standard forward, or buck converter that converts the energy stored in the capacitor to a regulated 12.

VDC output. Figure 3 shows the typical waveforms for the circuit (Fig. 11).

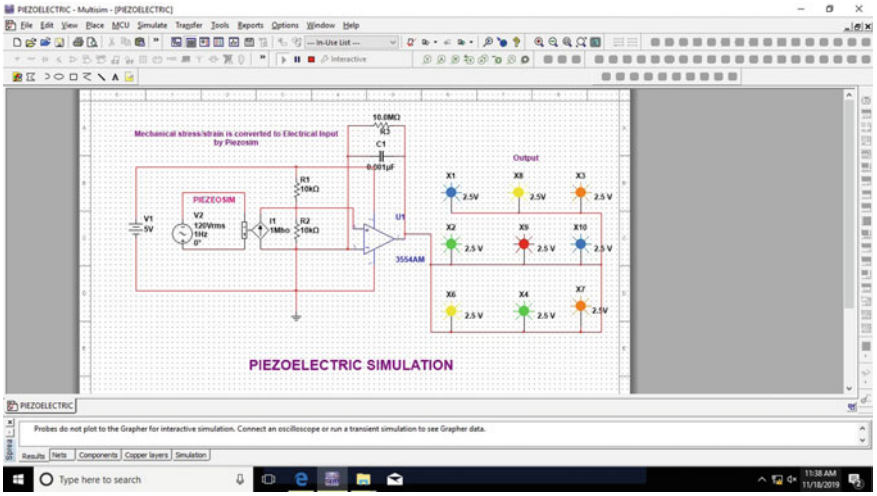


Fig. 11 Simulation

4 Simulation

The simulation for the proposed idea was concluded in Multisim 14.1, and results were satisfactory and as expected. The piezoelectric effect converts mechanical strain into electrical voltage. This strain can come from many different sources. Human motion, low-frequency seismic vibrations and acoustic noise are a few examples. The piezoelectric effect can be implemented to harvest mechanical energy from walking [2, 7, 9, 13].

5 Conclusion

The conclusion of the above is that the piezoelectric element was successful in conversion of mechanical stress/strain into electrical energy. The mechanical vibrations were recorded and converted into electrical signal say voltage through the piezoelectric element which led to the glowing of the various led probes connected together which in turn represents the various load that can be connected to the circuit for use implying the energy harvesting was successfully done. The entire simulation was carried in NI multisim which verified our assumption of the working of piezoelectric element and concluding the research. The evidence of the successful conclusion of the experiment is attached below as image file.

Acknowledgements We would like to express our special thanks of gratitude to our Vice Chancellor Dr. S. Konar, Dean of Student Welfare Dr. Anand Kumar Sinha, H.O.D. EEE Dr. P. R. Thakura,

H.O.D. ECE Dr. S. Pal for their precious guidance and support and also for providing us with all the facilities till completion of the project.

References

1. Case Studies in Enterprise Systems and Complex Systems
2. Trolier-McKinstry S (2008)
3. Stable & optoelectronic dipeptide assemblies for power harvesting
4. Sinha D, Amaratunga G (2015)
5. "Sensor Senses" by Robert Reeps
6. Becker RO, Marino AA
7. Review of power harvesting by piezoelectric materials
8. Harper D Piezoelectric
9. Piezoelectric crystal classes. Newcastle University, UK
10. Repas R Sensor sense: piezoelectric force sensors
11. Teubner BG Archived from the original on 2014–04–21. Berlin
12. Piezo drum kit quickstart guide—Sparkfun Electronics
13. Fundamentals of piezoelectricity. Oxford University Press

Kisaan Seva—A Web site for Serving the Farmers

Suchismita Sarkar, Archana Kerketta, and Vijay Nath

Abstract This chapter demonstrates a Web site which is a direct interaction of government and its schemes with the farmers. Through this Web site, farmers will be kept updated with the new schemes, market prices of the vegetables, etc. as proposed by the Government of India. In this way, farmers will no longer be dependent on middlemen to know the new schemes, prices and to sell their products through them. There will be direct link between the government and the farmers and hence, guard the farmers from being exploited by the middlemen.

1 Introduction

With an annual output of 260 million tons, India is among the largest producers of fruit and vegetables in the world. Agriculture plays a vital role in the field of an economy. It acts as the backbone of the economic system of our country. It not only provides food and raw material but also opportunities for employment to a very large proportion of population. The following facts clearly showcase the importance of agriculture in this country [1]. There are four types of agricultural marketing in India like sale in village, sale in mandi, sale in market and cooperative marketing. In agricultural marketing, transportation cost, inadequate market infrastructure, lack of market information, lack of processing units, storage facility and price fluctuation are the major source of problems. Apart from middlemen, enough storage facility, freedom from moneylenders, adequate transportation facilities, availability

S. Sarkar (✉) · A. Kerketta · V. Nath
Department of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: suchisarkar1999@gmail.com

A. Kerketta
e-mail: archanakerketta1997@gmail.com

V. Nath
e-mail: vijaynath@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_61

of loan and training facilities, etc., are required for satisfactory agricultural marketing. Among all these problems, transportation charges are concerned as the major one by the maximum number of farmers [2]. Even though India is a country with abundant agricultural land, still its agricultural marketing has been defective since decades. The Indian farmers fail to get reasonable price for the products even after their hard work due to exploitation by the middlemen [3].

But over the years, the system has created several layers of intermediaries,stretching the chain of supply and increasing the opportunity for cartels to form, which in turn drive prices down for farmers and up for consumers [4]. According to Goldman Sachs, middlemen have become like monopoly buyers in agriculture marketing, allowing them to take advantage of any problem, but they will not pass on the benefits to farmers or consumers [5].

Farmers' protests have shown up on the headlines consistently in the past years, and their distress over issues such as low prices, poor permissions for transport and government-assured purchases has been clearly evident. While the 14 volumes of the Dalwai Committee Report on doubling farmers' income released last year provided a road map for transition from a mere Green Revolution to an Income Revolution for farmers, a parliamentary panel report submitted in the Lok Sabha on January 3 underlines that the country is yet to solve the 'Riddle of Agriculture Marketing' [6].

Sumbul Mashhadi discusses the restrictive nature of the Agricultural Produce Market Committee (APMC) Act which is responsible for making the existence of middlemen in agriculture inevitable, stripping off farmers their right on their produce [7]. However, the very purpose of the act has been made redundant, and arguably even counter-productive, over time. Illiterate or semi-literate farmers, who cannot understand the concept of taxes in place, are at the advantage of middlemen and agents in these markets who have a license to operate [8].

Government schemes for farmers:

National Mission for Sustainable Agriculture (NMSA) aims at promoting Sustainable Agriculture through climate change adaptation measures. The major thrust is enhancing agriculture productivity especially in rain-fed areas while focusing on integrated farming, soil health management and synergizing resource conservation.

Pradhan Mantri Krishi Sinchai Yojana (PMKSY) has the motto of 'Har Khet Ko Paani' for providing end-to-end solutions in irrigation supply chain, viz. water sources, distribution network and farm-level applications. It adopts state level planning and accurate execution that allows states to create their own based on District Irrigation Plans and State Irrigation Plans [9].

E-NAM-National Agriculture Market (eNAM) is a pan-India electronic trading portal which interconnects the existing APMC mandis to create a unified national market for agricultural commodities.

The Paramparagat Krishi Vikas Yojana (PKVY) is an initiative to promote organic farming in the country, according to which farmers will be encouraged to form clusters and incorporate organic farming methods over large areas in the country.

Pradhan Mantri Fasal Bima Yojana (PMFBY) is the government sponsored crop insurance scheme that integrates multiple stakeholders on a single platform [10].

The income earned from crops depends on the prevailing market scenario, the greed of middlemen, the process of selling the produce and other several factors. Very often, the government-administered minimum support price (MSP) may not even fulfill the cost of production. Farmers are then caught in debt traps. They require money to pay back their loans, fund the next agricultural season and support their family at the same time. Hence, they take loans again. Loan recovery processes may be initiated, and these are often unethically aggressive in nature, even bordering on the criminal [11].

CACP is a scheme for crop pricing to directly benefit farmers. The proposed scheme promises to support farmers without disturbing the market [12].

Various measures are being taken by the West Bengal government to ensure that middlemen have no role to play during its grain-buying activities. The government is making a list of farmers who are recipients of benefits under the Krishak Bandhu Scheme [13].

The current agrarian crisis in India is a result of two factors: failing to recognize when the Green Revolution started giving low returns; and the economic impact of subsidies [14]. With the goal of formulating a unified national market for agricultural commodities by integrating Agriculture Produce Marketing Committees or APMCs across states in India, the 'National Agriculture Market' or NAM scheme was launched [15]. The government tried to modernize APMCs but ignored principal economic drivers in the APMCs. Any law which forces farmers to sell in a specified area or reduces the supply of traders through licensing will lead to exploitation of farmers [16]. Even the modernized 'model APMCs' of 2003 and 2017, retain two provisions that give rise to cartelization: trader licensing and market monopoly. Therefore, economics predicts that monopolistic behavior by the market committees and cartel formation by the traders will ensue [17].

APMCs are not only in India. Variations of this law were implemented in many other countries. However, over the years, South Africa, New Zealand, Australia and others have repealed such laws. It is high time that India abandons this fundamentally flawed legal policy in all its forms [18]. Farmers' incorporation in the whole process of marketing agricultural products can eliminate the need for middlemen totally, agriculture minister Mahmoud Hojjati said [19].

Agriculture is the major sector of the Indian economy and is the principal livelihood of about 58.4% of the country's population. However, in India the agricultural system largely remains inefficient and has been manipulated by middlemen [20]. One of the most successful propaganda under agricultural marketing initiative has been the Rythu Bazars in Andhra Pradesh state of the country. Presently, there are 106 Rythu Bazars operating in the 23 districts of Andhra Pradesh [21].

While horticulture is on average more remunerative than other crops, cultivation of perishable crops comes with its own set of problems. The retailer companies like Ninjacart, Crofarm and KrishiHub persuade farmers to join them by assuring them prices that are 15% higher than what local agents would pay. Farmers are kept updated about demand for a crop, current prices offered and other details either via apps or SMS [22]. Retailers and other wholesale customers can order through an

app where they are offered better prices, on-time deliveries and a quality assurance. There is also availability of product traceability [23].

After extensive research, it has been found that the total profit margin of middlemen in the entire chain adds up to 75%. There are also high rates of spoilage, as cargoes get repeatedly loaded and unloaded [24]. Food has to be made affordable to all keeping intact the quality, and our target should be to end hunger and malnutrition in India. In an agricultural country, which proudly proclaims 'Jai Jawan, Jai Kisan,' it is important that we truly respect the hard work of our farmers and make them pay off.

2 Methodology

The frontend of the Web site is built using HTML5 (Hypertext Markup Language), CSS3 (Cascade StyleSheet), JavaScript (ES6), Bootstrap and jQuery. HTML5 is the most latest and enhanced version of HTML. HTML is a markup language. It helps to define the layout of a webpage. CSS3 enhances the look of the Web site and makes it more attractive by implementing different formatting schemes to the markup language. JavaScript is responsible for the transition and fluidity of the webpage. ES6 does this with less amount of code. For the backend of the Web site, Flask is used. Flask is a micro web framework written in Python. MySQL database is used to store, retrieve and update the user data. It is an open-source relational database management system, so we can have a record of what all the registered customers are doing on the Web site [25].

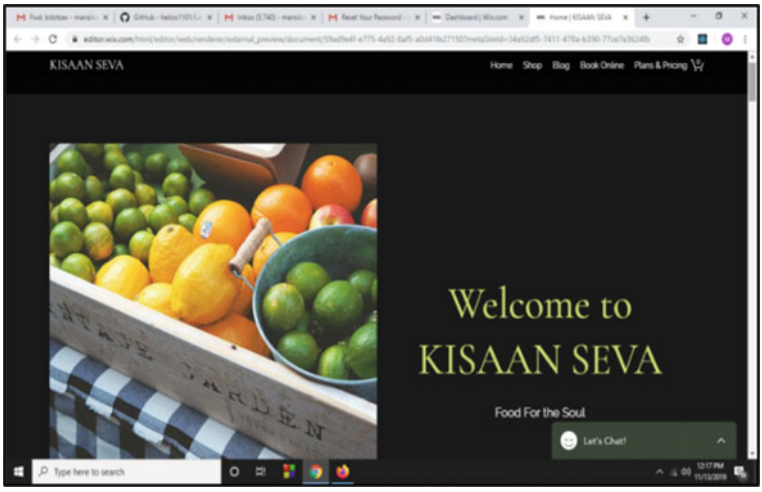


Fig. 1 Welcome page with sign in tab, home tab, add to cart tab and notification tab

As shown in Fig. 1, in the beginning, the Web site asks the user to 'Sign In' if he/she is already a member and if he/she is a new customer, then the Web site asks to 'Sign Up' for it. During signing up, the preferred language is asked from the customer so that we can assist the user in their language, and they feel more comfortable in using the Web site. On the home page of the Web site, there's a notification bar where Government can put up notifications that are intended for the farmers. The farmers will, therefore, remain updated with every amendments.

In Fig. 2, it has been shown that the farmers can put on display the available products and commodities. On clicking on the 'more' option, the consumer can view a wide range of available options with a particular seller. This has been shown in Fig. 3. Also, the consumer can view the types of services offered by the seller under the 'Services option' as shown in Fig. 4. As per Fig. 5, one can update their page with the upcoming commodities or if there is any change in the schemes or pricing, it can be notified to the user through this section. The farmers can update their blogs with different photographs as well as videos of the commodities they want to sell. This section can be viewed in Fig. 6. If a seller wants to open his/her shop for only a particular time interval, it can be mentioned in the 'Opening Hours' section, as shown in Fig. 7. It is set to '24 h open' by default. There are three different options to accept payments. Payments can be accepted through PayPal, Credit/Debit cards as well as cash, as preferred by the consumer. One can have a look at the payment section in Fig. 8. And most importantly, if a user has got any query regarding anything, he/she is free to contact the seller with the help of 'Contact Us' section. They can also give reviews regarding the products or also lodge complaints. This is also help the customers to efficiently and effectively choose the desired commodity or product

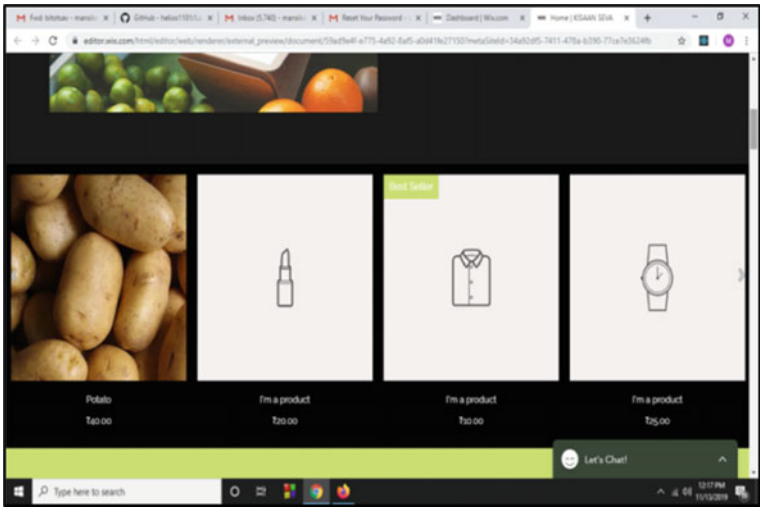


Fig. 2 A display of the available commodities

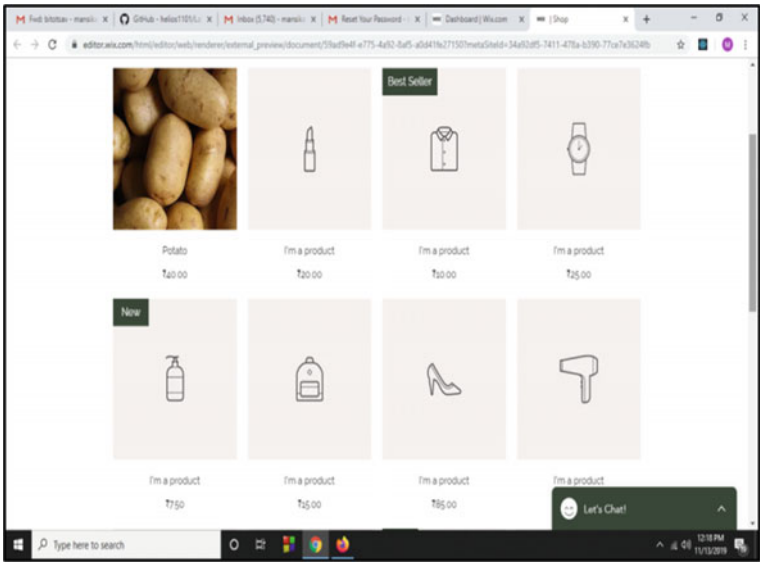


Fig. 3 A wide range of display of the products available along with their rates

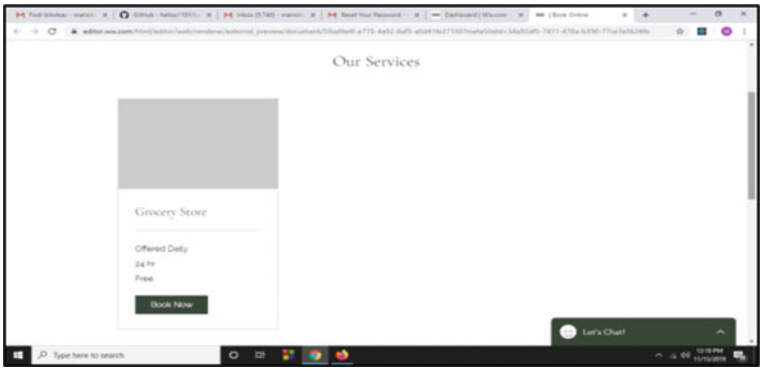


Fig. 4 Page showing different services available

from the right seller. And it will also create a sense of giving away the right product in the sellers. A fair environment will thus be maintained. This is shown in Fig. 9.

3 Results

The desired Web site with all the necessary details and requirements has been created.

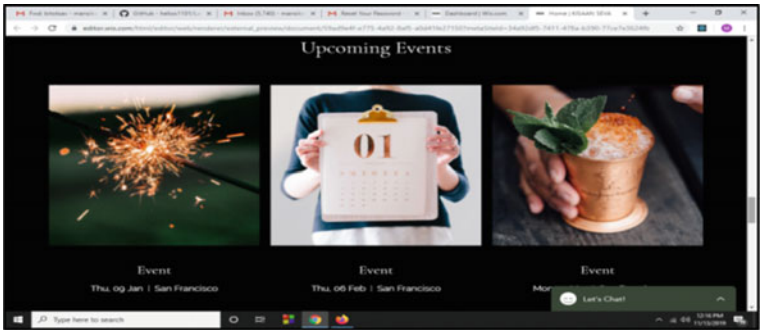


Fig. 5 Any upcoming event or new schemes can be added here

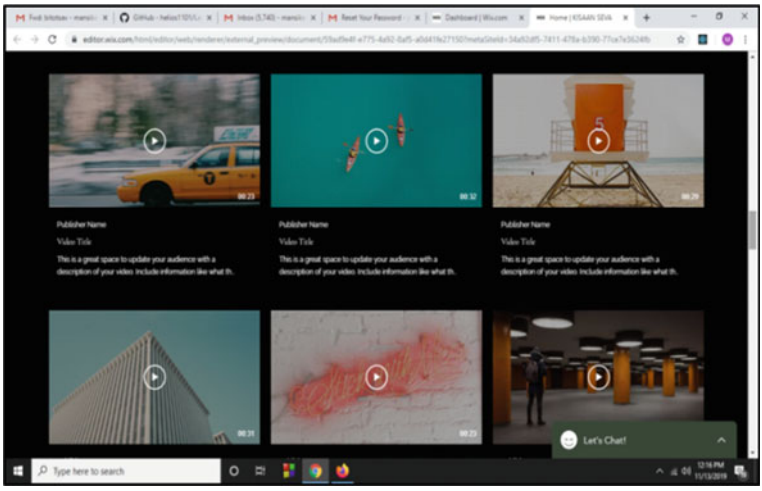


Fig. 6 Videos can be uploaded of different products

4 Discussions and Conclusions

The Web site is of great help for the farmers. Customers, vendors as well as Government can access the Web site. The Government can update the farmers with revised subsidies through notifications. Farmers can be updated and made aware of the grecent developments in the agriculture sector. They can also buy and sell the products online.

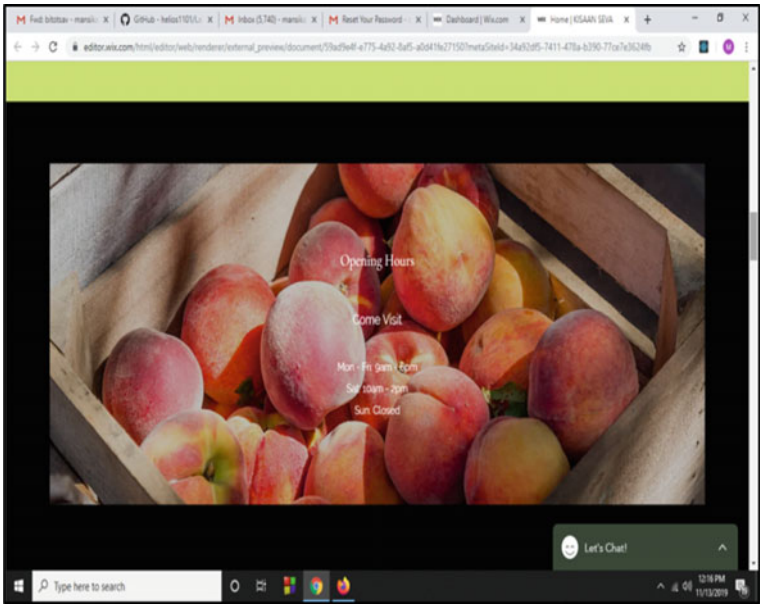


Fig. 7 Preferable time zone to allow shopping can be mentioned here

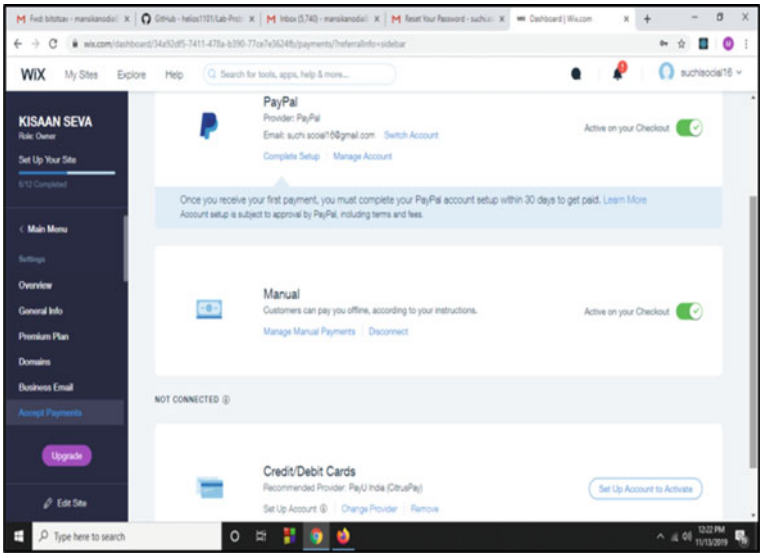


Fig. 8 Different payment methods

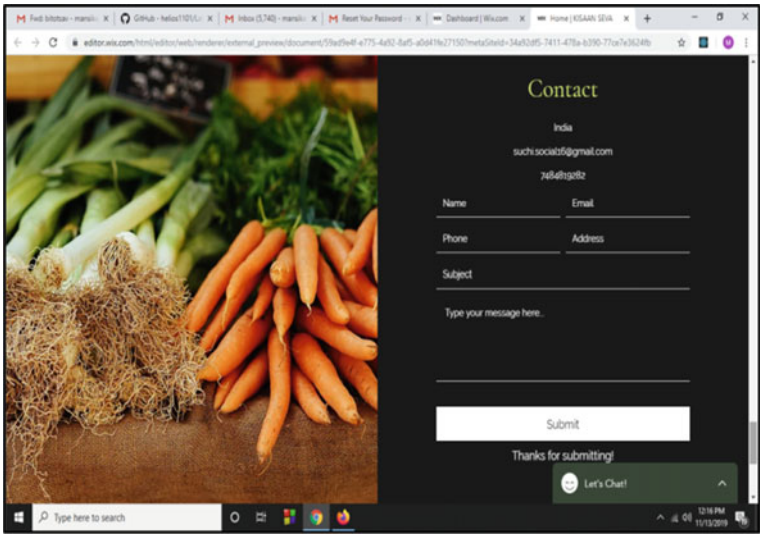


Fig. 9 User can ask queries though this section

Acknowledgements We would like to express our special thanks of gratitude to Dean Academic Program—Prof. Padmini Padmanabha, Dean Student Welfare—Prof. Anand Kr. Sinha, Head Of The Department—Prof. Srikant Pal, Vice Chancellor—Prof. S. Konar for providing infrastructure facility to carry out the my research work.

References

1. https://www.preservearticle.com
2. https://www.researchgate.net
3. https://www.allexamnotes.com
4. https://blogs.wjs.com
5. https://www.moneylife.in
6. https://www.downtoearth.org
7. https://www.spontaneousorder.in
8. https://www.academia.edu
9. https://www.agricop.gov.in
10. www.indiatoday.in
11. https://www.ncbi.nlm.nih
12. https://www.economictimes.com
13. www.uniindia.com
14. www.hindubusinessline.com
15. https://www.thebetterindia.com
16. https://theprint.in
17. https://linkspringer.com
18. https://moneylife.in
19. https://financialtribune.com

20. https://timesofindia.com
21. https://theepochtimes.com
22. https://linkedin.com
23. www.forbes.com
24. www.livemint.com
25. https://www.qrius.com

Design of Low Power Barrel Shifter Architecture by Using Proposed MUX Based CORDIC in CMOS Logic

Mohammed Wajid Khan, V. Vannammal Revathy, and S. Kaja Mohideen

Abstract Shifters usually play an important role in all the operations in VLSI computer systems. For example, barrel shifter is a fundamental block of many computing systems due to its operational operation that it can shift and rotate various multiple bits in one cycle. The barrel shifter can also be replaced for arithmetic and the logical shifters as it uses the rotation of the data. It also provides both the shifting of right and left arithmetically and logically. The design is purely MUX-based, and therefore, designing a MUX for low power to use it as a repetitive block in the barrel shifter will improve its efficiency. In this paper, an 8-bit barrel shifter using multiplexers (MUX) is implemented and is designed in cadence tool for demonstrating in power and its functional operations. The paper analyzes and optimizes the area and power of the barrel Shifter in 180 nm technologies. Cadence Virtuoso tool is used for implementing the block.

Keywords Barrel shifter · MUX · CMOS · DSP

1 Introduction

A. *Barrel Shifter*

The shifting of data in a circuit is one of the major requirements of key computer operations; whereas shifting of a single bit is slow process, barrel shifter is a combinational circuit with n data inputs and n data outputs with control inputs that specify the shifting of the input data as the output and bit barrel shifter require n, n-bit multiplexers. But n-bit barrel shifter requires n number of n-bit multiplexers. If n is increased the circuit complexity also increases, i.e., circuit overhead, it leads to occupy more area and high power consumption and also shows the effect on speed of the operation [1]. The latest technology

M. W. Khan (✉) · V. Vannammal Revathy · S. K. Mohideen
Deparment of Electronics and Communication, BS Abdur Rahman Crescent Insitute of Science & Technology, Chennai, India
e-mail: abwajidkhan114@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_62

used for constructing integrated circuits is complementary metal-oxide semiconductor (CMOS). The technology is being used in various digital and analog logic circuits.

Barrel shifter is a digital circuit that shifts data word by a specified number or bits [1]. It can be implemented as a sequence of multiplexers, and in such an implementation, the output of one MUX is connected to the input of the next MUX in a way that depends on the shift distance.

A barrel shifter forms a part of microprocessor CPU. Barrel shifter is often required for performing data shift and rotation in many key computer operations from address decoding to computer arithmetic. A significant reduction in area and power required by the barrel shifter circuit is achieved by implementing rightward operations as operations in leftward directions [2]. A significant reduction in delay is possible by reducing the length of critical path.

B. *Applications*

- Several microprocessors incorporate it as a part of their ALU to provide fast shift operations.
- Barrel shifter circuits are essential elements in the design of data paths for DSP applications.
- Barrel shifters are often required for performing data shifting and rotation in many key computer operations from address decoding to computer arithmetic [2].
- It is used extensively in floating-point units, scalars, and multiplications by constant numbers.
- A barrel shifter is a digital circuit that can shift a data word by a specified number of bits in one clock cycle. It can be implemented as a sequence of multiplexers (MUX), and in such an implementation the output of one MUX is connected to the input of the next MUX in a way that depends on the shift distance [2].

2 Literature Review

Many researchers have done work in it with different architectures.

In 2007, Khan et al. [3] realized the ternary Toffoli gate and modified Fredkin gate. Realization of the quantum circuits is done using generalized ternary gates and Feynman gates and then is being replaced with their equivalent realization using Muthukrishnan-Stroud gates. The ternary quantum gates realized by them are more efficient and have less quantum cost. In 2010, Kotiyal et al. [4] proposed a proficient architecture and design of a reversible ternary barrel shifter. The ternary barrel shifter is being realized using the modified Fredkin gates (MFG) and the ternary Feynman gates. In this, they used the multiple-valued reversible logic in order to design the (4:2) ternary barrel shifter which was being attempted first time in the literature. In order to calculate the quantum cost, they have used the MS-gate and shift gate which is being used in the implementation of reversible barrel shifter. In this quantum cost, ancilla bits and garbage output are being calculated. In 2012, Nia [2] proposed optimized

(4, 2) reversible bidirectional logical barrel shifter, optimized (8, 3) reversible right barrel shifter, and GRS-bit generation and optimized (8, 3) reversible normalization logical left barrel shifter for floating-point arithmetic for the first time in the literature. The proposed optimized binary shifters are designed using Feynman gates, Fredkin gates, and Peres Gates. Some parameters such as the amount of garbage outputs, the number of constant inputs, size of the circuit and quantum cost are being calculated. In 2015, Lisa and Babu [1] proposed the two new designs of ternary barrel shifter: First design is ternary unidirectional barrel shifter, and the second design is ternary bidirectional barrel shifter. They even proposed the ternary peres gate which is being used in the implementation of the proposed design the new technique to calculate the gate complexity is also being proposed. And the comparison of garbage output, ancilla bit and quantum cost are also being done. The compared results proved that the proposed work is more efficient than the existing work.

3 Existing Architecture

A barrel shifter is part of a microprocessor CPU which can typically specify the direction of shift left or right, the type of shift circular, arithmetic, or logical and the amount of shift (typically 1 to $n - 1$ bits, but sometimes 1 to n bits). Barrel shifters are generally used for the digital signal processors and general-purpose processors to manipulate the data [2]. Barrel shifter consists of an array of transistors, as shown in Fig. 5 in which the number of rows equals the word length of the data, and the number of columns equals the maximum shift width. The control wires are routed diagonally through the array [2].

The design of the barrel shifter is almost symmetric and can be done using repetitive combinational logic blocks. 2:1 multiplexer can be effectively used to design n-bit barrel shifter [5]. This shows the block schematic of a 2×2 barrel shifter using four 2:1 multiplexers. If each multiplexer block is optimized for energy dissipation, then the simulation time of the entire barrel shifter is reduced by a factor of $n\log 2n$ times because the simulation of only one multiplexer is enough to estimate the overall energy dissipation and delay [6, 7].

4 Proposed Architecture

In this paper, we proposed a barrel shifter which further implemented in DSP applications; firstly for making as block in cadence, we proposed inverter and by using different gates and we proposed MUX and then further made an entire block of barrel shifter using cadence tool. Figure 2 shows MUS-based 2×2 barrel shifter, and using this, rotate and shift operation is been utilized and barrel shifter is made as a block which can be used in many applications. Shifting of the data in the barrel shifter can be as per the user requirement [8].

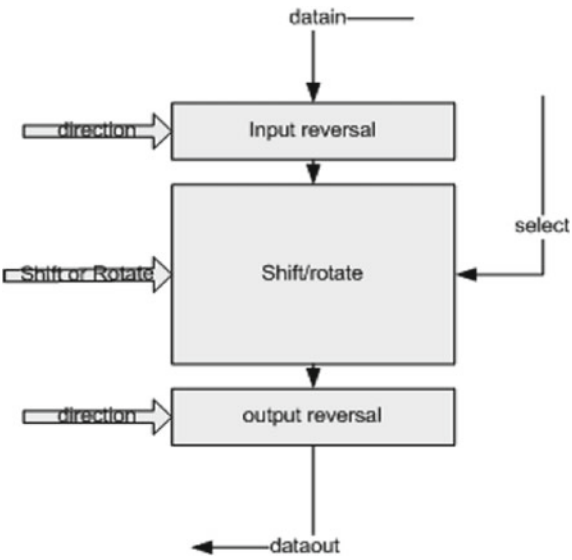


Fig. 1 Block diagram of existing barrel shifter

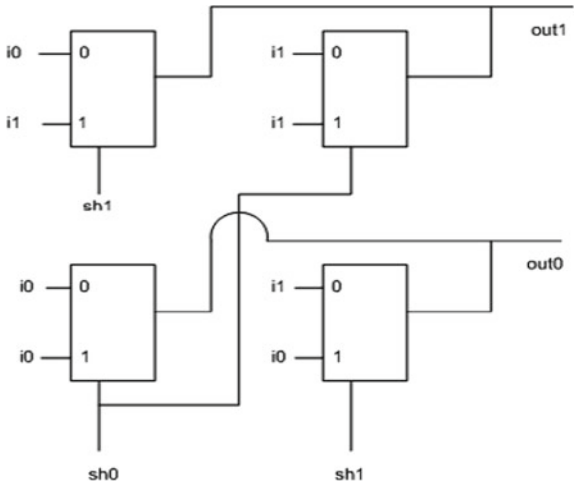


Fig. 2 MUX-based 2 × 2 barrel shifter

The left and right operations are implemented through inversion of the input and output vectors, so the basic multiplexing function can perform both operations. The number of multiplexing stages is relative to the width of the input vector [4]. A shifter/rotator performs both arithmetic shift and rotation operations [8].

The operation is shown by using block how the shift and rotate operation is been done as shown the block of operation from rotate and shift the test bench schematic is shown in Fig. 4; the example is explained by using 2 blocks by taking 10-bit data

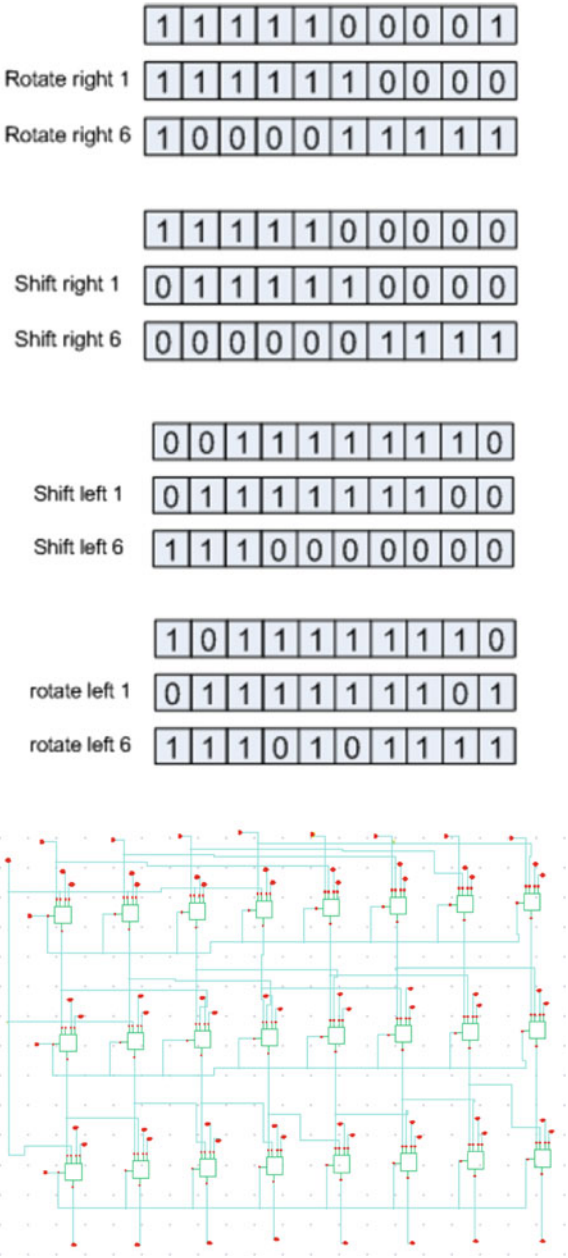


Fig. 3 Schematic of 8-bit logical right shifter

as 1,111,100,001, and the bits are rotated first and then shifted; similarly, the data obtained are checked with shift and rotate operation which gives similar output.

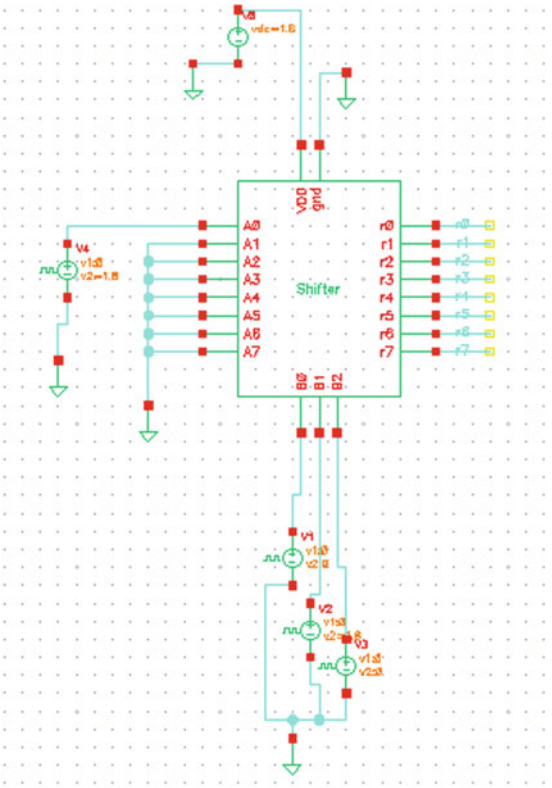


Fig. 4 Test bench of 8-bit logical right shifter

As shown in Fig. 3 schematic block, there is three stages of multiplexers for 4-bit, 2-bit, and 1-bit shift/rotation. A multiplexer output decides the operation to be performed, i.e., shift or rotation. Inputs to this MUX are 0 and a7 [4]. Logical right shift operation is performed if the MUX selects 0 [9, 10].

5 Comparison

The sinusoidal waves of frequency 400 MHz are applied as power clocks. For each circuit, binary inputs are applied to check the correctness of the circuits.

Power consumption of all circuits is measured at 400 MHz clock frequency.

The work is compared with the existed work, and in support, the power dissipation is further shown in Table 1.

Table 1 Comparing the power dissipation for all the three architectures

Architecture	Power dissipation (MW)	
	CMOS	Proposed technique (MUX-based)
Barrel right shifter	11.62	7.96
Barrel right rotate	16.46	12.33
Barrel right shifter/rotate	21.55	18.23

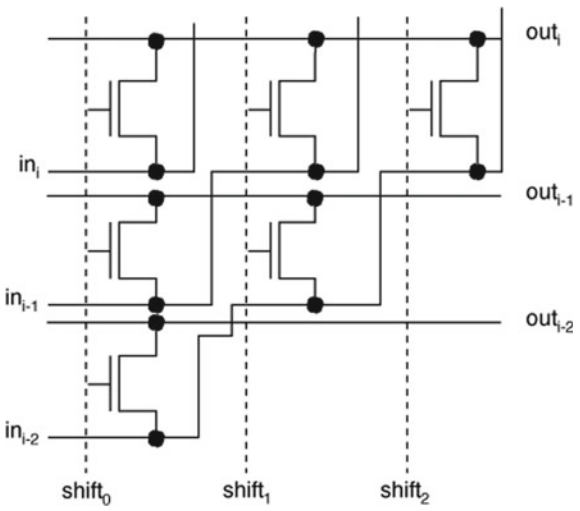


Fig. 5 Schematic of the barrel shifter

6 Designing and Implementation

The 8-bit barrel shifter circuit is simulated in 180 nm CMOS process technology in Cadence Virtuoso. Figure 5 shows the schematic representation of barrel shifter. It is clearly seen that for 8-bit barrel shifter, 6 transistors are used which compensate the area and power. The schematic shows the CMOS logic implementation by using various gates for both shift and rotate.

7 Results and Discussion

Simulation Result of 8-bit Right Shifter

The result shows that entire barrel shifter is simulated by using cadence, and the demonstration clearly shows that the block is working perfectly.

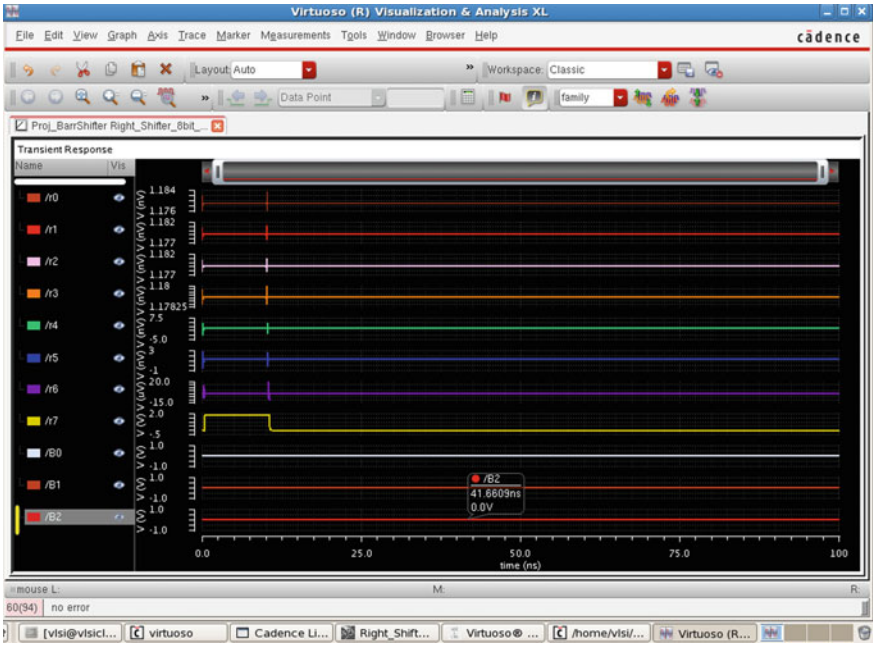


Fig. 6 Simulation result of zero-bit shifting

a. **Simulation Result of Zero-Bit Shifting:**
For the input a7a6a5a4a3a2a1a0 = 1000 0000, control inputb2b1b0 = 000.The output r7r6r5r4r3r2r1r0 = 1000 0000, Fig. 6
b. **Simulation Result of One-Bit Shifting:**
For the input a7a6a5a4a3a2a1a0 = 1000 0000, control input b2b1b0 = 001. The output r7r6r5r4r3r2r1r0 = 01,000 0000, Fig. 7
c. **Simulation Result of Two-Bit Shifting:**
For the input a7a6a5a4a3a2a1a0 = 1000 0000, control input b2b1b0 = 010. The output r7r6r5r4r3r2r1r0 = 0010 0000, Fig. 8
d. **Simulation Result of Four-Bit Shifting:**
For the input a7a6a5a4a3a2a1a0 = 1000 0000, control input b2b1b0 = 100. The output r7r6r5r4r3r2r1r0 = 0000 1000, Fig. 9
e. **Simulation Result of Seven-Bit Shifting:**
For the input a7a6a5a4a3a2a1a0 = 1000 0000, control input b2b1b0 = 100. The output r7r6r5r4r3r2r1r0 = 0000 0001, Fig. 10.

8 Conclusion

The 8-bit barrel shifter using multistage structure reduces the effect of junction capacitance and increases the speed of an operation for RISC processor. It can also

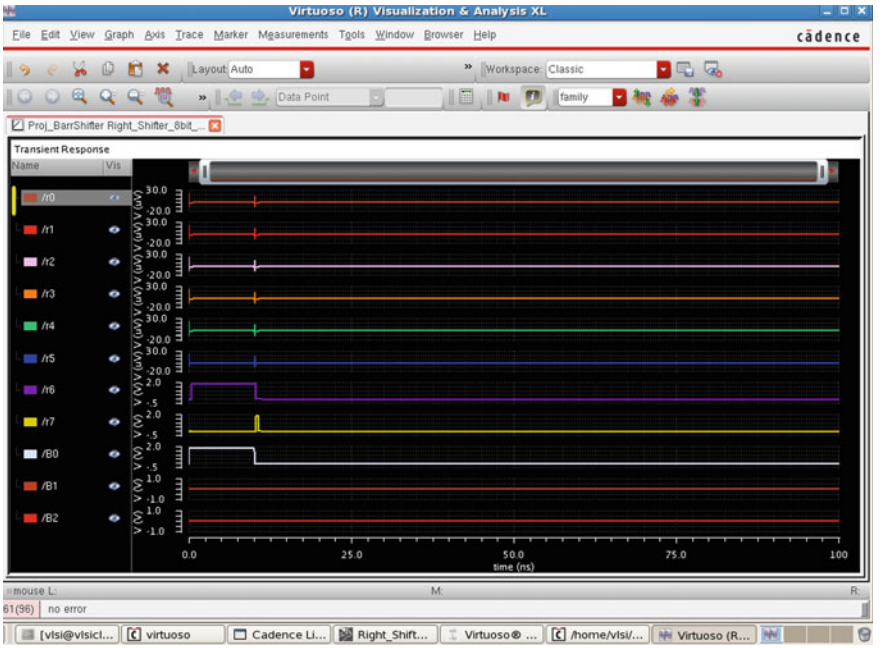


Fig. 7 Simulation result of one-bit shifting

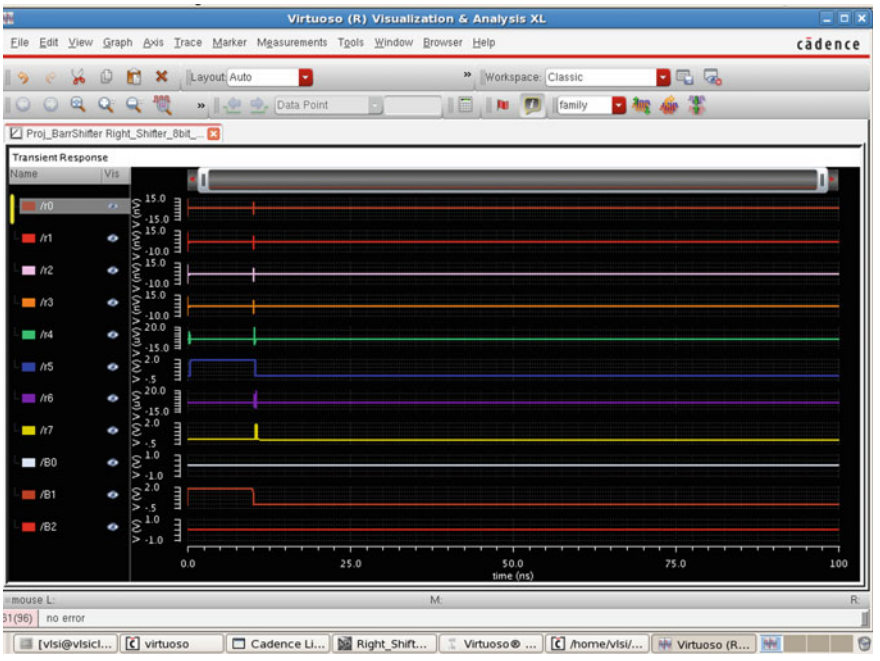


Fig. 8 Simulation result of two-bit shifting

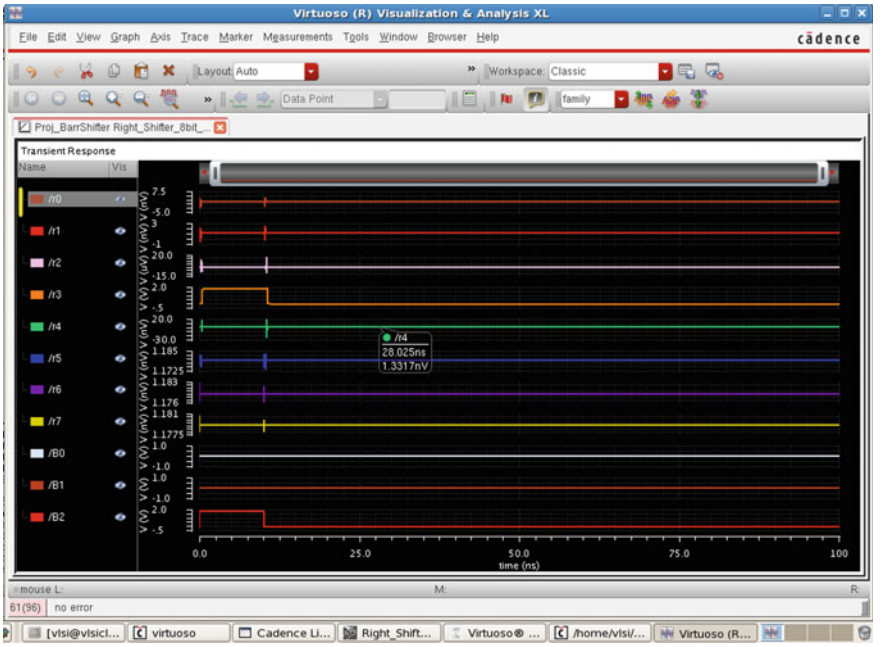


Fig. 9 Simulation result of four-bit shifting

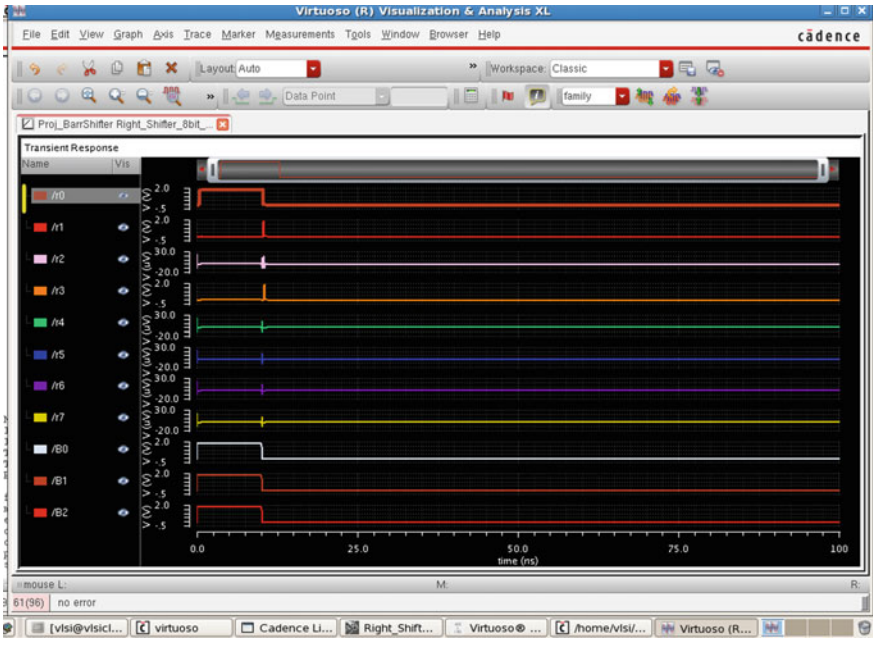


Fig. 10 Simulation result of seven-bit shifting

reduce the power consumption for low chip area to minimize the fabrication cost. In this paper, operation of barrel shifter was designed using cadence. The circuits are designed using static CMOS logic. CMOS logic can be used for the circuits where low power consumption is the main goal. Barrel shifters designed using CMOS logic can be used in DSP processors and microprocessor where high speed and low power are needed [11].

We have presented an approach to realize the unidirectional barrel shifter which consumes less power and the less no. of gates in its implementation. This work can form an important move in CMOS technology. It has the application in the various fields like quantum computing, nanotechnology, DSP technology, and in designing of low power CMOS circuit.

Reference

1. Lisa NJ, Hasan Babu HM A compact representation of a quantum controlled ternary barrel shifter. 978-1-4799-8391-9/15/$31.00 ©2015 ieee
2. Nia NH (2012) Design of an optimized reversible ternary and binary bidirectional and normalization barrel shifters for floating point arithmetic. Life Sci J 9(3)
3. Khan AI, Nusrat N, Khan SM, Khan MHA (2007) Quantum realization of some ternary circuits using Muthukrishnan-stroud gates. In: 37th international symposium on multiple-valued logic (ismvl'07) IEEE
4. Kotiyal S, Thapliyal H, Ranganathan N (2010) design of a ternary barrel shifter using multiple valued reversible logic. In: 10th IEEE international conference on nanotechnology joint symposium with nano Korea
5. Srujana V, Babu G (2013) Design methodology of bidirectional barrel shifter based on reversible logic. Int J Smart Sens Ad hoc Netw (IJSSAN) 3(1) ISSN no. 2248-9738 (print)
6. Wenyu X (2008) Verilog digital system design tutorial. Beijing: Beijing university of aeronautics and astronautics press, 212–215
7. Rajalakshmi R, Aruna Priya P (2014) Design and analysis of a 4-bit low power universal barrel-shifter in 16nm finfet technology. In: 2014 ieee international conference on advanced communications, control and computing technologies, Ramanathapuram, pp 527–532
8. Thomson T, Tam H (1997) Barrel shifter, US Patent 5,652,718
9. . Angelopoulos, Pitas I (1996) Fast parallel DSP algorithms on barrel shifter computers. IEEE Trans Signal Process 44(8):2126–2129
10. Pillmeier MR, Schulte MJ, George Walters E III Design alternatives for barrel shifters
11. Anuar N, Takahashi Y, Sekine T (2009) Adiabaticlogic versus cmos for low power applications. In: Proceedings itc-cscc 2009, pp 302–305

Efficient HDL Implementation of Turbo Coded MIMO-OFDM Physical Layer

Yanita Shrimali and Janki Ballabh Sharma

Abstract MIMO-OFDM is outstanding framework for fast data transmission over remote channel. In this chapter, turbo coded MIMO-OFDM (TC-MIMO-OFDM) physical layer implementation using VHDL is presented. In the proposed architecture, in order to achieve high throughput and low complexity, efficient Log-MAP turbo encoder/decoder and pipelined FFT processor is employed. Single delay feedback-based pipelined memoryless FFT/IFFT processor helps in achieving improved area and power efficiency, whereas high speed and good error correction capacity are obtained by Log-MAP turbo decoder. Simulation results obtained using Xilinx ISE Design suite are compared with state of the art architectures verifies the efficiency of the proposed system.

Keywords MIMO · OFDM · Turbo code · FFT

1 Introduction

To upgrade the speed and most prominent measure of data transmission MIMO-OFDM system have been created [1].

In MIMO-OFDM system framework, OFDM method is combined with multiple input multiple output signal processing method and has been adapted in different wireless communication application systems [2]. MIMO-OFDM is a communication technology having multiple antennas for transmission and receiving end with different coding and decoding schemes at input and output, respectively [2]. Error correcting coder such as viterbi encoding-based MIMO-OFDM system has already been developed [2–4]. In [3], MIMO-OFDM transceiver system has been designed

Y. Shrimali (✉) · J. B. Sharma
Department of Electronics Engineering, Rajasthan Technical University, Kota, India
e-mail: shrimaliyanita@gmail.com

J. B. Sharma
e-mail: jbsharma@rtu.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_63

using convolutional encoding and veterbi decoding. FPGA implementation architecture of MIMO-OFDM physical layer using conventional coding/decoding methods has been reported in [4].

Recently, turbo coding schemes have been utilized in MIMO-OFDM systems to achieve high throughput [5, 6].

However, SOVA and MAP decoding algorithm for turbo coding/decoding and their VHDL implementation have been reported [7–9]. MAP decoding scheme is considered as more efficient than SOVA [7–9]. FFT processors are used to generate sub-carrier in OFDM system, therefore, FFT processor is also vital in MIMO-OFDM physical layer system framework. Different architectures of FFT processor for OFDM application, such as parallel, pipelined with feedback (MDF and SDF), and with memory/without memory, have been developed [10]. An improved pipelined memoryless FFT processor architecture using single-path delay feedback scheme, and its efficient FPGA implementation has been reported in [11–13].

The major contribution of the proposed research work is the use of efficient pipelined single delay feedback (SDF) memoryless FFT processor architecture and turbo decoder using Log-MAP algorithm to improve the throughput, power and FPGA-based hardware requirement and complexity of MIMO-OFDM physical layer. In OFDM systems, the use of turbo encoder and FFT processor provide high throughput in comparison with convolution and other conventional codes [10]. Therefore, a software (MATLAB)-based scheme for turbo coded MIMO-OFDM (TC-MIMO-OFDM) physical layer has been reported in [5, 6]. Turbo coded MIMO-OFDM systems with different receiver and detector architectures have been reported in [14–16]. Similarly, the performance of OFDM system improved by the use of improved FFT processor architectures in comparison with conventional architectures. Low power FPGA implementation of TC-OFDM physical layer has been reported in [10], but, FPGA implementation of TC-MIMO-OFDM physical layer has not been explored.

This paper presents VHDL realization of turbo coded MIMO-OFDM physical layer framework. Proposed system uses efficient FFT/IFFT processor and MAP algorithm-based turbo coder. Xilinx ISE Design suite 14.2 is used implement the proposed system and the simulation results are compared with [4]. Efficacy of the proposed design is verified by the comparison analysis results.

Organization of the chapter is as follows, in Sect. 2 review of MIMO-OFDM system model is described. Details of the proposed TC-MIMO-OFDM architecture are described in Sect. 3. Synthesis and simulation results are presented in Sect. 4, whereas in Sect. 5 conclusion of the work is presented.

2 Overview of MIMO-OFDM

Conventional wireless communication systems, single input and single output (SISO) signal processing technique are used. But, this method suffer with the limitation that, its capacity cannot be increased to a great extent without increasing the transmitted

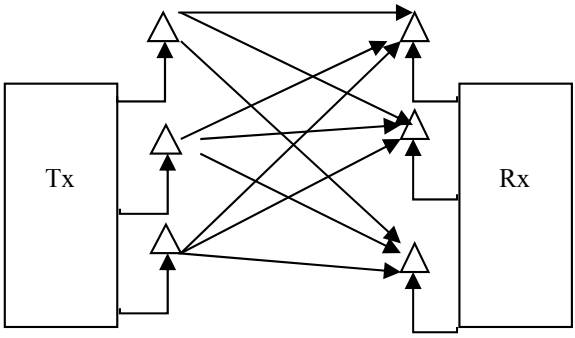


Fig. 1 MIMO principle

spectrum power [1, 2]. OFDM system is also a SISO system. Moreover, MIMO system uses multiple antennas for transmitting and receiving (as shown in Fig. 1), with appropriate channel decoding scheme [2]. A MIMO system exploits the spatial diversity in multipath scattering system, in order to enhance diversity gain, capacity gain and to reduce signal fading. Multiple types of realizations of MIMO system have been reported and details are available in [1].

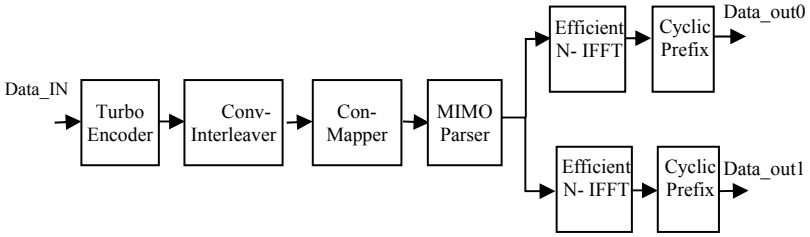
3 Proposed Efficient Turbo Coded MIMO-OFDM Physical Layer

This section describes proposed TC-MIMO-OFDM physical layer system framework. In the proposed system, implementation, pipeline memoryless SDF architecture-based FFT/IFFT processor [10] and turbo coder with Max-MAP algorithm [7] are employed. Proposed system architecture is depicted in Fig. 2, where Fig. 2a depicts the transmitter and Fig. 2b shows the receiver system architecture.

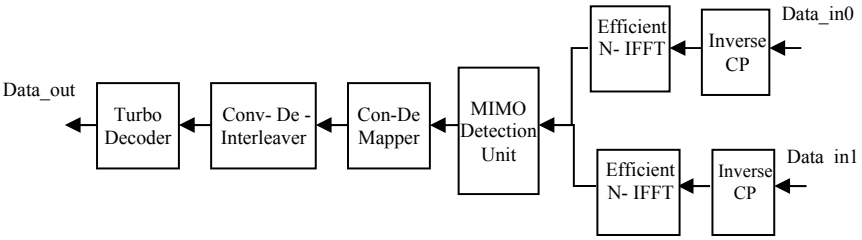
To obtain improved performance in OFDM physical layer transmitter and receiver, turbo coder based on MAP algorithm and FFT/IFFT processor without memory and based on SDF pipelining process is used. Used FFT/IFFT processer configuration reduces the power requirement and system complexity. High throughput, low complexity, high speed, etc., are some of the advantages offered by the proposed system architecture.

At transmitter, two input data sequences are encoded firstly by (rate 1/3) turbo encoder and parity bit0 and bit 1 are produced according to rate systematic bit Fig. 3a. The encoded bits are interleaved by a block convolution interleaver shown in Fig. 3b.

Thereafter, in order to minimize burst error probability, block convolution interleaver is employed to interleave the encoded data bit. This is followed by a quadratic amplitude modulation (16-QAM) constellation mapping, this data is mapped to QAM symbols in interleaved form [6]. First, pilot symbol is inserted according to pilot pattern. The mapped information is fed to parser split into two output using spatial multiplexing [2]. Proposed system is 2×2 MIMO-OFDM physical layer system



(a) Proposed 2X2 efficient TC-MIMO-OFDM transmitter



(b) Proposed 2X2 efficient TC- MIMO-OFDM receiver

Fig. 2 Proposed efficient turbo coded MIMO-OFDM physical layer transceiver

so it has two transmitting and two receiving antenna at the transmitter to transmit information and receiver to receive information. Efficient IFFT receive information from MIMO parser using spatial multiplexing and then fed to cyclic prefix. N data bit-sized cyclic prefix is inserted to avoid the inter-symbol interference [4].

At 2 × 2 MIMO-OFDM receiver, receivers are received information from two different transmitting antennas at inverse cyclic prefix (CP) block to remove preamble after that it convert into frequency domain.

Next, the information data are collected at MIMO detection unit and inverse operations of multiplexing, mapping and interleaving are performed. System is then decoded by a turbo decoding process, by employing a soft-decision MAP decoder [5] turbo decoder. Most likely transmitted sequence for a received data sequence is determined by the turbo decoder using Max-MAP algorithm [7].

The details of main blocks used in the proposed TC-MIMO-OFDM are as follows:

Turbo Encoder/Decoder

Turbo encoders/decoder (shown in Fig. 3a) use parallel link strategy-based recursive systematic convolutional (RSC) encoders isolated by a interleaver [7].

Rate 1/3 and coefficients [111; 101] are chosen for the conventional encoder.

MAP decoder-based turbo decoder is employed to decode input sequence bit stream. Max-Log-MAP algorithm [6] is used with each MAP decoder [5]. Inputs like parity bit (par0), systematic bit(sys) and deinterleaved bit (w3) are used in MAP

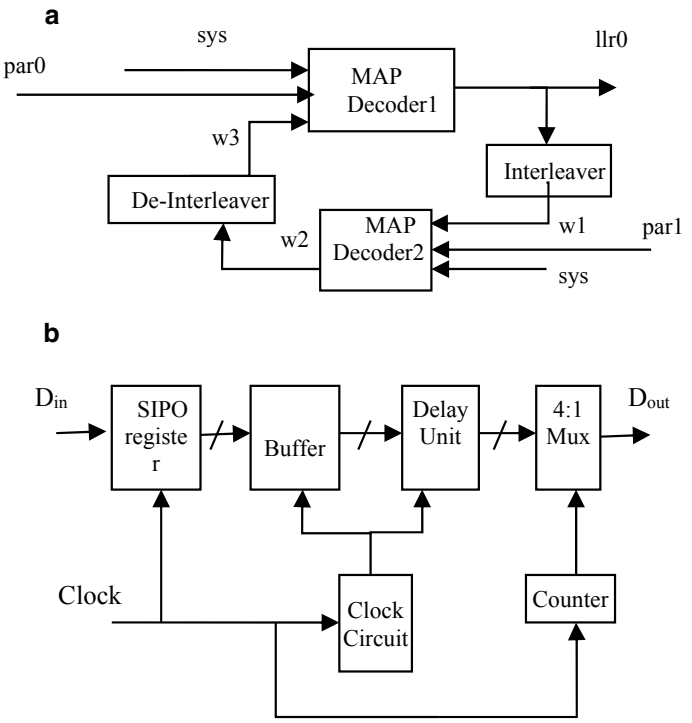


Fig. 3 **a** MAP algorithm-based turbo decoder block diagram. **b** Block diagram of convolutional interleaver

decoder1 to generate llr0 (extrinsic information).Thereafter, the output of the MAP decoder 1 is applied to MAP decoder 2 after interleaving.

In decoder 2 parity bit 1(par1), input systematic bit sys) and interleaved bit (w2) are used. Inverse interleaving operation is performed by deinterleaver. The deinterleaved information (w3) of decoder 2 output w2 is applied to decoder 1. In order to decode the received input sequence, above steps are repeated iteratively.

Interleaving/De-Interleaving

As mentioned in [4–7], "interleaving process protects the data from burst errors during transmission. For this, the incoming data is broken into blocks and the bits within a block are rearranged such that adjacent bits become non-adjacent to each other. By this adjacent bits within an OFDM symbol are placed on non-adjacent sub-carriers." At the receiver, de-interleaving rearranges the bits into original sequence.

Constellation Mapper/De-mapper

OFDM is digital modulation as well as multiplexing technique. So we have chosen PSK as digital modulation technique. 16-PSK (16-QAM) has been used to evaluate the efficacy of the proposed system.

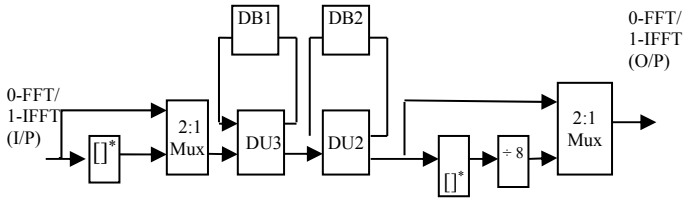


Fig. 4 Four-point SDF pipelined FFT/IFFT processor

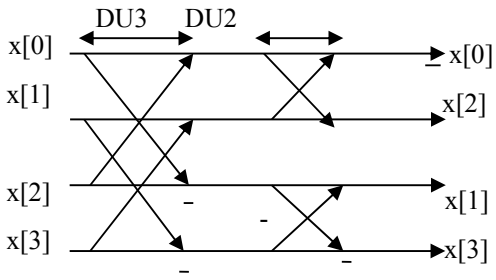


Fig. 5 Radix-2 DIF butterfly diagram of four-point FFT

Cyclic Prefix/Inverse Cyclic Prefix

The cyclic prefix acts as a guard interval to protect the OFDM signals from inter-symbol interference. It is created by copying the end part of the symbol. Inverse cyclic prefix is used to remove prefixed data bits that are used at the transmitter.

Efficient FFT/IFFT Processor

FFFT/IFFT is the key of MIMO-OFDM physical layer systems for generation of multiple sub-carriers. For this, fast speed, low complexity processors are required. Therefore, a four-point memoryless FFT processor with single-path delay feedback pipeline configuration is used in this proposed system. Figure 4 depicts pipelined FFT/IFFT architecture.

Four parallel 16 bit input sequence information obtained from the parser subblock is applied to 2 × 1 multiplexer with the help of selection line (0). Further, processing is performed with the help of element unit (DU1, DU2 and DU3) and the respective delay-line buffers, details of the architecture are available in [14].

Further, for FFT operation at TC-MIMO-OFDM transmitter, it is received at the end of the 2 × 1 multiplexer (2:1 mux).

For IFFT operation at TC-MIMO-OFDM receiver selection line of multiplexer is 1, and one additional processing unit is employed to perform conjugate operation ([]*). For this, 2s complement is applied on the imaginary part of complex valued number. Hardwired shift operation is utilized to implement divide by four operations. Proposed FFT/IFFT system details are available in [11–13] for ready reference.

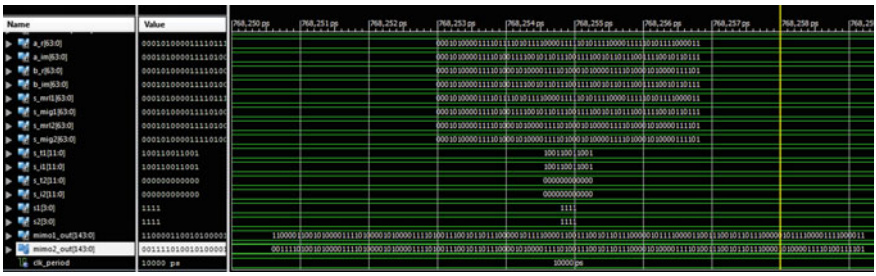


Fig. 6 Turbo coded MIMO-OFDM transmitter simulation result

The architecture as shown in Fig. 5 has two butterfly desire unit DU3 and DU2. The DU2 and DU3 are reported in [12, 13]. The architecture is using complex constant multiplier with reconfigurability, which eliminates the need of ROM.

4 Implementation Results and Comparative Analysis

Proposed TC-MIMO-OFDM transmitter and receiver is implemented using Xilinx ISE simulator ISIM 14.2. Figures 6, 7, 8 and 9 show the simulation and synthesis results for 143 sub-carriers with using rate (1/3) coding. Figure 8 shows that in the implemented TC-MIMO-OFDM receiver 4 bit output is generated from received 143 sub-carriers. 127 subcarrier convert into frequency domain to time domain or vice-versa through efficient 4 point 16 bit FFT/IFFT processor and 16 bit preamble insert by cyclic prefix. To modulate the sub-carrier, 16-QAM modulation technique is used. Simulation and synthesis results of intermediate steps are shown in Figs. 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21.

Table 1a, b shows the simulation results, where in Table 2, the comparative analysis of the proposed design are presented. It is evident that the proposed design has low complexity and overall low hardware requirement.

5 Conclusion

In this work, VHDL implementation of fast and efficient TC-MIMO-OFDM system model is reported. Turbo decoder and efficient four-point radix-2 FFT processor are employed in the proposed design to investigate their performance in MIMO configuration. The simulation results show that the proposed TC-MIMO-OFDM architecture is having improved throughput and complexity with reduced hardware. Proposed design can be utilized in wireless communication applications.

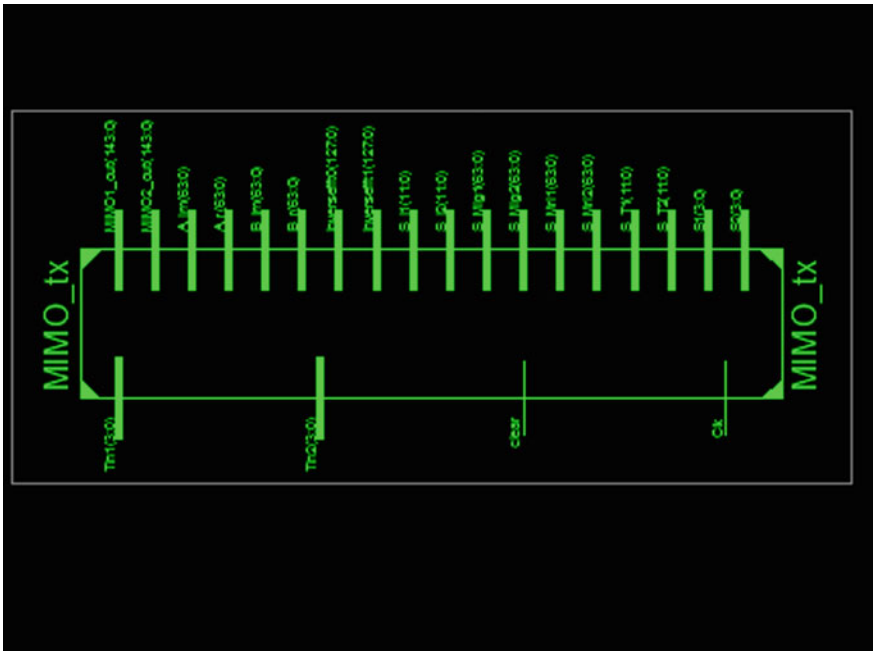


Fig. 7 Turbo coded MIMO-OFDM transmitter synthesis result

Fig. 8 Turbo coded MIMO-OFDM receiver simulation result

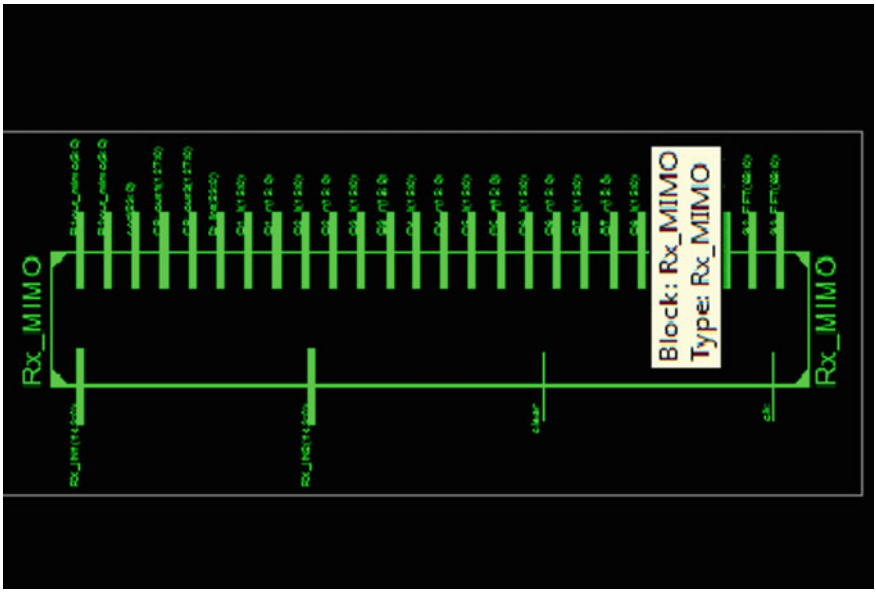


Fig. 9 Turbo coded MIMO-OFDM receiver synthesis result

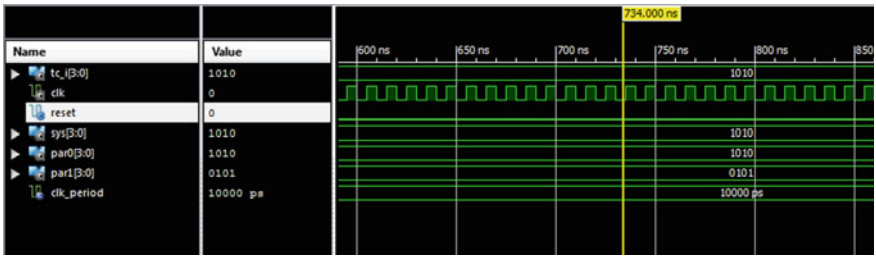


Fig. 10 Simulation result of turbo encoder (1/3 rate)

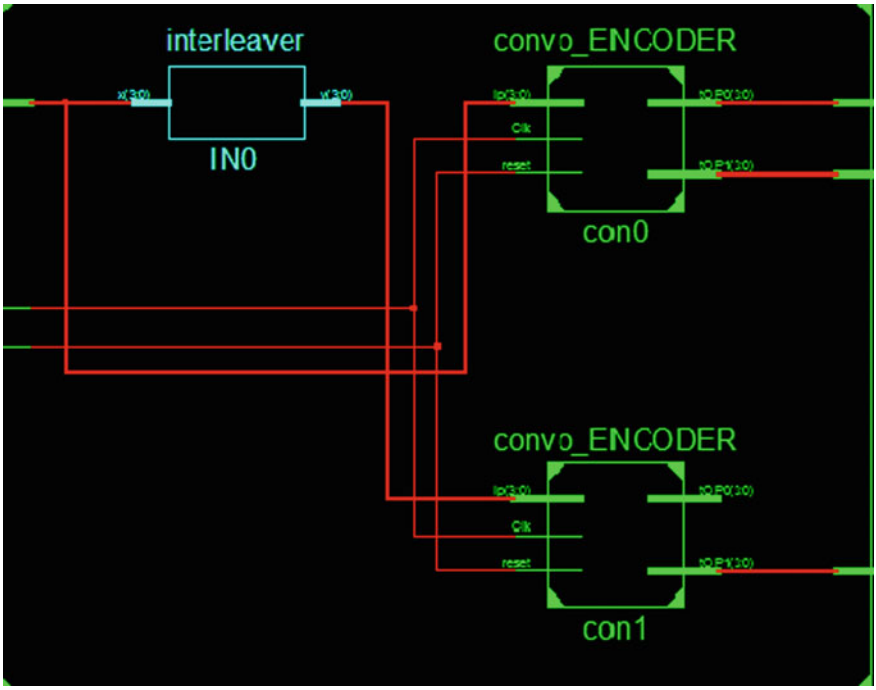


Fig. 11 Turbo encoder of TC-MIMO-OFDM physical layer transmitter system synthesis result

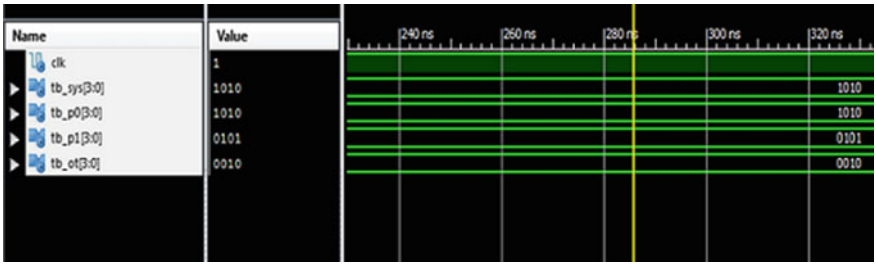


Fig. 12 Turbo decoder simulation result

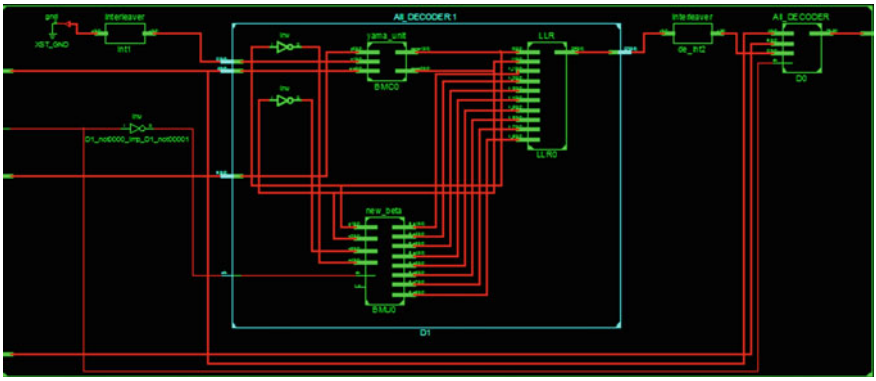


Fig. 13 Synthesis result of turbo decoder of TC-MIMO-OFDM physical layer receiver system

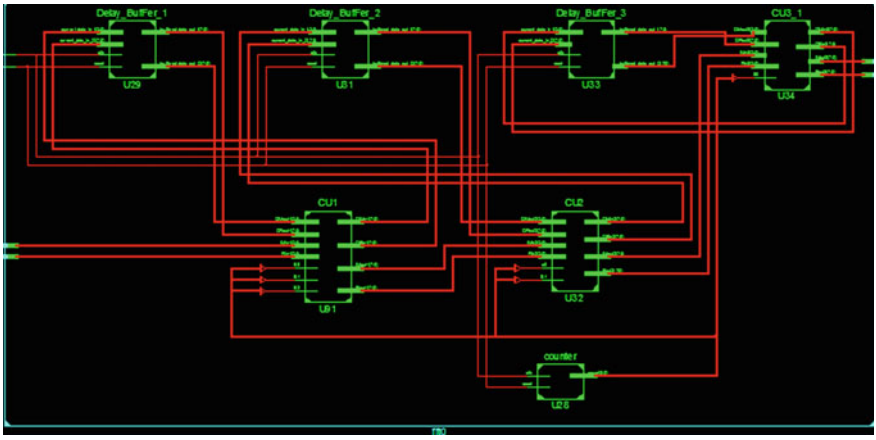


Fig. 14 Efficient FFT pipeline processor of TC-OFDM physical layer receiver system synthesis result

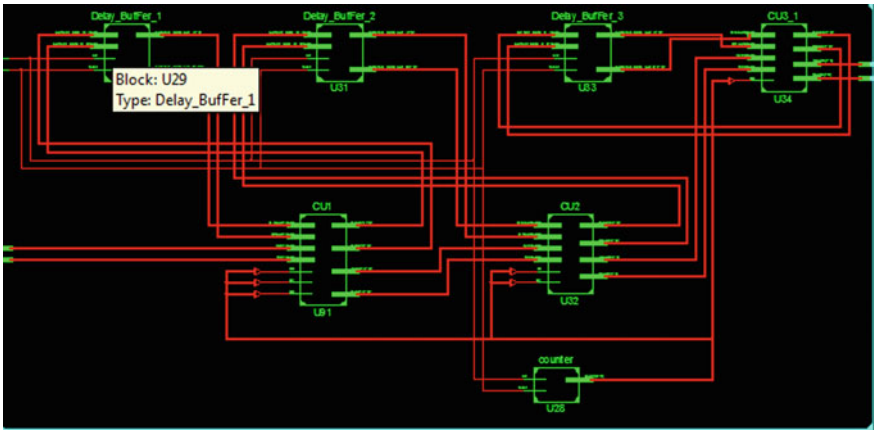


Fig. 15 Efficient IFFT pipeline processor synthesis result

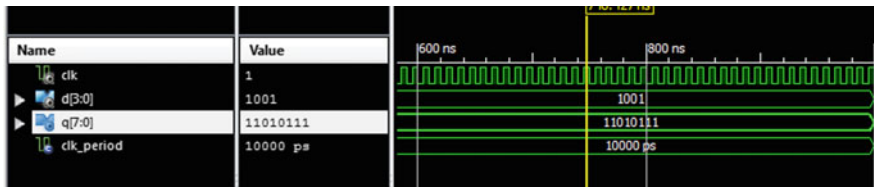


Fig. 16 Mapper simulation result

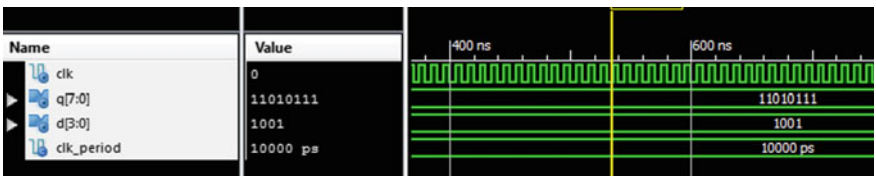


Fig. 17 Simulation result of de-mapper

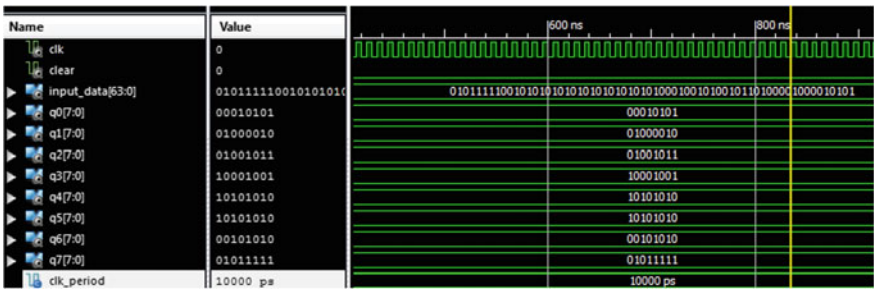


Fig. 18 SIPO simulation results

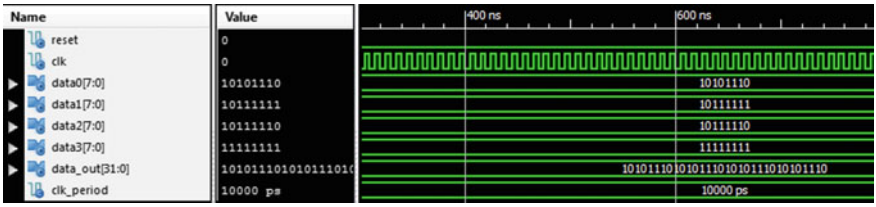


Fig. 19 PISO simulation results

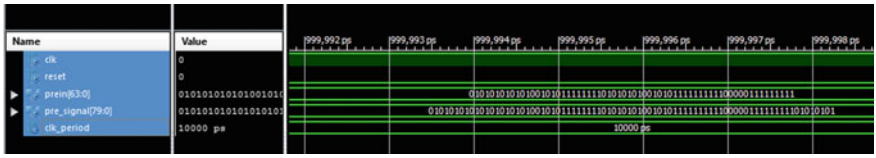


Fig. 20 Cyclic prefix simulation result

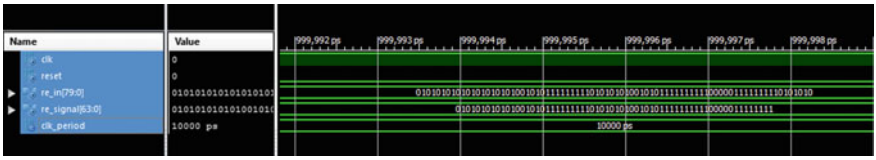


Fig. 21 Inverse cyclic prefix simulation result

Table 1 **a** Proposed turbo coded MIMO-OFDM transmitter device utilization summary, **b** proposed turbo coded MIMO-OFDM receiver device utilization summary

Device utilization summary (estimated values)			
Logic utilization	Used	Available	Utilization (%)
(a)			
Number of slices	4557	10,752	42
Number of slice flip flops	2912	21,504	13
Number of 4 input LUTs	8261	21,504	38
Number of bonded IOBs	1122	448	250
Number of GCLKs	1	32	3
(b)			
Number of slices	4287	10,752	39
Number of slice flip flops	2148	21,504	9
Number of 4 input LUTs	8078	21,504	378
Number of bonded IOBs	1082	448	241
Number of GCLKs	1	32	3

Table 2 Comparative analysis of proposed scheme

Design	No. of slices	No. of flip flop	LUTs	RAM/ROM
(a) Transmitter				
[4]	3176	4876	4621	13
[17] Total (Transmitter and Receiver)	28,571	–	–	134
[18]	3090	4784	4505	11
[19]	3087	4790	4509	11
Proposed	4557	2912	8261	0
(b) Receiver				
[4]	3622	5032	4899	10
[17]*	Included with transmitter			
[18]	3282	4692	4389	8
[19]	3282	4820	4523	8
Proposed	4287	2148	8078	0

*Values of total hardware (transmitter and receiver) requirement is provided in Table 2a

References

1. Tsoulos G (2006) MIMO system technology for wireless communications, 1/E, CR press, Boca Raton
2. Mahmudul Hasan M, Sagar SM (2013) Wireless communication through long term evolution (LTE) over satellite channel by using MIMO-OFDM model. In: 2nd international conference on advances in electrical engineering (ICAEE 2013), pp 170–175
3. Srinandhini K, Vaithianathan V (2014) FPGA implementation of MIMO-OFDM transceiver. In: International conference on communication and signal processing, pp 353–357
4. Park J, Tokunbo O (2012) Efficient FPGA-based Implementations of MIMO-OFDM physical layer. Circuits Syst Signal Process 31(4):1487–1511
5. Mathews L, Pillai SS (2016) Performance of turbo coded FBMC based MIMO systems. IEEE Int Conf Comput Intell Comput Res 2016:1–5
6. Gede Puja Astawa I, Kurniawan Y, Pratiarso A, Yoedy M, Hendy B, Ahmad Z (2017) Performance of single-RF based MIMO-OFDM 2 ×2 using turbo code. In: 2017 IEEE international conference on radar, antenna, microwave, electronics, and telecommunications, pp 83–88
7. Shahabuddin S, Janhunen J, Bayramoglu MF, Juntti M, Ghazi A, Silv´en O (2012) Design of a unified transport triggered processor for LDPC/Turbo decoder. In: IEEE international conference on embedded system, pp 288–295
8. Vijyata RS, Meena, Sharma JB VHDL implementation of low-power turbo decoder. In: International conference on electrical, computer and electronics (UPCON-2016) proceedings by IEEE, 9th-11th December 2016. https://doi.org/10.1109/UPCON.2016.7894723
19. Vijyata RS, Meena, Sharma JB Performance analysis and implementation of high speed MAP decoder for long term evolution. In: International conference on micro electronics and telecommunication engineering (PIICON 2016) proceedings by IEEE, 25th–27th November 2016
10. Shrimali Y, Sharma JB, Meena RS (2018) VHDL implementation of low power turbo coded OFDM physical layer for wireless communication. Int J Eng Technol 7:955–958 (In Press)
11. Kalra B, Sharma JB (2017) Vedic multiplication based efficient OFDM FFT processor. J Intell Fuzzy Syst 32(4):3121–3128

12. Gupta T, Sharma JB (2017) A high speed single path delay feedback pipeline FFT processor using vedic multiplier. In: IEEE international conference on information, communication, instrumentation and control (ICICIC-2017)
13. Yu C, Yen MH (2011) A low-power 64-point pipeline FFT/IFFT processor for OFDM applications. IEEE Trans Consumer Electron 57(1):40–45
14. Jin WW, Xioi G, Fu WX, Hu YX, Ming ZC, Kai-Kit W (2012) Dual-turbo receiver architecture for turbo coded MIMO-OFDM systems. Sience China Inf Sci 55(2):384–395
15. Tomasoni A, Siti M, Ferrari M, Bellini S (2010) Low complexity, quasi-optimal MIMO detectors for iterative receivers. IEEE Trans Wirel Commun 9(10):3166–3177
16. Iruthayanathan N, Vishvaksenan KS, Rajendran V, Kumar SM (2015) Performance analysis of turbo-coded MIMO–OFDM system for underwater communication. Comp Electr Eng 43:1–8
17. Haene S, Perels D, Burg A (2001) A real-time 4-stream MIMO-OFDM transceiver: system design, FPGA implementation, and characterization. IEEE J Sel Areas Commun 26(6):877–889
18. Park JS, Jung H, Prasanna VK (2006) Efficient FPGA-based implementations of the MIMO-OFDM physical layer. In: Proceeding of the international conference on engineering of reconfigurable systems and algorithms, pp 273–279
19. Park JS, Ogunfunmi T (2010) FPGA implementation of the MIMO-OFDM physical layer using single FFT multiplexing. In: Proceeding of IEEE international symposium on circuits and systems (ISCAS)

Design of a 2–30 GHz Low-Noise Amplifier: A Review

Krishna Datta, Srikanta Pal, and Vijay Nath

Abstract LNA is the first component after an antenna in any RF receiver system and therefore plays a key role in determining most of the key parameters of the RF receiver like noise figure, gain, linearity, and stability. For designing an LNA, its effect on every of these parameter has to be kept in mind. In this paper, the steps of designing a wideband LNA in the RF band 2–30 GHz have been studied with references from various previous works done on wideband LNAs and some application-based approaches on the discussed band. The technology studied has been restricted to Cadence CMOS technology.

Keywords LNA · Wideband · UWB · ZigBee · Bluetooth · Topology

1 Introduction

As the want of data with higher and higher speed is increasing, there comes the need for highly sensitive receivers with very large bandwidth for faster wireless communication systems. The difficulty in maintaining a decent trade-off between noise, gain, bandwidth, linearity, stability, and power consumption with a decreased size for wideband topology can be improved by the performance of an LNA. To improve these trade-off, different researchers have studied and published various topologies and designs for wideband LNAs. A low-noise amplifier (LNA) is an electronic amplifier, which is found in RF transmitter or receiver, and its job is to amplify a very low-power signal and not degrading its signal-to-noise ratio significantly. LNAs are used to amplify weak signals that are just above the noise floor.

K. Datta (✉) · S. Pal · V. Nath
VLSI Design Group, Dept. of Electronics and Communication Engineering, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: krishnadatta596@gmail.com

S. Pal
e-mail: spal@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_64

Table 1 Standard radar frequency letter band numenclature

Band designation	Nominal frequency range (GHz)
S	2–4
C	4–8
X	8–12
Ku	12–18
K	18–27
Ka	27–40

$$\text{Freciever} = \text{Flna} + \frac{\text{Frest} - 1}{\text{Glna}} \tag{1}$$

By using LNA so close to the signal source, the effect of the noise from the subsequent stages of the receiver chain in the circuit is reduced by the signal gain created by an LNA. The LNA design is important because of various reasons which are discussed below:

(i) NF of the receiver system is mostly dominated by the NF of the LNA,
(ii) The LNA gain helps in reducing the NF, that is contributed by the rest of the components.
(iii) Linearity of the LNA can control how much the receiver sensitivity is compromised.

This can be proved by the Friis's noise formula [1] given in Eq. (1). A good LNA has low noise figure, good gain, and a large enough intermodulation and compression point (IIP3 and P1dB) and impedence matching. The frequency range 2–30 GHz includes.

S, C, X, Ku, K, Ka bands as defined in the IEEE standard letter designations for radar frequency bands (Table 1).

So, the LNA designed for the whole band 2–30 GHz will be giving good performance for all the above-defined frequency bands. Some examples of distributed amplifiers are shown in [2–4] for their wide input matching characteristics. But these cases suffer from high power consumption, low gain, and large chip size. In [5], a UWB LNA is designed, whose input network in embedded in three-section Chebyshev filter, which had helped achieve good gain, linearity, low noise with good wideband performance but increased chip area. In [6], there is cascade of two different topologies.

This paper is organized as follows. Section 1 gives the working idea of an LNA and introductory gist of the reference literatures which have been studied and reviewed here. Section 2 gives the core LNA topologies that have been most used until now for wideband applications. Section 3 covers some ways of the processes used to improve upon the performance parameters over those core topologies. Section 4 gives the comparison of all these works, in order to determine the best topology. Section 5 is the conclusion drawn from this study. Section 6 is the references that have been used.

2 LNA Topologies

Low-noise amplifier is the first block of any RF receiver and therefore is the most important par also. It should be matched with the antenna output for the maximum transfer of power, because the already weak signal received by the antenna cannot be compromised further by poor matching. The properties of an antenna are its excellent input and output matching and large gain. In order to optimize the low-noise amplifier design, a topology should be selected, suitable for low-power and low-voltage applications. For optimum performance, common topologies used for LNA design are discussed. In case of shunt series feedback common source topology, it exhibits trade-off among gain, small noise figure, low power consumption, decent input and output matching. On the other hand, for common gate topology, the gain is lesser than 10 dB and shows very low power consumption and significant noise, which is contributed by the MOSFET. For resistor termination common source topology, the noise figure is increased by the thermal noise of the resistors. The minimum noise figure possible in this topology is 3 dB. In inductive degeneration common source topology, it shows low power consumption but the isolation is improved in cascode inductor source degeneration topology which can give the same LNA performance but with very low power consumption. The cascode inductor source degeneration topology gives more gain and a less noise figure. Therefore, there are several basic types of topologies for low-noise amplifier. Figure 1 below shows the topologies of LNA:

After deciding on the suitable topology, the corresponding input and output matching network is designed and maximum transfer of power from antenna is to the amplifier and from the amplifier is to the next stage of the circuit (Fig. 2).

3 3. Wideband LNA Topologies

The most common core topologies that have been used for wideband applications for LNA design are (1) inductive degenerated common source topology (IDCS) and (2) common gate(CG) topology. Some literatures that have used IDCS as core topologies are [7, 8] and some literatures that have used common gate are [9]. The CG has much better noise performance at higher frequencies, when compared to the IDCS topology. But at lower frequencies, the IDCS topology shows better noise performance. One more advantage of IDCS topology over CG topology is that it mostly shows more gain caused by the input matching network. In IDCS, the source impedance is matched by the series resonance while in CG topology a parallel one is used. The quality factor is generally greater for a series resonance than the one at the resonance frequency, thus increasing the transconductance of the input transistor. And also, the quality factor of a parallel resonance is same as the one at the resonance frequency. Yet, IDCS becomes more sensitive to parasitic and process variation since it has high quality factor at the input matching network. In CG topology, maximum of

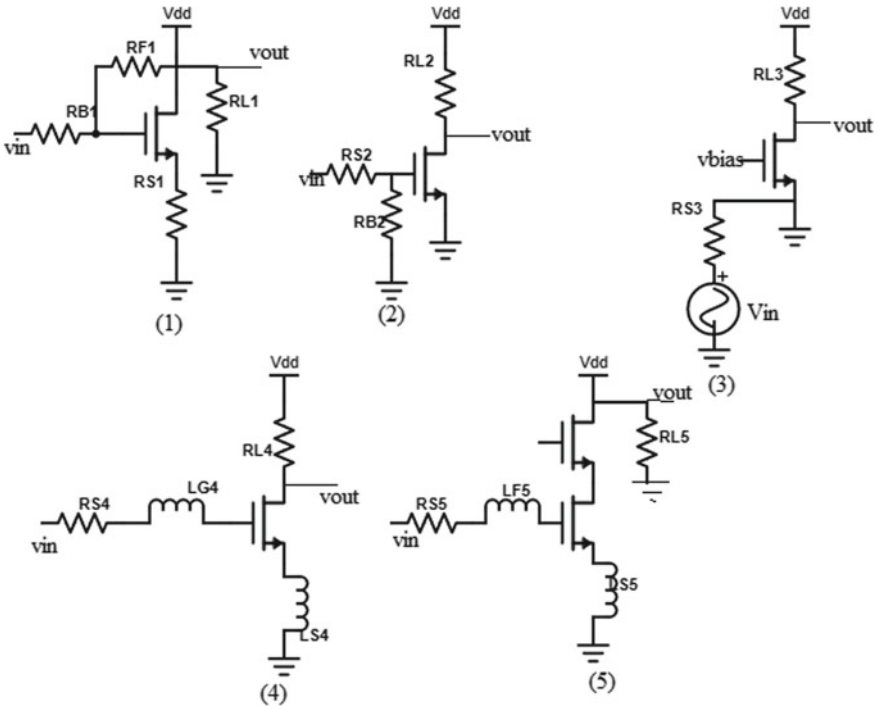


Fig. 1 Fundamental topologies of LNA: 1—Shunt series feedback common source; 2—resistive termination common source; 3—common gate; 4—inductive degeneration common source; 5—cascode inductor source degeneration

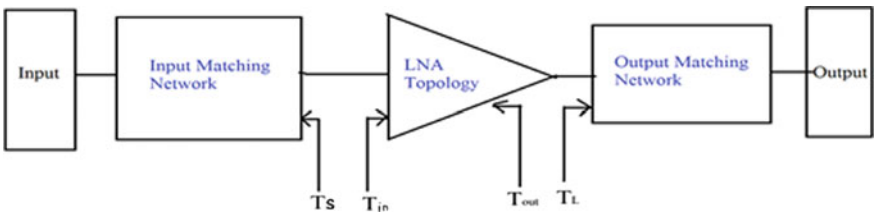


Fig. 2 A typical LNA design

the parasitic at the input can be absorbed into its structure, thus reducing their effect on the circuit performance. This comparison is summarized in Table 2. By studying [10] Table 2, CG LNA has far more advantages over IDCS LNA. If the problem of noise performance and effective transconductance value for CG can somehow be overcome or controlled, the CG LNA can prove to give far better performance than IDCS LNA.

Table 2 Advantages (+) and disadvantages (−) of LNA stages

Parameter		IDCS	CG
Effective g_m		+	−
Noise factor		+	−
DC power		−	+
Parasitic		−	+
Reverse isolation		−	+
Sensitivity		−	+
Input matching		−	+

In the literature [6], a combination of topologies has been used, CG for wideband input matching and second common source (CS) stage for its high gain to make up for the low gain of the first stage.

4 4. Study of Different LNA Designs

In [6],the CG amplifier is used as the first stage instead of CS amplifier because CG has lower quality factor than CS amplifier and thus can achieve input matching for wideband operation. The CG stage has to have lesser number of passive elements than CS stage in order to attain the same bandwidth. CG amplifier has a drawback of lower gain. Therefore, the CS stage is used as the second stage, which ultimately increases the overall gain of the LNA. Source degeneration of M_3 is used to tackle the trade-off factor between N.F & nonlinearity (Fig. 3).

FOM or figure of merit is a standard to compare various performance metrics of LNAs with same type of functions. As given in [6], FOM uses gain, noise figure, bandwidth, and power consumption to calculate its value:

$$\text{FOM} = \frac{(B \cdot W)\text{GHz} \times \text{Gain(dB)}}{(F-1)\text{dB} \times \text{PmW}} \tag{2}$$

Another very good example of a wideband LNA is discussed in [12], where a gain, a cascade of cascaded complementary CG (CCG), and CS topologies are used. CG has higher gain at lower frequencies and CS has higher gain at higher frequencies. The CCG technique helps in achieving smaller chip size, lower DC power and also achieves wider bandwidth and better linearity.

5 Application-based designs

LNA designed in the large frequency range 2–30 GHz can be used in multiple applications like:

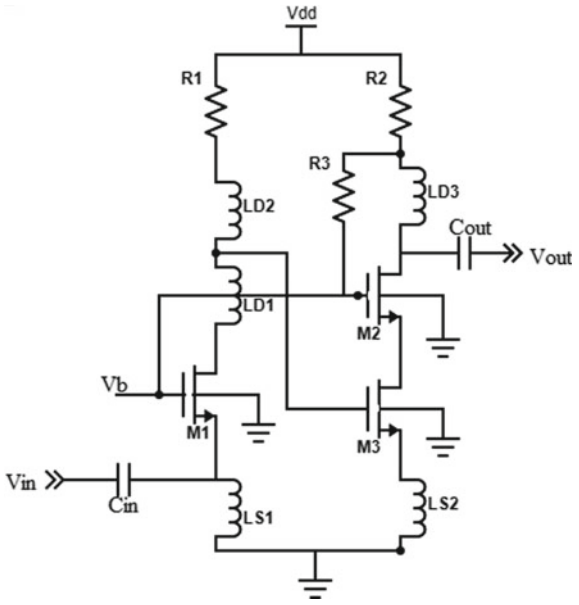
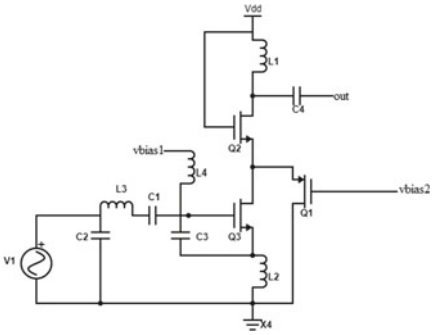


Fig. 3 Proposed LNA in [6, 11]

1. Bluetooth (2.402–2.48 GHz)
   An example for LNA designed for specifically Bluetooth application is shown in [13]. In this paper, an LNA with cascode topology is employed using 45 nm CMOS technology. The MOSFETs Q2 and Q3 form the core IDCS topology with Q2 being the input transistor and Q3 being the cascode transistor, which therefore increases the gain and improves isolation between the input and output. The output port is taken from the source of the cascode transistor. The combination of C1, C2, L3 forms the input impedence matching network. The combination of L2, L4, and Q3 is responsible for keeping the input impendence as close to 50 Ω as possible. The whole combination of C1, C2, L2, L3, L4 and transconductance (gm^3) and gate-to-drain capacitance (Cgs3) of Q3 should be such that the input impedence, Zin = 50 Ω, and the imaginary parts of Zin cancel each other out. Since the parasitic capacitance is in the order of femto farads, C3 is employed so that the inductor values do not reach unrealizable value and thus a sound design can be made. L1 is used for gain peaking at 2.45 GHz and transistor Q1 is used to improve nonlinearity by using intermodulation distortion (IMD) technique. By this technique, Q1 is said to be a folded PMOS IMD sink, whose biasing and width were chosen such that it cancels the third-order nonlinearities introduced by Q3.
   For Bluetooth application, narrowband performance is required and it is to be centered at 2.45 GHz. For narrowband application, FOM is given by

$$\text{FoM} = \frac{\text{Gain(dB)}Xf\text{(GHz)} \times \text{P1dB(dBm)}}{(N \cdot F - 1)\text{dB} \times \text{Pdc(mW)}} \tag{3}$$

Here, FOM gives more importance to the LNA power consumption, since BLE systems has a limitation of low power. Gain, f, P1 dB, N.F, and Pdc are the |S21|, operational frequency, noise figure, DC power consumption, and 1-dB compression point, respectively.



2. Ultra-wide band (UWB)
   According to the definition given by FCC, an UWB signal is any emitting signal that has a fractional bandwidth, whose value is greater than or equal to 0.20 or has a bandwidth greater than or equal to 500 MHz, unrelated to the fractional bandwidth. The operational frequency bands of UWB, according to the FCC, are below 960 MHz, 3.1–10.6 GHz, and 22–29 GHz. The allowed effective isotropic radiated power (EIRP) is less than −41.3 dBm/MHz. A LNA specifically designed for UWB band 3.1–10.6 GHz has been studied and presented in [5]. Here, in order to make reactive part of the input impedance resonate over the whole band, i.e., 3.–10.6 GHz, doubly terminated three-section passband Chebyshev filter structures are used. Thus, a good noise performance and also a wideband input match are achieved. Noise match is almost equal to power match for MOS devices [14].
3. ZigBee
   Ghadimipoor and Garakani [15] discuss a simple design of LNA for ZigBee applications working at 2.4 GHz. This is using IDCS topology with RC feedback for noise cancelation. Varactor, which uses a frequency tuning technology, that helps in dealing with the problem of frequency drifting due to process variation and unexpected parasitics, is used instead of capacitors in LC tank circuits placed at the input and output side. The varactor is implemented by using a MOS capacitor, that is being voltage-controlled. The drain source acts as the bottom plate of the varactor area and the gate acts as the upper plate of capacitor. The body of the MOS is connected to a controlling voltage through a resistor, which blocks the signal leakage.
4. WiMax
   Wang and Chen [16] discuss an LNA designed using core topology CG, for its excellent wideband input matching. But this topology comes with high noise contribution and low gain which are accomplished by two common source stages,

which act as a noise canceling and also gain enhancement technique. This is done by increasing the transconductance of the MOSFETs in the common source stages.

5. Satellite applications

Wang and Chen [17] have designed an LNA for satellite applications. The LNA must work in a wide dynamic range. So, for even performance, it must exhibit variable gain. This is achieved by two gain control techniques: (1) parallel current distribution technique and (2) adjustment of impedence at the output by altering the Q-factors of the parallel LC tank circuits. The first method has cascoded common source (CS), several parallel common gate (CG), and output impedence stages. Gain is controlled by changing the current flowing through each of the common gate stages. The latter method uses negative resistance modes to change the parasitic series resistance of inductance of load tank circuit. The wideband interference rejection is enhanced by decreasing the gain in current distribution mode and enhancing the gain in the negative resistance modes.

6 Comparison Table

Due to restriction in space, the comparison table is added at the end of the paper (Table 3).

7 Conclusion

This paper provides a summary of LNA design for wideband applications that have been discussed. Some examples have been provided, where similar work has been done inside the 2–30 GHz band, with some application-based approaches, whose frequency bands are fixed according to their individual operation. A comparative study have also been provided among the studied papers.

Table 3 Cross-platform comparative performance

Parameter	[6]	[5]	[13]	[17]	[16]	[15]	[12]
Technology	0.18 μm	0.18 μm	45 nm	0.18 μm	0.18 μm	0.18 um	65 nm
LNA topology used	CG + CS	IDCS	IDCS	CS + CG	CG + CS	IDCS	CG + CS
Power supply (V)	1	1	1	1.8	1.8	1.2	1
Bandwidth (GHz)	8.5–20	2.3–9.2	2.4–2.5	2.6	1–5	2–2.7	7.6–29
Gain (dB)	11.13	9.3	14.1	14.5–34.2	21–25	15.1	10.7
N.F (dB)	2.1–3.2	4	1.42	1.25	2.6–3	2.1	4.5–5.6
S11 (dB)	< −9.44	<−9.9	−18.4	< −18.9	< −7	< −10	
S12 (dB)	< −60	< −43	–	–	–	–	–
Area (mm^2)	0.116	–	–	–	–	0.66	0.3
IIP3 (dBm)	0.96	−6.7	18.6	4.52	–	−2	–
1-dB compression point (dBm)	–	−15	−9.87	–	–	–	–
Power consumption (mW)	5.4	9	1.98	23.85–28.17	14.5	12	12.1
FOM	15.1–28.4	–	4	–	–	–	–

Acknowledgements A piece of thanks goes to Prof. M. K. Mishra, Vice-chancellor, BIT Mesra Ranchi for providing us with infrastructure and facility to carry out the research work.

References

1. Friis HT (1944) Noise figures of radio receivers. Proc IRE 32(7):419–422
2. Heydari P (2007) Design and analysis of performance-optimized CMOS UWB distributed LNA. IEEE J. Solid-State Circuits 42(9):1892–1905
3. Zhang F, Kinget PR (2006) Low-power programmable gain CMOS distributed LNA. IEEE J. Solid State Circuits 41(6):1333–1343
4. Liu R, Lin C, Deng K et al (2003) A 0.5–14 GHz 10.6-dB CMOS cascade distributed amplifier. In: VLSI Circuits symposium on digest. Kyoto, Japan, June 2003, vol. 17, pp 139–140
5. Bevilacqua A, Niknejad A.M (2004) An ultra-wide band CMOS low-noise amplifier 3.1–10.6-GHz wireless receivers. IEEE J Solid-State Circuit 39(12):2258–2268
10. Pandey S, Gawande T, Inge S, Pathak A, Kondekar PN (2018) Design and analysis of wideband low-power LNA for improved RF performance with compact chip area. IET Microwav Ant Propagat 12(11):1816–1820
7. Kim CW, Kang MS, Anh PT et al (2005) An ultra-wideband CMOS low noise amplifier for 3–5 GHz UWB system. IEEE J. Solid State Circuits 40(2):544–547

8. Lin YS, Chen CZ, Yang HY et al (2010) Analysis and design of a CMOS UWB LNA with dual-RLC-branch wideband input matching network. IEEE Trans. Microw. Theory Tech. 57(2):287–296
9. Li Z, Wang Z, Zhang M et al A 2.4 GHz ultra-low-power current-reuse CG-LNA with active gm-boosting technique. IEEE Microw. Wirel. Compon
10. Idzdihar MI, Yusop N, Chachuli SAM, Ismail MM (2014) Design and analysis of low noise amplifier using cadence. J Theore Appl Inf Technol 69(1):151–160
11. Pandey S, Gawande T, Inge S, Pathak A, Kondekar PN (2018) Design and analysis of wide-band low-power LNA for improved RF performance with compact chip area. IET Microwaves Antennas Propag 12(11):1816–1820
12. Qin P, Xue Q (2017) Design of wideband LNA employing cascaded complimentary common gate and common source stages. IEEE Microwave Wirel Compon Lett 27(6):587–589
13. Sreekumar R, Mehdi N, Sotoudeh HH (2017) Cascode stage based LNA for bluetooth applications in 45 nm CMOS technology. In: 2017 new generation of CAS (NGCAS). IEEE, pp 145–148
14. Ballweber BM, Gupta R, Allstot DJ (2000) A fully integrated 0.5–5.5-GHz CMOS distributed amplifier. IEEE J. Solid-State Circuits 35:231–239
15. Lee CC, Shen Y, Wah CL, Faan HH, Kim FT (2016) ZigBee LNA design for wearable healthcare application. In: 2016 IEEE 14th international conference on industrial informatics (INDIN). IEEE, pp. 1134–1136
16. Ghadimipoor F, Garakani GH (2011) A noise-canceling CMOS low-noise Amplifier for WiMAX. In: 2011 international conference on electronic devices, systems and applications (ICEDSA). IEEE, pp 165–169
17. Wang J-J, Chen D-Y (2016) LNA with wide range of gain control and wideband interference rejection. Int J Electron 103(10):1748–1758

Ad Hoc Network Using UAVs in Indian Farms: A Review

Shivanta Sahoo, Yash Gupta, Vijay Nath, and Srikanta Pal

Abstract Agriculture is repeatedly affected by threats like wild animals, insects, droughts and floods. It is important to safeguard the farmland from various natural disasters and maintains the quality of crops to meet the consumer demand. For this purpose, the Government of India (GOI) has presented a number of rural plans all through the nation to assist the agriculture sector. The farmers need to monitor their crops to ensure the final product is undamaged and well grown. With the help of UAVs, the information related to the farmland could be gathered and sent to the farmer from time to time. But many of the Indian farms are located in deep remote areas with erratic geographical conditions where telecommunications or Internet services are difficult to maintain. So, the motivation of this research is to provide the farmers of our country with proper access to communication and monitoring farmland by using UAVs connected together by an ad hoc network. The ad hoc network will be able to maintain direct line of sight communication with uninterrupted connectivity in hilly areas and mountains. These connected UAVs can scan the crops and act as surveillance from intrusions.

Keywords Agriculture · Ad hoc network · Unmanned aerial vehicles (UAV) · Communication technologies

S. Sahoo (✉) · Y. Gupta · V. Nath · S. Pal
Wireless Communication, Department of ECE, B.I.T. Mesra, Ranchi, Jharkhand 835215, India
e-mail: shivantasahoo@gmail.com

Y. Gupta
e-mail: yashgupta142010@gmail.com

V. Nath
e-mail: vijaynath@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_65

1 Introduction

In India, agriculture accounts for only about 17% or so in our GDP, and about 13% in our exports, it is a very important sector. This is because our agricultural output has a major role to play both in our service sectors and also in industrial sectors by way of producing the raw materials and demands for industrial products. Over 50% of our workforce are employed directly in the agriculture sector and only when their income levels improve, the country as a whole can achieve mid-income status. As it is, based on the population and output, one can say that the average income level of people employed in agriculture in only one-fourth of the average earning of those employed in service/industry sectors. Here, we review a number of solutions and current developments under e-farming. The improvements under e-farming have given us many options to increase productivity by using sensors and networks. Here we suggest developing an ad hoc network using UAVs to establish a communication environment and use them for digital imaging and data collection at the time of natural calamities like floods, droughts, forest fires and landslides.

Saurabh Dwivedi et al. provided a method to educate and provide information to the farmers relating to different fields like market value of products, information on fertilizers, information of new farming techniques and various schemes provided by government to farmers on loans and capital investments [1].

Tamoghna Ojha et al. in their research article provided an introduction of available wireless sensor networks with a brief case study. The existing solutions are discussed, and then new methods are proposed with future developments. We also get an insight about how wireless sensor networks can be used in farming and its potential for agricultural purposes [2].

Takashi Furuta et al. provided a model for the farmland with agricultural support system for rice plantation whose field servers are mobile. The servers can work with a mobile battery, have a camera and can use free local networks. Servers are the cores of the computing network, and in a big farmland, the servers pose a hindrance to the farm work. So the movable field server provides a possible solution [3].

Jeongeun Kim et al. presented a review on hardware configurations of UAVs for agriculture use. Control systems for operating agricultural UAV were introduced, and applications of UAVs were classified. Discussion was made in-depth about the limitations and technology trends. A multi-robot system was suggested to provide solutions and improve the limitations of a traditional smart farming system [4].

Santiago González et al. gave a plan for monitoring architecture used in agriculture. The prototype used low cost hardware to deploy the wireless sensing network. Geolocation and videos of prescribed area are done with the use of GPS sensor and webcam. Practical demonstration was done to simulate the operation. A method was suggested to optimize the consumption by means of an energy-aware routing protocol, which will keep the network operational while extending the time of life on nodes with higher connectivity level and energy demand [5].

Pavan Sikka et al. presented a large heterogeneous sensor network having 60 nodes. This sensor network was powered by solar and could operate for over 6 months

including mobility. It had 2 types of sensors and 4 different classes of sensors. Farmland animals were tracked using GPS. The operated network was designed in a way to ensure 24/7 connectivity is maintained with the Internet via a dedicated high-gain radio link [6].

Evsen Yanmaz et al. provided a high-level architecture for the design and operation of UAVs and ground stations under drone networks. The network capabilities of UAVs were designed with the principles of the proposed architecture. Observations of the results were made for an effective improved design of drone networks, particularly for dynamic applications [7].

Paolo Tripicchio presented vision-based technique to analyze the characteristics of soil. This method was used in the plowing type discrimination. Two different algorithms were developed. Experimental tests were shown, and the methodology was able to provide a significant classification of the farmland's plowing depths [8].

Peerasak Serikul et al. did a research, meant to propose a cultivating application dependent on the Internet of things. The prototype of a smart capsule was created to estimate the humidity in paddy bags. Node MCU ESP8266 microcontroller and the SHT21 humidity sensor were utilized to send data to the Blynk server over a Wi-Fi network. The C+ + code for the microcontroller was written using Arduino IDE. The Blynk mobile application was utilized to monitor and display real-time humidity data on a digital screen. Humidity data were then collected and analyzed to develop a paddy crop storage system. The system developed monitored the humidity of paddy crop in order to avoid excessive humidity and maintained the crop quality [9].

Prusayon Nintanavongsa demonstrated how the movements of sensor affect the network communications within a farmland. The sensors are placed on UAVs. The farmland was inspected with these sensor-equipped UAVs. Through an experimental demonstration, the effect of overall network throughput by mobility of sensors was done as well as determination of optimal UAV mobility profile was made. The profile made yielded an optimal network throughput [10].

Sinung Suakanto et al. structured reasonable model and framework for decision support of smart farming with network sensor applications to help farmers with necessary tasks. A theoretical model included six parts responsible for management, planning, monitoring, information distribution, decision support, and control action. A comprehensive model using Internet of things (IoT) approach was proposed which was applied to farmland. This model was given to provide solution to farmers dealing with the problems of task management and planning, environment factors measurements, and information distribution [11].

Anton Louise et al. made an approach for multi-depot, fuel constrained coverage path planning. Target coverage was fragmented into smaller areas on the basis of number of available charging depots. Furthermore, each region then divided into number of cells with area equivalent to the camera field of view when UAV was flying above the farmland at a height of 3 m. A number of possible routes were generated and then fed into optimization scheme to find out the optimal path keeping in consideration the fuel levels and the availability of recharging depots in the farmland. The optimization yielded a considerable improvement in obtaining the route that would give the shortest distance that the UAV should follow to cover the given

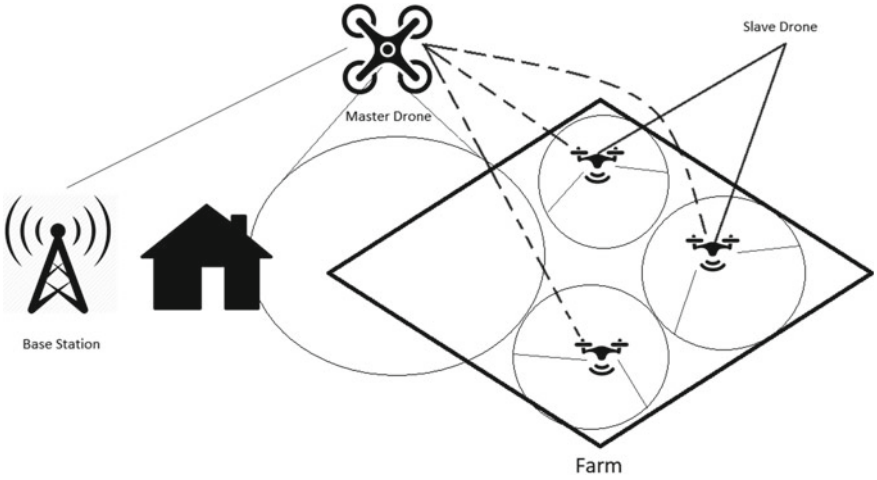


Fig. 1 Proposed UAV network

area of interest within the farmland. This approach was proved to be useful to increase the efficiency of UAVs for farmland monitoring [12].

Philipp Lottes et al. addressed the problem of detecting crops like sugar beets as well as the typical weeds using a sensor camera placed on a UAV with lightweight equipment's. The proposed system performs crop recognition, plant-tailored feature extraction, and classification to get an estimate of the distribution of sugar beets and weeds in the field. The proposed system also tells about crop recognition, feature extraction, and classification for aerial images depending on object-features and key points in combination with a random forest. The monitored data was captured on two farmlands, and there was an illustration made to show the approach allowing for identification of the crops and weeds in the field [13].

2 Methodology

Figure 1 presents a simple method where we have a base station, backbone drone, and slave drones being connected together to form a backhaul network.

3 Result and Discussion

Although there are a wide variety of UAVs available in market, their use is limited due to high cost. Traditionally farmers look for cost-effective solution for monitoring the farms. Here we present the idea of using ad hoc network created by mini UAVs can

extend the network coverage area and also can establish a temporary communication network.

4 Conclusion

A number of solutions and current developments under e-farming were reviewed and the proposed network system having UAVs to construct an ad hoc network is provided to help the farmers in the time of any natural calamity and also to provide basic surveillance via wireless sensing network.

Acknowledgements We are thankful to our Vice-Chancellor, Birla Institute of Technology, Mesra, Ranchi, for providing infrastructure facility to carry out this research work. We also grateful to our HOD, Department of ECE and other staff members of BIT Mesra for their motivation and support.

References

1. Dwivedi S, Parshav V, Sharma N, Kumar P, Chhabra S, Goudar RH (2013) Using technology to make farming easier and better: simplified e-farming support (SEFS)
2. Ojha T, Misra S, Raghuwanshi NS (2015) Wireless sensor networks for agriculture: the state-of-the-art in practice and future challenges. Comput Electron Agric 118:66–84
3. Furuta T, Nakagawa H, Kitamura T, Kawakami Y, Kurosawa K, Kogami K, Tanaka MS (2016) Agricultural support system equipped with short-range wireless communication system. In: 2016 IEEE technological innovations in ICT for agriculture and rural development (TIAR). IEEE, pp 58–64
4. Kim J, Kim S, Ju C, Son HI (2019) Unmanned aerial vehicles in agriculture: A review of perspective of platform, control, and applications. Special section on new technologies for smart farming 4.0: Research challenges and opportunities.https://doi.org/10.1109/ACCESS.2019.2932119
5. González S, Vargas TR, Arce P, Guerri JC (2016) Energy optimization for video monitoring system in agricultural areas using single board computer nodes and wireless ad hoc networks. In: 2016 XXI symposium on signal processing, images and artificial vision (STSIVA). IEEE, pp 1–7
6. Sikka P, Corke P, Valencia P, Crossman C, Swain D, Bishop-Hurley G (2006) Wireless adhoc sensor and actuator networks on the farm
7. Yanmaz E, Yahyanejad S, Rinner B, Hellwagner H, Bettstetter C (2018) Drone networks: communications, coordination, and sensing
8. Tripicchio P, Satler M, Dabisias G, Ruffaldi E, Avizzano CA (2015) Towards smart farming and sustainable agriculture with drones. In: 2015 international conference on intelligent environments
9. Serikul P, Nakpong N, Nakjuatong N (2018) Smart farm monitoring via the Blynk IoT platform: case study: humidity monitoring and data recording. In: 2018 16th international conference on ICT and knowledge engineering
10. Nintanavongsa P, Pitimon I (2017) Impact of sensor mobility on UAV-based smart farm communications. In: 2017 international electrical engineering congress, Pattaya, Thailand, 8–10 March 2017

11. Suakanto S, Engel VJ, Hutagalung M, Angela D (2016) Sensor networks data acquisition and task management for decision support of smart farming. In: 2016 international conference on information technology systems and innovation (ICITSI). IEEE, pp 1–5
12. De Ocampo ALP, Bandala AA, Dadios EP (2018) Coverage path planning on multi-depot, fuel constraint UAV missions for smart farm monitoring. In: 2018 IEEE region ten symposium (Tensymp)
13. Lottes P, Khanna R, Pfeifer J, Siegwart R, Stachniss C (2017) UAV-based crop and weed classification for smart farming. In: 2017 IEEE international conference on robotics and automation (ICRA). IEEE, pp 3024–3031

Web Application Based on PHP for Farmers

Jayanth Siddarth Ponugoti, Satya Sai Teja Karuturi, and Vijay Nath

Abstract This paper demonstrates the Web application called *Kisan Bank* designed especially for the farmers to aid them to have the connection directly with the government for exchange of their stock without interference of any intermediate parties. The application was designed using PHP [1–3] and HTML [4] for secure and clean Web site. For having better user interface JavaScript [5] and CSS [6] are used. The Web site also uses the MySQL [7] database, jQuery [8] and Google API [9].

Keywords Web development · Web application · Web site · HTML · PHP · CSS · JavaScript · MySQL · Google API · jQuery · Farmer · Government · Agriculture

1 Introduction

India is mainly an agricultural country. Almost 70% of the total population lives in the rural areas. The main occupation in these areas is agriculture.

According to an Indian survey at least 60–70% of people are dependent on agriculture for their employment either directly or indirectly. It is an important source of economy and contributes almost 16–17% of nation GDP.

But in recent years farmers are leaving their occupation and migrating to urban areas for employment due to heavy losses incurred by them doing farming. This is caused due to the lack of proper connection between the government and the farmers.

There is lot of communication gap between the government and farmers, and policies designed by the government are not reaching to them. Due to this, they do not know what the actual price that they can sell their crops. This situation is utilized by the intermediate groups (individual/firms), and they buy crops at a low price and sell them at high price.

J. S. Ponugoti (✉) · S. S. T. Karuturi · V. Nath
Embedded Systems Group, Department of ECE, B.I.T. Mesra, Ranchi, Jharkhand 835215, India
e-mail: pjayanth99@gmail.com

S. S. T. Karuturi
e-mail: saitejakaruturi0405@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_66

This condition can be removed by making direct communication between the farmers and government and cutting the involvement of intermediate parties. It means government directly buy crops from the farmers by giving them the present market price for the crop.

Our Web site deals with the above scenario which helps the farmers to sell the crops directly to government and allow the government to pay directly to farmer and acknowledge to him. The paper flow consists of the following pattern, the Web site design, approach and elements were discussed in Sect. 2, the application interfaces and language are used and their purpose were shown in Sect. 3, the final nature and interface of Web site are shown in results in Sect. 4, and conclusion is in Sect. 5 followed by references.

2 Design Approach

The basic idea of the Web site is to create a platform between farmers and the government to have direct connection between them without any interference from third party. So, it is designed to have a database that stores all the information sent by the farmer about his crops and the status of it from the government site.

It contains two login sites: one for the government and the other for the farmer. The farmer site consists of login and register page while the government site contains only login page.

The farmer can register himself by providing his Aadhar image, bank details, and IFSC code. After registration, the farmer can login to account and can send request to the government by sending information about crop. The request is sent to the government site.

The government verifies the details of the farmer and can acknowledge his request which was sent to him by the government site to farmer site.

The elements of the Web site.

- Home page
- Government login
- Farmer login
- Farmer registration.

Fields of farmer registration.

- Full name
- User name
- Password
- Phone number
- Account number
- IFSC code.

Farmer site attributes.

- Total requests
- Accepted requests
- Pending requests
- Rejected requests.

Government site attributes.

- Verifications pending.

3 Programming and Interfacing

Many programming languages and interfaces are used to develop the Web site. They are.

- PHP
- HTML
- JavaScript
- CSS
- Google Application Programmer Interface (Google API)
- MySQL
- jQuery.

PHP: Hypertext Preprocessor which is a general-purpose scripting language especially suited for Web development and can be embedded into HTML. It is used for backend in the Web site.

HTML: Hypertext Markup Language is used to display designed documents in the browser.

CSS: Cascading Style Sheets is used to describe presentation of a document written in HTML.

JavaScript: It is used as a programming language for the database used in the Web site.

Google API: It is used for the purpose of pie charts in the Web site.

MySQL: It is a structured query language used for the purpose of data base.jQuery: It is the JavaScript library used to have dynamic behavior in static Web pages.

4 Results

See Figs. 1, 2, 3, 4, 5, and 6.

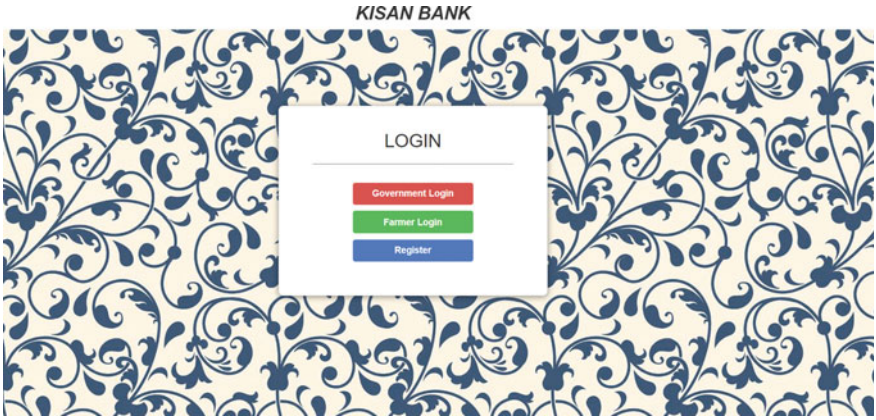


Fig. 1 Home page [3–6]

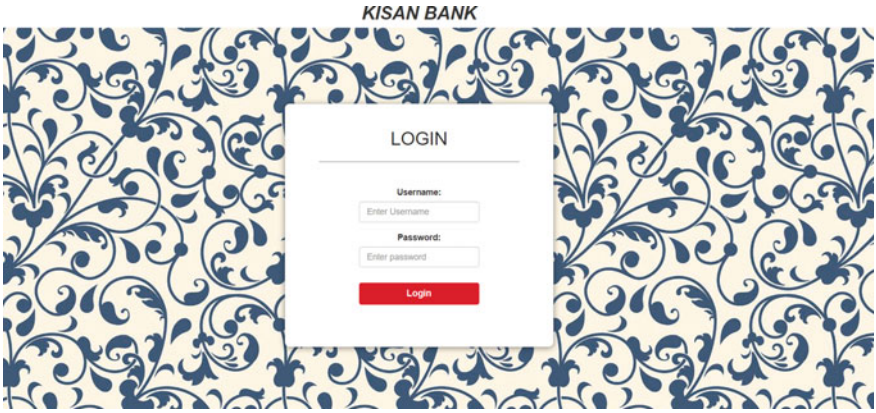


Fig. 2 Login page [3–6]

5 Conclusion

By this Web site, we can help the farmers to bridge the gap between them and the government. This also helps in the farming which contributes mostly to the national welfare. This project consists a lot of programming languages which have lot of interactive pages which most of the programmers are using. Hence, there is a lot of scope for the development.

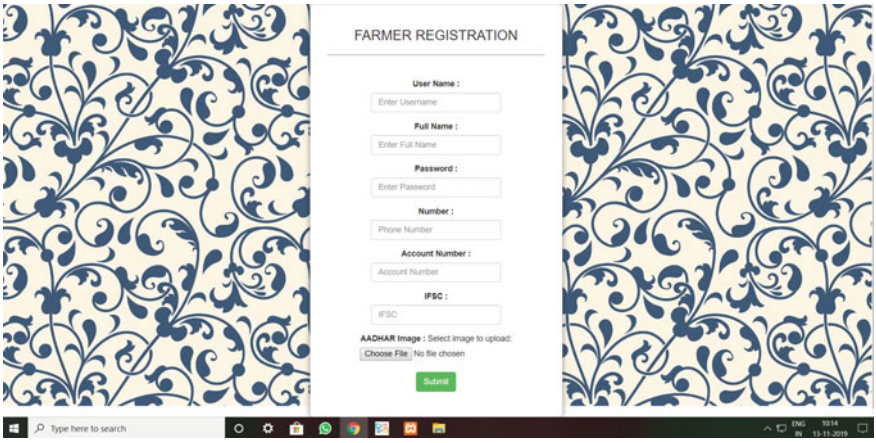


Fig. 3 Farmer registration page [3–6]

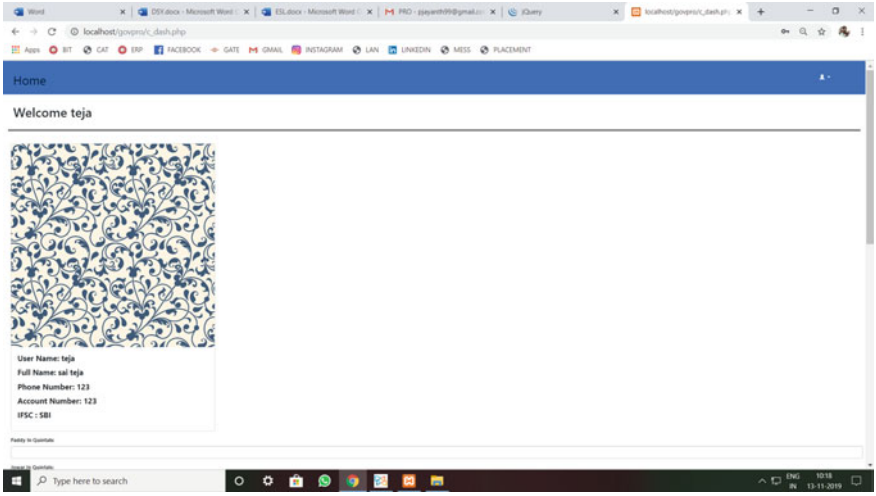


Fig. 4 Farmer site [3, 7, 8]

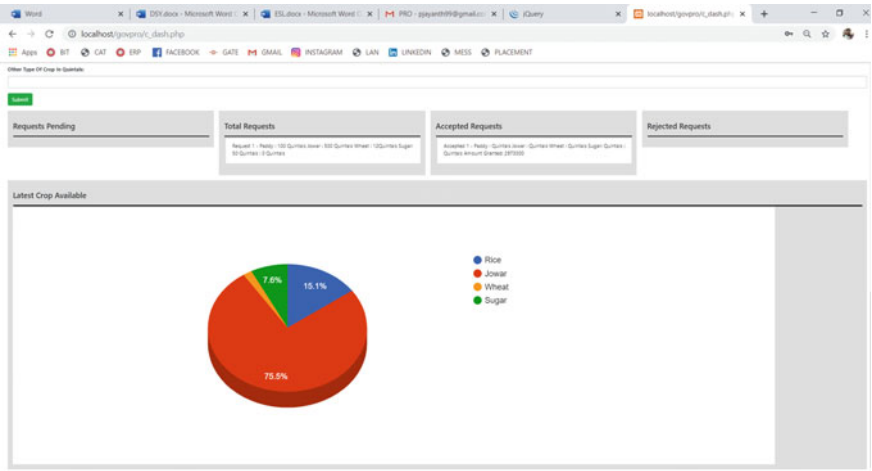


Fig. 5 Farmer attributes [3, 7–9]

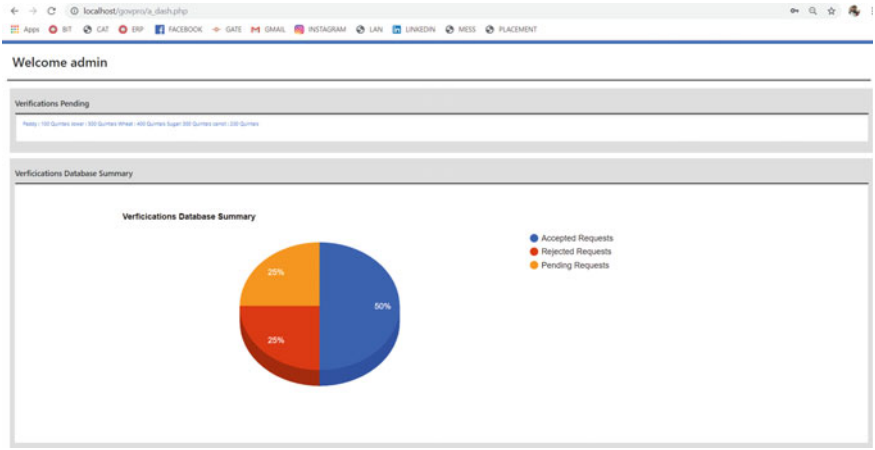


Fig. 6 Government attributes [3, 7–9]

References

1. https://www.geeksforgeeks.org/
2. https://www.tutorialspoint.com/
3. https://en.wikipedia.org/wiki/PHP
4. https://en.wikipedia.org/wiki/HTML
5. https://en.wikipedia.org/wiki/JavaScript
6. https://en.wikipedia.org/wiki/Cascading_Style_Sheets
7. https://en.wikipedia.org/wiki/MySQL
8. https://www.w3schools.com/jquery/
9. https://en.wikipedia.org/wiki/Google_APIs

Prediction and Classification of Semi-structured Medical Data Using Artificial Neural Fuzzy Inference System

Puneet Singh Duggal and I. Mukherjee

Abstract Soft computing and machine learning techniques have been used in the medical domain for prediction and classification of diseases. In this paper, a programmed determination system based on artificial neural network (ANN) and adaptive network based on fuzzy inference system (ANFIS) for male richness is presented. The model consolidated the adaptive capacities of neural network and subjective methodology of fuzzy rationale. A few conclusions concerning the saliency of highlights on classification of the ripeness dataset were gotten through examination of the ANFIS. The presentation of the ANFIS model was assessed as far as preparing execution and classification exactness's and the outcomes confirmed that the proposed ANFIS model has potential in grouping the fruitfulness information. The right conclusion execution of the ANFIS-based programmed finding system for clinical prediction is assessed by utilizing classification exactness, affectability and explicitness examination, individually. The classification precision of this ANFIS-based programmed finding system for the analysis of male richness was gotten in about 93.16%.

Keywords Fertility data · Adaptive network based on fuzzy inference system (ANFIS) · Data analytics

1 Introduction

The fertility rates have a striking decrease in last two decades [1–3]. It has been seen that this decrease is because of changes in conduct identified with financial perspectives, fuse of ladies in the process of childbirth and the following deferral in the age at which an individual chooses to have posterity, and the inescapable utilization

P. S. Duggal (✉) · I. Mukherjee
Department of CSE, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: duggal@gmail.com

I. Mukherjee
e-mail: imukherjee@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_67

of contraceptives [4, 5]. Despite the fact that obviously the social viewpoints have been contributing essentially to worldwide decrease in ripeness rate with the decay of regenerative well-being brought about by unfriendly natural factors [6]. In the previous decades, Authors in [7] played out a meta-examination on the chance of decrease in fundamental quality. Few examinations show a reduction in semen boundaries of men [8, 9]. Computerized reasoning has become a significant instrument in medical data analytics. A significant number of these applications have progressed to the appearance of expert systems and decision support systems in a few unique regions. To distinguish the hypertension patients and locate the vital boundaries for the expectation whether an individual is influenced or not, authors in [10] have proposed another model dependent on the affiliation rule and neural system to analyze the bosom malignancy sicknesses. To assess the exhibition of above said model, three overlap techniques are utilized and contrasted it with neural system model in which affiliation rule in addition to neural system model gives better grouping rate. Authors in [11] have created two multilayer neural system models for finding tuberculosis malady in that, one model has utilized a solitary shrouded layer while extra model has utilized two concealed layers and the presentation of these models is estimated on precision boundary, wherein multilayer neural network with two shrouded layers gives preferred exactness over different models. Kumar [12] has proposed a standard base characterization model for anticipating various kinds of liver illnesses. The standard base-grouping model is mix of rules and various information mining procedures in which rules have determined utilizing standard test outcome esteems, and information mining strategies are utilized to survey presentation of model. It is discovered that choice tree strategy gives better outcomes among rule acceptance, bolster vector machine, counterfeit neural system, and credulous Bayes. Author in [13] has applied choice multilayer perceptron, trees, and bolster vector machines techniques to assess their presentation in the forecast of the fundamental quality from information of the natural components and way of life. Outcomes reason that multilayer perceptron and bolster vector machine show the most noteworthy precision with forecast exactness estimations of 86% for a portion of the fundamental boundaries. This paper manages the investigation of male richness rate and furthermore applies the computerized reasoning methods to anticipate the male fruitfulness rate and discover natural elements and life propensities boundaries, which influence the semen quality.

2 Semi-structured Data

In data analysis, different types of data pose a roadblock as these cannot be processed by a single method or tool. Apart from this, data can be a mix of the above-mentioned types. Also, data can be linier, high dimensional, or heterogeneous in nature.

Semi-structured data has non-sensitive core component data, flexible content, and individual data (Fig. 1). Its usage are content item-extending data and use case data [14].

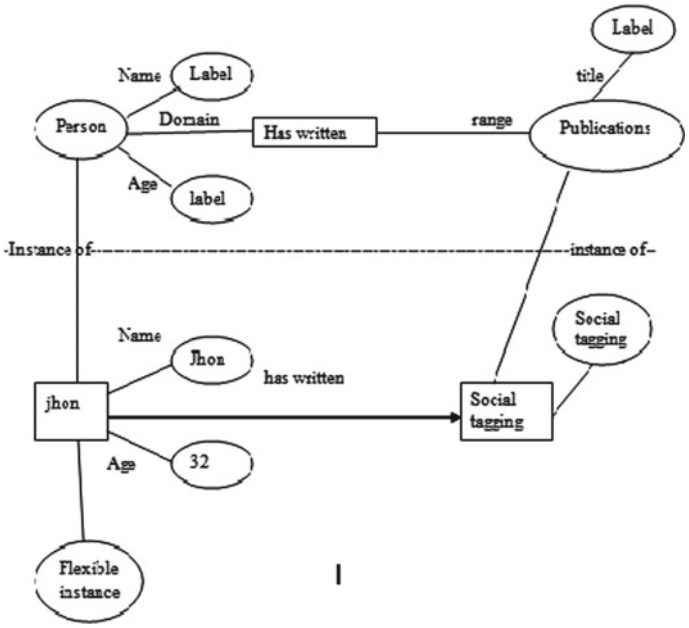


Fig. 1 Semi-structured data instances

Apart from semi-structured data, traditional structured and new age unstructured data are there for researchers for analysis. Structured data like relational data have been worked upon in the data analytics to get a meaningful output. With the advent of Internet, small devices, and sensors, conventional meaning of data has changed, unstructured data which do not follow the schema and structures have posted a challenge to the researchers. These are briefly explained in the comparative Table 1.

3 ANFIS and R

R supports the implementation of different layers of AFIS architecture [15, 16] that implements the adaptive network-based fuzzy inference system. Different phases/layers of the architecture are coded into the R IDE. The inbuilt robust architecture or R makes it easy for analysis of the data eve if it is semi-structured. It is used to solve regression, prediction, and classification tasks.

Table 1 Comparison of unstructured, semi-structured, and structured data

	Unstructured	Fully structured	Semi-structured
Technology	Character and binary data	Relational database tables	XML/RDF
Transaction management	No Transaction management no concurrency	Matured transaction management, various concurrency techniques	Transaction management adapted from RDBMS, not matured
Version management	Versioned as a whole	Versioning over tuples, rows, tables, etc.	Not very common, versioning over triples or graphs is possible
Flexibility	Very flexible absence of schema	Schema-dependent, rigorous schema	Flexible, tolerant schema
Scalability	Very scalable	Scaling DB schema is difficult	Schema scaling is simple
Robustness	–	Very robust, enhancements since 30 years	New technology, not widely spread
Query Performance	Only textual queries possible allows	Structured query allows complex joins	Queries over anonymous nodes are possible

4 Experimental Setup

The experiments were performed on a Pentium Core i3 with virtualization enabled 1.8 GHz CPU with 8 GB RAM-based machine. The characteristic of the data is shown in Table 2.

Table 2 Characteristics of data

Name of attribute	Attribute role	Attribute type	Attribute statistics	Attribute range	Missing values
Diagnosis	Prediction	Binominal	mode = N (88) least = A (12)	(88), A (12)	0
Season	Regular	Nominal	Avg. = -0.072 ± 0.797	[−1.00; 1.00]	0
Age	Regular	Real	avg. = 0.669 ± 0.121	[0.500; 1.000]	0
Childish diseases	Regular	Nominal	Avg. = 0.870 ± 0.338	[0.000; 1.000]	0
Accident or serious trauma	Regular	Nominal	Avg. = 0.440 ± 0.499	[0.000; 1.000]	0
Surgical intervention	Regular	Nominal	Avg. = 0.510 ± 0.502	[0.000; 1.000]	0
High fevers	Regular	Nominal	Avg. = 0.190 ± 0.581	[−1.00; 1.00]	0
Alcohol consumption	Regular	Nominal	Avg. = 0.832 ± 0.168	[0.200; 1.000]	0
Smoking habit	Regular	Nominal	Avg. = -0.350 ± 0.809	[−1.00; 1.00]	0
Number of hours spent sitting	Regular	Nominal	Avg. = 0.407 ± 0.186	[0.060; 1.000]	0

*A. **Process Steps for ANFIS** in performing Data Analysis in R*

Step 1 input data
Step 2 Separate Training Data, Testing Data and Fit Data
Step 3 Define Training Class
Step 4 Define Data Range
Step 5 Apply ANFIS
    Step 5.1 Set Rules
    Step 5.2 Set Number of iterations
    Step 5.3 Set step size
    Step 5.4 Set the type of Membership Function used

Step 6 Configure Control
Step 7 Generate Fuzzy Model
    Step 7.1 Train Data (using ANFIS model & Control applied)
Step 8 Prediction using test data
Step 9 Fitting the model using data training
Step 10 Calculate Error
    Step 10.1 Calculate Mean Square Error
    Step 10.1 Calculate Root Mean Square Error
Step 11 Compare Error (MSE, RMSE)
Step 12 Compare between simulation and real data

5 Results and Discussion

From the plot of fitting phase (Fig. 2), the training of the data using AFIS is perfectly accurate. This means that the initial two steps of the ANFIS which are training and

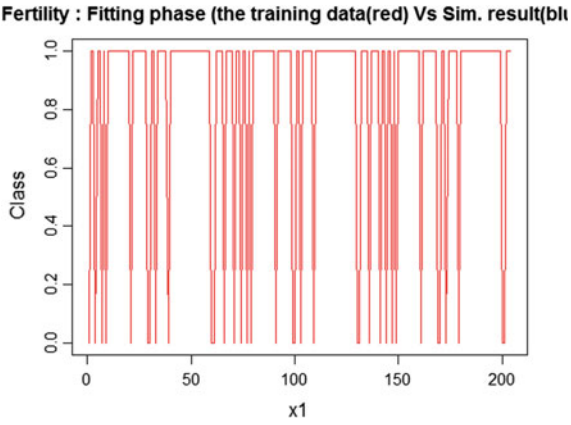


Fig. 2 Fitting phase of training

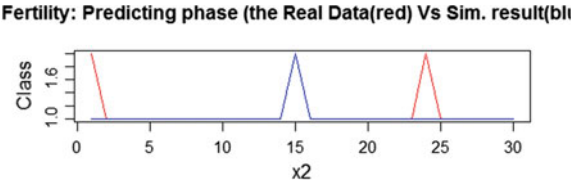


Fig. 3 Comparison of the real data versus the sim results

applying fuzzy rules are fitting to the curve of input data and this will ensure that the prediction phase will give accurate results.

A. Results from the R R Engiene

```
print("FRBCS.W: percentage Error on Fertility")
[1] "FRBCS.W: percentage Error on Fertility"
print(err.frbcs)
[1] 13.33334
print("tree: percentage Error on Fertility")
[1] "tree: percentage Error on Fertility"
print(err.tree)
[1] 66.6666
```

It is seen from the prediction plot (Fig. 3) that the accuracy for the prediction is 66.666/80 *100, i.e. **83.33%,** the prediction error comes out to be 16%.

6 Conclusion and Future Discussion

R has came out to be a powerful tool in the data analytics domain. With the coming of big data problems, different categorization of data on the basis of variety of data to be analyzed has pulled up a different kind of challenge. With the coming of unstructured and semi-structured data, traditional data mining and statistical methods have a handicap in dealing with them. Storing and working on unstructured variety needs a combination of existing and new technologies. R with Hadoop is a relatively new way that will enable researchers and programmers to manipulate Hadoop data which are stored in data centers on a distributed array of HDFS and HBASE.

The future scope of this work will be an extension of the above work using the R + HDFS combination on fully unstructured data. Authors are sure that this methodology will provide new insights in the already solved problems and will give more accurate solutions.

References

1. Inhorn MC (2003) Global infertility and the globalization of new reproductive technologies: illustrations from Egypt. Soc Sci Med 56:1837–1851
2. Lutz W, O'Neill BC, Scherbov S (2003) Demographics. Europe's population 295 at a turning point.Science 299:1991–1992
3. Grant J, Hoorens S, Sivadasan S, Loo MV, Davanzo J, Hale L, Butz W (2006) Trends in European fertility: should Europe try to increase its fertility rate … or just manage the consequences? Int JAndrol 29:17–24
4. Skouby SO (2004) Contraceptive use and behavior in the 21st century: a comprehensive studyacross five European countries.Eur J Contracept Reprod Health Care 9:57–68
5. Cibula D (2008) Women's contraceptive practices and sexual behaviour in Europe. Eur J Contracept Reprod Health Care 13:362–375
6. Skakkebaek NE, Jorgensen N, Main KM, Rajpert-De Meyts E, Leffers H, Andersson AM, Juul A, Carlsen E, Mortensen GK, Jensen TK, Toppari J (2006) Is human fecundity declining? Int J Androl 29:2–11
7. Carlsen E, Giwercman A, Keiding N, Skakkebaek NE (1992) Evidence for decreasing quality of semen during past 50 years.BMJ 305:609–613
8. Auger J, Kunstmann JM, Czyglik F, Jouannet P (1995) Decline in semen quality among fertile men in Paris during the past 20 years. N Engl J Med 332:281–285
9. Splingart C, Frapsauce C, Veau S, Barthelemy C, Royere D, Guerif F (2011) Semen variation in a population of fertile donors: evaluation in a French centre over a 34-year period. Int J Androl 35:467–474
10. Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. Expert Syst Appl 36(2):3465–3469
11. Er O, Temurtas F, Tanrıkulu AC (2010) Tuberculosis disease diagnosis using artificial neural networks. J Med Syst 34(3):299–302
12. Kumar Y, Sahoo G (2013) Prediction of different types of liver diseases using rule based classification model. Technol Health Care 21(6):417–432
13. David G, Girela JL, Juan JD, Jose Gomez-Torres M, Johnsson M (2012) Predicting seminal quality with artificial intelligence methods. Expert Syst Appl 39(16):12564–12573
14. Manola F, Miller E (2015) Resource description framework (RDF): concepts and abstract syntax. https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/. Accessed 25 Dec 2015
15. R FRBS package documentation, https://www.inside-r.org/packages/cran/frbs/docs/
16. https://www.ukessays.co.uk/essays/computer-sciences/the-fuzzy-if-then-rules.php

K-NN Classification of Very Large Dataset in Spark Framework

Ritesh Jha, Vandana Bhattacharjee, and Abhijit Mustafi

Abstract The amount of data is increasing every day through the sources of social media, IoT devices, day-to-day bank transaction, and many other sources. To handle such kind of data, it is typical task and it comes in big data domain. Big data has characterized 5-Vs known as velocity, volume, value, veracity, and last one veracity. The use of machine learning techniques to handle this complexity is becoming a very sought after research area. Classification is another challenging task in big data, and this is the problem that we are aiming to solve using K-NN approach under spark framework. We compared the result of K-NN with hadoop and spark with and without (HPC) and found that training and testing time drastically decrease in spark as compared to hadoop and accuracy also has been increased.

Keywords Big data · K-NN · HPC · Spark

1 Introduction

With the emergence of IoT, information technology in various fields leads to a creation of very large data called the big data which has substantial challenges to store and process or analyze the data [1]. There are various energy-efficient routing protocol which have been proposed to collect the data. The data is growing day-by-day, and the format of the data is also not fixed the format of big data generally be structured or unstructured like records; images may be structured or unstructured [2]. To analyze the big data, we need some machine learning techniques; but this

R. Jha (✉) · V. Bhattacharjee · A. Mustafi
Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: riteshjha@bitmesra.ac.in

V. Bhattacharjee
e-mail: vbhattacharya@bitmesra.ac.in

A. Mustafi
e-mail: abhijit@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_68

machine learning techniques not easily be applied on the big data as data is stored in distributed environment [3]. For storing, the big data two popular platforms most widely used are Apache Hadoop and spark [4]. Data mining plays an important role in analyzing the data.

The main objective of any data mining or machine learning algorithm in the domain of big data is to find the hidden information or pattern. Two most widely used techniques for this are classification and clustering. In both methods problem arises when data is imbalanced and multi-class. In this paper, K-NN algorithm has been applied for multiclass classification. Further, the comparative study of results has been presented.

The organization of paper is as follows: Section 2 presents the related work; Section 3 explains the methodology including the spark frameworks; Section 4 describes the experimental setup with HPC and algorithm. Section 5 discusses the results, and finally, Section 6 presents the concluding remarks for the paper.

2 Related Work

To deal with the medical imaging data, the k-NN has been applied in the paper. In this paper, the authors had been applied both clustering and classification techniques. Firstly, the k-means technique divided the whole datasets into different segments, and then K-NN had applied on each segments for classification. The results showed that it improves the accuracy and efficiency.

Another work is done in the area of traffic flow prediction. In this paper, the authors had applied various algorithms like naive Bayes, multilayer perceptron model, NN regression to predict the traffic flow.

Accuracy had been found to be best in K-NN techniques.

For the varying number of maps and number of reducers for runtime [5]. The authors compares the result in terms of speed up concepts by applying the K-NN technique. The results showed the run-time decreases as the number of mappers and reducers has been varied.

For finding the duplicate or nearby objects in images, the work is done in this area by using the nearest neighbors approach [6]. The authors had applied the NN technique to find the duplicates in billions of images which are most widely used in domains of OCR and face recognition.

To reduce the cost of K-NN search, the work is done in this domain in [7], where authors had used the LSH technique. In this technique, the same kind of objects was put in the same bucket. The benefit of this approach reduces the set of the K-NN algorithm. The result had shown that this technique is accurate and efficient for high-dimensional big data.

3 Methodology

This section describes the methodology, and the framework is used in this work (Fig. 1).

A. Spark Framework

The driver program assigns the job across all workers. Each worker has task and executor to execute the task or job. Each worker works simultaneously to execute the job, and final output from different workers is collected by driver program (Fig. 2).

B. Existing k-NN Sequential Procedure

To be more exact, all (or most) of the training data are needed during the testing phase.

k-NN has two phases:
Training Phase
Testing Phase
Training is done with features along with class labels.
Testing requires data without class label.

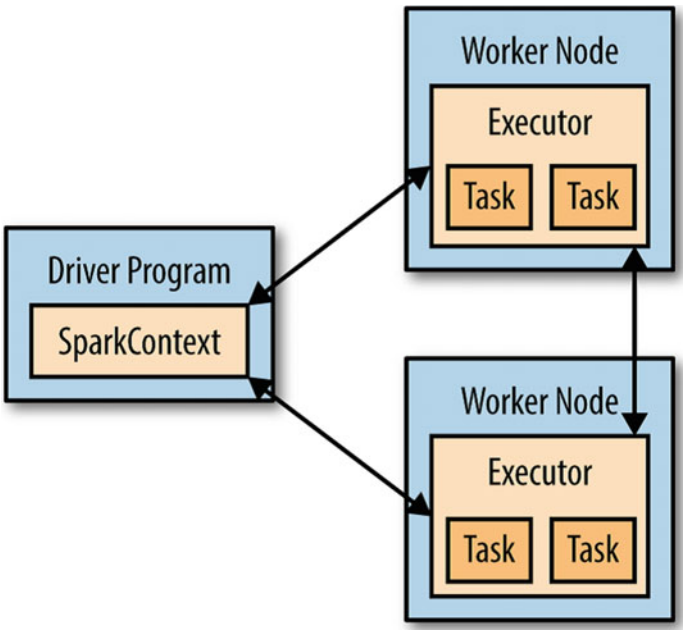


Fig. 1 Apache framework [12]

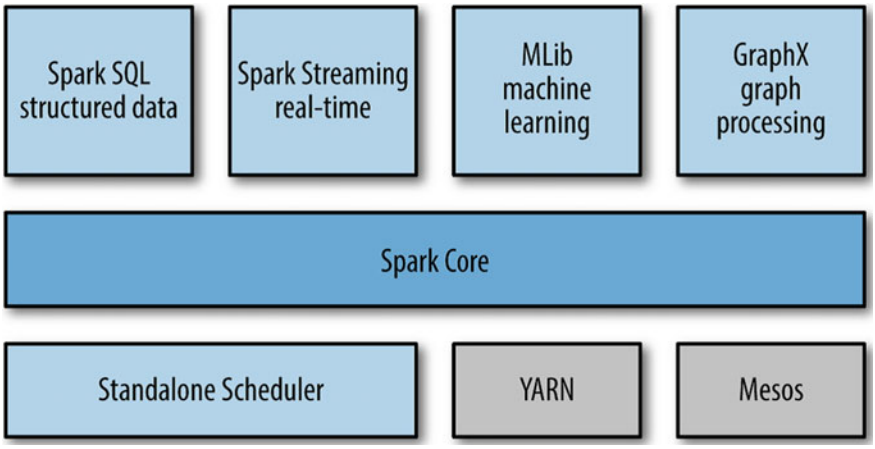


Fig. 2 Spark stack [4]

Algorithm : 1(Sequential Approach)
Input: Data, k

1) Load the Dataset
2) Split the data into (training (70%) and testing (30%))
3) Calculate the Euclidian Distance of the Test Data from each instance the Training Data.
4) Perform majority vote amongst the "k" nearest neighbors.
5) Assign the class label of the majority vote winner to the new class label from the testing Dataset.

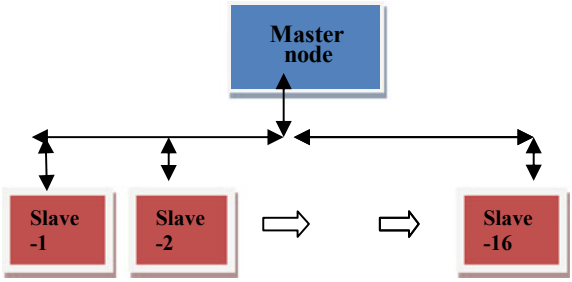
Repeat until all for each instance of the Testing Data is classified

Here, the data is initially loaded into spark RDD [17]. We split the data into fraction 0.7 (train) and 0.3 (test). The K-NN calculates the distance(Euclidian) from each training data instance to each test data in each partition, and then it performs the majority of votes for class label for given *k*.

Algorithm : 2
Input : Data set, K(no of nearest neighbors)

1. 1. Load the data into RDD.
2. Split the data into 70% (train) and 30%(test)
3. Calculate the Euclidian distance in each partition along with class label.
4. Broadcast the test data in each partition.
5. Predict the class label of test data in each partition.
6. Collect the result at master.
7. Calculate the majority of class label as value of k.
8. Evaluate the Accuracy

Fig. 3 HPC configuration



4 Experimental Setup

The experiment has been conducted on high-performance computing server (HPC), which has two master nodes and 16-slaves nodes.

In spark framework, one master node is the main node and it communicates with other slaves node which is shown in Fig. 3. The responsibility of the master node is to distribute the data among its slaves. Once the data has been distributed, each slave node has the executor which works on the data. Once the processing of the data is over, master node collects the result from all slave nodes. Apart from the distribution of data, master node handles the slave node failure efficiently. In case if any node failed, the data is transferred to nearby slave node. The difference between hadoop and spark is as follows:

1. The hadoop read and write the data in disk manner.
2. Spark read and write the data in RAM, that is, called the in-memory data retrieval.
3. Hadoop is best suitable for batch processing.
4. Spark is used for both batch and real-time streaming processing.

5 Results

Poker dataset was used for the purpose of this experiment. The dataset contained 1,000,000 instances in the dataset (size: 23 MB). The K-NN has been applied on the following on the poker dataset. The results are shown in Table 1 for various values of *K*.

From Table 1, it has been seen that the optimal value of K is 8. The same dataset has also been tested in hadoop platform with 4-nodes and results has been obtained as follows [8, 9] (Table 2):

The result for hadoop-based k-NN implementation has been collected for various values of K-NN. The optimal values were found to be k=8. It had been also found that total runtime including training and testing time was 41040 which is much higher than the result obtained in the spark framework in Table 1.

Table 1 Experimental results

k	Training time (s)	Testing time (S)	Accuracy
2	7.06	54.87	60.37
3	7.01	60.02	60.21
4	7.01	71.18	61.42
5	7.41	61.55	61.31
6	7.32	74.88	62.04
7	7.22	109.25	61.79
8	7.41	109.73	62.39

Table 2 Experimental results for hadoop platform

k	Run Time (s.)	Accuracy (%)
8	41040	47.93

Table 1 shows the improvement in results—reduction of testing time and increase in accuracy.

Figure 4shows how the accuracy very with runtime with varying number of K-NN. The peak value of accuracy is 62.39 for which the training and testing time are 7.41 and 109.7.

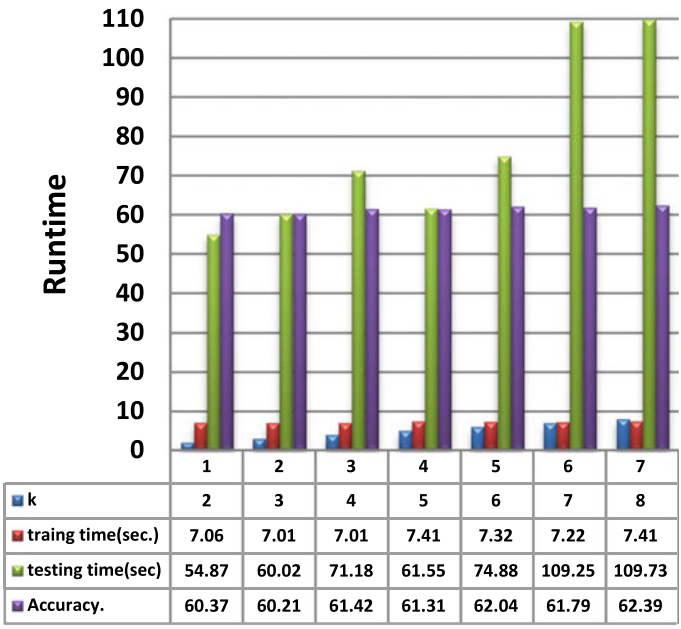


Fig. 4 Plot K versus traing time, testing time, and accuracy

6 Conclusion

The k-NN algorithm would be limited to small or medium data in its sequential form. The proposed method uses parallelization of the k-NN model for large dataset into order to improve efficiency. The results on Hadoop and spark framework further prove the efficiency of the spark-based model.

References

1. Mohammadi M, Al-Fuqaha A, Sorour S, Guizani M (2018) Deep learning for IoT big data and streaming analytics. IEEE Commun Surveys Tutorials 20:2923–2960
2. Zhou L, Pan S, Wang J, Vasilako AV (2017). Machine learning on big data: opportunities and challenges, Neurocomput Elsevier 237:330–361
3. Zhou L, Pan S, Wang J, Vasilako AV (2017) Machine learning on big data: opportunities and challenges. Neurocomput Elsevier 237:330–361
4. Zhang S, Deng Z, Cheng D, Zong M, Zhu X (2017) Efficient kNN classification algorithm for big data. Neurocomput Elsevier 195:143–148
5. Xia D, Li, H, Wang B, Li1 Y, Zhang Z (2016) A mapreduce-based nearest neighbor approach for big-data-driven traffic flow prediction. IEEE Acess 4:2920–2934
6. Maillo J, Ramírez S, Triguero I, Herrera F (2017) kNN-IS: an iterative spark-based design of the k-nearest neighbors classifier for big data. Knowle-Based Syst 117:3-15. ISSN 1872–7409 (In Press)
7. Liu T, Rosenberg C, Rowley HA (2007) Clustering billions of images with large scale nearest neighbor search. In: 2007 IEEE workshop on applications of computer vision (WACV '07). IEEE Conference
8. Ji J, Chung Y (2017) Research on K nearest neighbor join for big data. In: Proceedings of the 2017 IEEE international conference on information and automation (ICIA) Macau SAR, China, July 2017
9. Jha R, Bhattacharya V, Singh V, Mukherjee I (2018) K-NN Classification for large dataset in hadoop framework. In: International conference on nanoelectronics, circuits and communication systems. Springer, Singapore, pp 495–501

Website Development for Trading Between Farmers and Government

Pooja Asati, Sparsh Koushik, and Vijay Nath

Abstract Considering that 70% of the rural household depends primarily on agriculture in India, the need of digitization in the field of agriculture becomes important. Maintaining all the records of the farmers produce can ease out the process. Websites can be used by the consumers or the government for buying the produce from farmers in an efficient way. With this website, farmers and government (or consumers) can create their accounts as seller and buyer, respectively. Third-party buyers can also use the database to buy crops of interest. Authentication is done through username and password to get the separate ID. Users can use this ID to access the database relevant to them. This project will help in creating a database that will in turn help in better supply of crops from farmers to government (or consumers) without the need of any intermediate party.

Keywords Produce · Website · Trade

1 Introduction

The process of obtaining crops from farmers by government has seen less advancement over the years. With the advent of digital era and easy availability of Internet even to the remote villages, digital platforms can be used for entry of quality and quantity of crops by the farmers [1]. The database of these platforms can help the

P. Asati · S. Koushik · V. Nath (✉)
VLSI Design Group, Department of ECE, B.I.T. Mesra, Ranchi, Jharkhand 835215, India
e-mail: vijaynath@bitmesra.ac.in

P. Asati
e-mail: poojaasati04@gmail.com

S. Koushik
e-mail: sparshkoushik@gmail.com

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_69

government make better decisions on fair price and mechanism for obtaining the crops. Advanced computer science clustering algorithms can be used to group the farmers on the basis of location and crops to arrange for transport in a cheaper way [2–4].

This is very beneficial to the government to keep records of all the crops and their prices set by the different farmers in one single database. This website would make our process more efficient and easier. A government can view the prices and the list of crops anytime, whereas farmers can update the details whenever required. Account details of farmers are provided for money transactions. Not just the government, third-party buyers can also be given access to this database for fairer market.

Several applications are helping existing government schemes and reaching farmers in rural India [10, 11]. The information provided by these applications is mainly on the weather forecast, knowledge of fertilizers, seeds, and machinery and gives news regarding the government's agricultural policies and schemes. There is currently no scheme by which farmers can sell their crops to the end consumers (or government) directly by setting their prices of crops and then communicating with them further.

The main objective of the website is to facilitate the trading of crops between farmers and the government (or consumers) [5–8]. The website is made user-friendly, so that farmers can easily use it. It will help farmers to sell crops at a reasonable price directly to the government without the need of an intermediate body [9]. The government will have the location-based knowledge of the crops available with the farmers, this would in turn make collection of crops easier and efficient. The requirement involves quantity, expected price of crops, location, and bank details of a farmer.

This paper is constructed in following manners: The proposed work is explained in Sect. 2. The project description is given in Sect. 3. Section 3 gives the data flow chart and website description. System implementation is demonstrated in Sect. 4. The implementation shows the interface of the website and the functions supported.

2 Proposed Work

The proposed work aims at developing a website that can facilitate trading between government and farmers. It captures the quantity and price of crops set by different farmers. Using this website, farmers can easily update information about crops on the server which can be accessed by others. Initially, farmers have to use their login credentials, after which they can set price and quantity as desired. The government can see which farmer has updated what crop, its quantity, and expected price.

A database would be created. The database would have information on what crops are up for sale, what is the expected pricing of the crops, and also the location where the crop's available. The buyer which may be the government or any third-party trader can access the database to know which crops are available at a location, the quantities in which they are available, and the price that the farmer is expecting for

those crops. Using a location and quantity-based filters, the buyer can choose which farmers would be suitable for trading.

The website includes building different architecture:

1. The front end of the website has been developed by HTML, CSS, JavaScript.
2. The back end or the server coding has been developed by nodejs.
3. The database is created by MongoDB.
4. The application has been created through React Native.
5. The website has been hosted on Amazon Web Services with the IP address 13.233.152.146.

HTML is a markup language used to create webpages. It describes the structure of a website. CSS is used for describing the formatting and look of a document written in a markup language. It is basically used for styling the webpages. JavaScript is a high-level, interpreted, and dynamic programming language. It allows webpages to respond to a user activity and dynamically update themselves, and all without requiring a webpage reload to change its appearance [12, 13].

Node.js is a platform which is built on Chrome's Javascript runtime for easily building scalable and fast network applications. Node.js is an open-source, cross-platform runtime environment for server-side development [14–16].

Web application requires a database for storing and retrieving data when needed. MongoDB is a popular cross-platform document-oriented NoSQL database. Node.js has ability to work with MongoDB databases. MongoDB enables faster access of data due to its nature of using internal memory for the storage. It provides an improved security to the database [17].

React Native is a popular JavaScript framework for writing mobile applications for IOS and Android [19]. It is a UI focused, which makes the application to load quickly and give a smoother feel. It provides faster development, the ease to be maintained, and time saving. AWS or Amazon Web Services is a secure, comprehensive cloud services platform, offering different functionalities [18]. We have used AWS for hosting Web applications.

3 Project Description

The project description provides the following details: the data flow of the website, the structure of the website, and the benefits served by this project.

1. **Data Flow Chart**
   The data flow diagram shows that farmer provides the relevant details of crops which can be viewed in the government portal. Government can see the details of the crops of all the farmers who have registered through this website. The location of the farmers is also available (Fig. 1).

2. **Website**

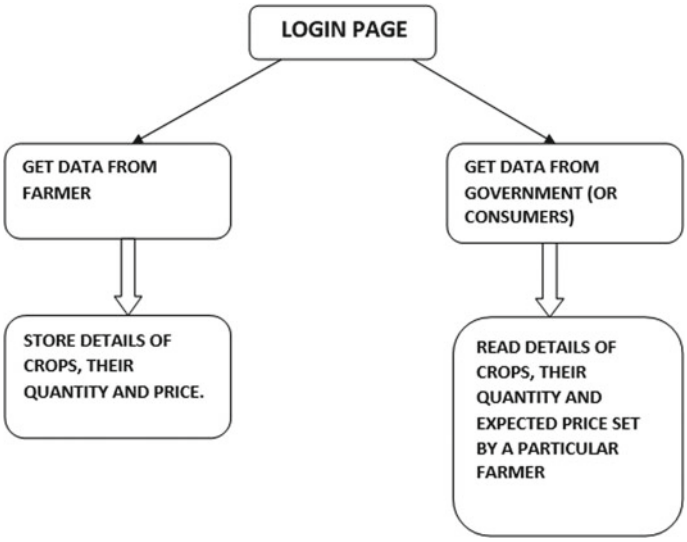


Fig. 1 Data flow diagram of website

The website contains following pages:

1. Login Page
2. Sign up Page
3. Dashboard for farmer
4. Dashboard for Government (Buyer)

1. **Login Page**
   There are two edit text box and two buttons. One edit text is for entering username, and another one is for submitting a password. If the user is an existing one, they can use sign in button or else the user need to register using sign up button.
2. **Sign Up or Registration Page**

This page is used only for new users. There are six edit text, one round button, and two buttons.

Edit Text:

1. Username
2. Full Name
3. Email
4. Location
5. Password
6. Confirm Password

Round Button:

1. Admin
2. Farmer

Button:

1. Sign Up
2. Sign In

3. **Dash Board for Farmer**

This page provides the previous details of crops uploaded by farmer and an additional button for adding crops to the list. Using an 'ADD CROP' button, farmers can upload the crop name, its quantity, and expected price. This is then uploaded using 'SUBMIT CROP' button. There is a 'LOGOUT' button at the top which is used for signing out of the farmer's account.

4. **Dashboard for Government**

This page provides the different list of farmers and their contact details who wants to sell their produce. There is also one button 'CROP LIST' which on clicking displays the list of crops, its quantity, and the expected price uploaded by a particular farmer. There are additional two buttons at the top. One is 'HOME' button that is used to navigate to the first page in government portal. Second is the 'LOGOUT' button that is used for signing out of the account.

4 System Implementation

The following section shows the various pages of the website that has been created. To access the website, one has to access the following IP address - 13.233.152.146.

After opening the website, one should login or sign up based on whether they already have an account or not (Figs. 2, 3, 4, 5, and 6).

5 Conclusion

The model of a website has been developed which is aimed at providing a data entry platform for the producers of crops. The database created by this entry platform can be used by the government or third-party buyers for forming policies for acquisition of the crops. The website provides option of either being a farmer or an administrator. The farmers are the sellers, and the admins are the buyers. The implementation of the website platform can be extended to an Android application to further smoothen the process of data collection. The database has been attached within the website, and this database provides opportunities for a location-based clustering which would make the transportation cheap.

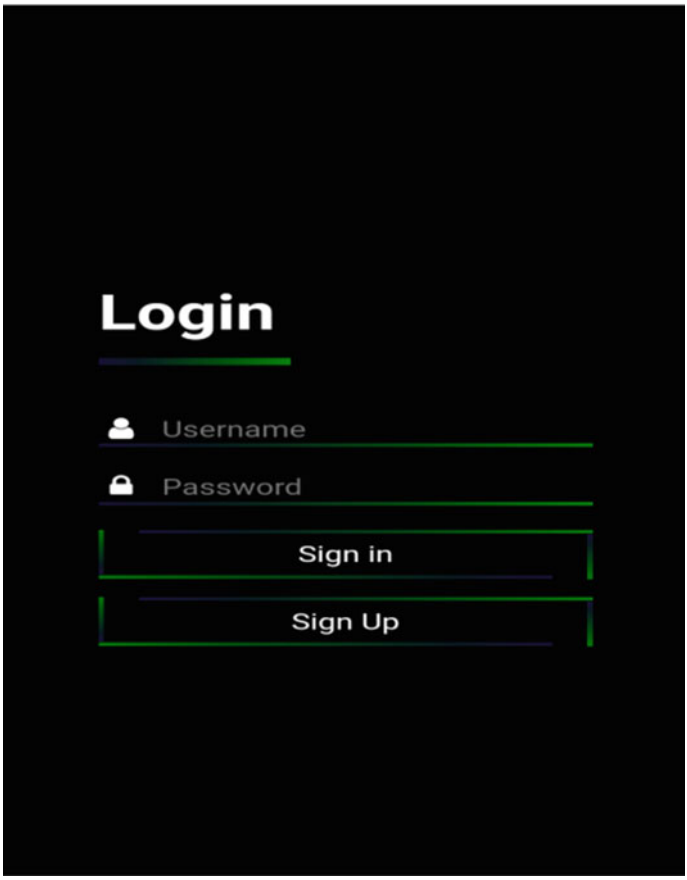


Fig. 2 Login page of website

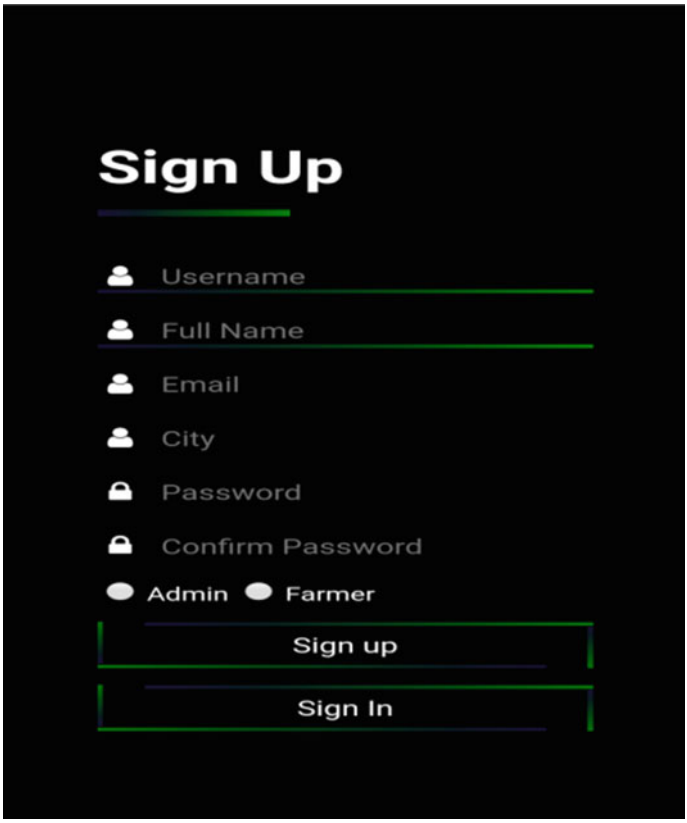


Fig. 3 Sign up page of website

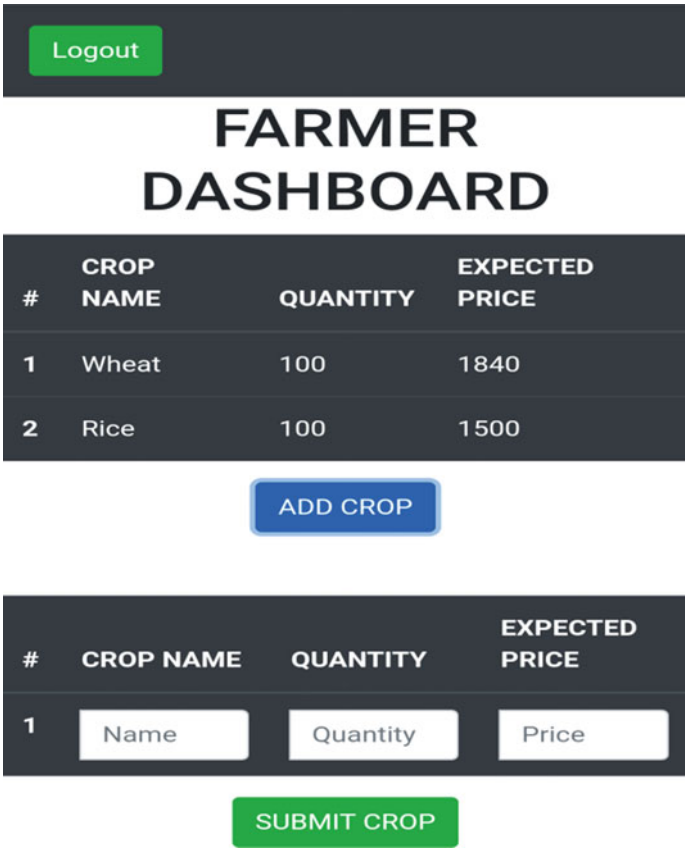


Fig. 4 Farmer dashboard

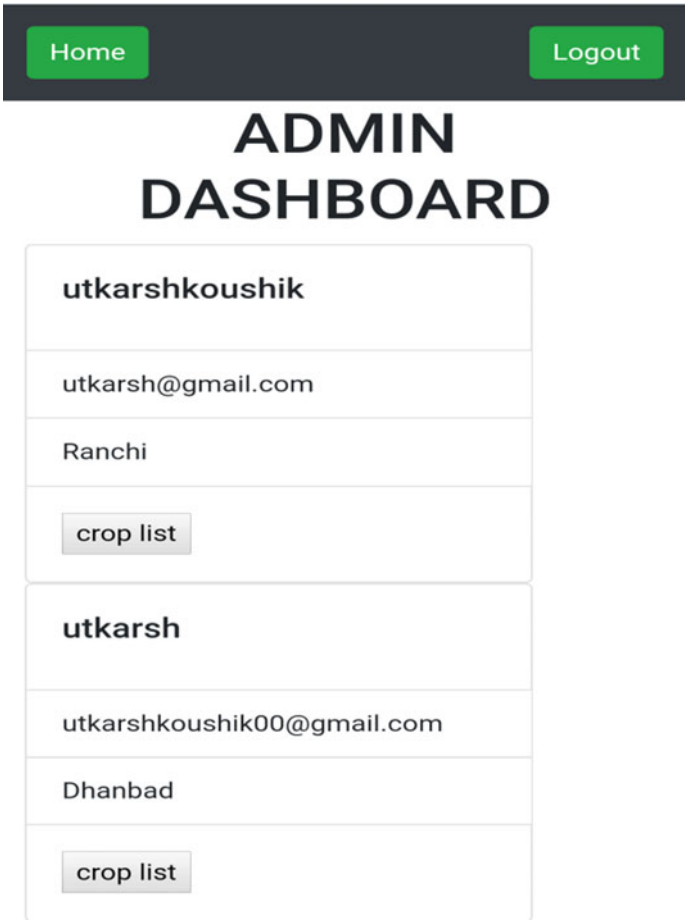


Fig. 5 Administrator dashboard

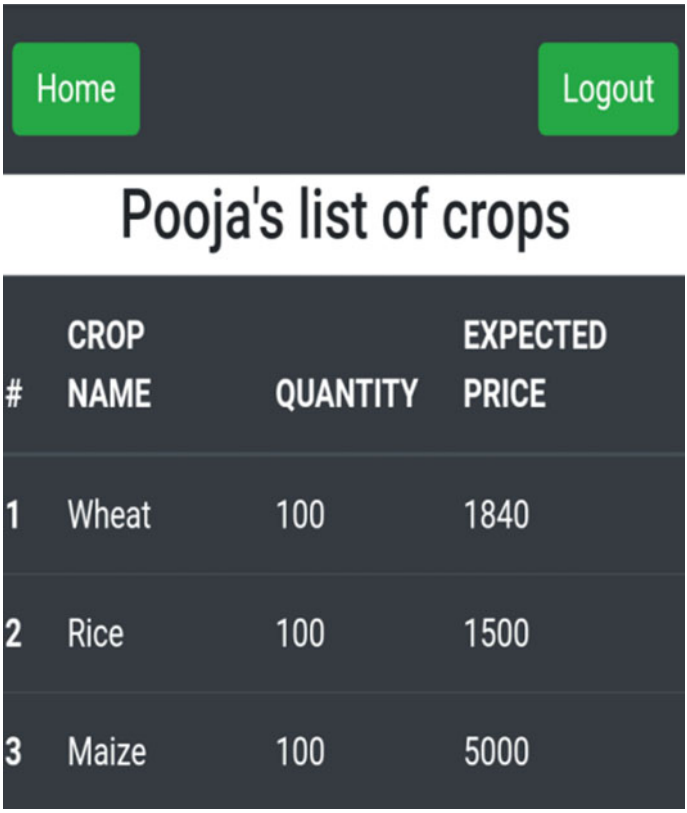


Fig. 6 Database view in admin login

Acknowledgements We are thankful to our Prof. Vijay Nath, VLSI Design Group, Department of ECE, for providing the support and guidance for this research work. We extend our gratitude to the management and professors of our institute for providing all the necessary facilities and never-ending support for the work.

References

1. Gangeshwer DK (2013) E-commerce or internet marketing: a business review from Indian context. Int J u-and e-Ser Sci Technol 6(6):187–194. ISSN: 2005–4246
2. Preciado JC, Linaje M, Comai S et al (2007) Designing rich internet applications with web engineering methodologies. IEEE International Workshop on Web Site Evolution, IEEE Computer Society, pp 23–30
3. Zhang M (2011) Application of computer technology in modern agriculture. Agricul Eng 1(4):26–28
4. Mittal SC (2001) Role of information technology in agriculture and its scope in India. Fertilizer News 46(12):83–87

5. Jiang F (2009) Research on the creative model of the rural supply chain of and strategies. On Econ Probl 12:81–84
6. Deng J, Qi C, Peng X (2009) Consideration on the construction of agricultural supply chain based on the food quality and safety—take citrus as example. Hubei Agricul Sci 48(7)
7. Li X, Tang L (2009) Research on brand strategy of agricultural supply chain management. Log Eng Manage 31(5):65–66
8. Fang L (2009) Research on information management of agricultural supply chain. Chinese Academy of Agricultural Sciences, pp 14–2
9. Ze-long Z, Yu T (2014) Discussion about agricultural e-commerce situation and optimization.In: Management science & engineering (ICMSE), 2014 international conference on information technology, computer, and electrical engineering, pp 88–95
10. Suprem A, Mahalik N, Kim K (2013) A review on application of technology systems, standards and interfaces for agriculture and food sector. Comput Stand Inter 35:355–364
11. Behera SB, Panda B et al (2014) Information communication technology promoting retail marketing. in agriculture sector in India as a study. In: International conference on intelligent computing, communication & convergence (ICCC-2014)
12. Fu C (2016) Exploration of Web front-end development technology and optimization direction. In: International conference on electronics network and computer engineering
13. B Patel S Roy D Bhattacharyya Tai-Hoon K (2017) Necessity of big data and analytics for good e-governance. Int J Grid Distrib Comput SERSC 10(8):11–20
14. Wilson JR Node.js the right way:practical server side javascript that scales. The Pragmetic express. ISBN-13: 978–1937785734
15. Lei K, Ma Y, Tan Z (2014) Performance comparison and evaluation of web development technologies in PHP, Python and Node,js. In: Computational science and engineering (CSE). IEEE, Page(s) 661–668
16. Tilkov S, Vinoski S (2010) Node.js: using javascript to built high-performance network programs. Int Comput IEEE 14(6): 80–83
17. Dickey J (2015) Write modern web apps with the mean stack: mongo, express. In: Angular JS, Node JS (eds) 1st edn. Pearson
18. Saakshi N, Arushi J, Prachi (2015) Cloud computing security: amazon web service. Int Conf Adv Comput Commun Technol
19. Kaushik V, Gupta K, Gupta D (2019) React native application development. In: International conference on cyber security and privacy in communication networks

Telemetry-Based System for Data Acquisition of Agricultural Practices in Rural Areas

Shantanu Jahagirdar, Narendra Manglani, Saket Rai, and Vijay Nath

Abstract The data acquisition telemetry system is a versatile and economical means to accumulate data from multiple sensors at remote locations over an extended period of time. The data obtained is normally transferred to the final destination and saved for further analysis. This paper introduces the design and implementation of a simple and economical telemetry system to collect and transfer agricultural crop-related data from remote farming locations to the servers where it is evaluated. The proposed data transfer system is by utilization of package-switching networks (GPRS and Internet) to transfer the collected data. Data that will be collected includes soil temperature, soil pH, moisture content, and other similar information and will be collected through suitable sensors and transducers.

Keywords Telemetry · Wireless communication · Sensors · GPRS

1 Introduction

i. Government Initiatives

To promote direct interactions of farmers with consumers in fresh produce, there have been farmers' markets in India in the form of

S. Jahagirdar (✉) · N. Manglani · S. Rai
Department of EEE, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: shantanu2jahagirdar@gmail.com

N. Manglani
e-mail: narendramanglani04@gmail.com

S. Rai
e-mail: saketrai3005@gmail.com

V. Nath
Department of ECE, Birla Institute of Technology, Mesra, Ranchi, Jharkhand 835215, India
e-mail: vijaynath@bitmesra.ac.in

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_70

Some of the important schemes in agriculture are [1]:

(a) **Soil health card scheme**

The scheme has been launched in 2015 to assist state government to provide soil health cards to all the farmers. The soil health cards provide information to farmers on nutrient status of their soil along with recommendation on appropriate dosage of nutrients to be applied for improving soil health and its fertility [2].

(b) **Paramparagat Krishi Vikas Yojana (PKVY)**

PKVY was implemented to promote organic farming in the country and to improve soil health and organic matter content and increase the net income of farmers.

(c) **National Agriculture Market (e-NAM)**

This scheme helps farmers to provide an e-marketing platform at a national level and support creation of infrastructure to enable e-marketing.

(d) **Rainfed Area Development Programme (RADP)**

Rainfed Area Development Programme (RADP) was implemented as a sub-scheme under (RKVY).

The purpose of a telemetry system is to collect data at a place that is remote or inconvenient and to relay the data to a point where the data may be evaluated. Telemetry has the advantage of reducing the cost of manually reading, checking, and controlling remote devices [3].

The use of wireless technology enables systems to be located virtually anywhere without depending on other factors which may prove to be challenging and humanly inaccessible [4].

Telemetry system has a wide range of applications, such as

- Environmental monitoring
- Space applications
- Security applications
- Load management
- Agriculture.

Its main aim was to improve the quality of life of the farmers. It helped farmers to increase agricultural productivity of rainfed areas.

i. **Telemetry System Categories**

In order to evaluate the type of communication required for a telemetry application [5], it is important to consider telemetry systems in specific categories, i.e.,

(a) High integrity, High data throughput, Continuous transfer
(b) High integrity, Low data throughput, Continuous transfer
(c) High integrity, Low data throughput, Intermittent transfer
(d) Low integrity, High data throughput, Continuous transfer
(e) Low integrity, High data throughput, Intermittent transfer

Fig. 1 Rythu bazaar in Andhra Pradesh [4]

(f) Low integrity, Low data throughput, Intermittent transfer.

Based on these different categories, different communication links are established. These directly affect the complexity of the communication system.

Based on the required data, the concerned system needs the type f), i.e., Low integrity, Low data throughput, Intermittent transfer.

As the crop data required does not change much with respect to time [6], communication can be intermittent and frequent data transfer is not necessary. The data collected is not expected to be complex and thus can be collected easily.

The sensors which are to be installed to collect the data have been explained in detail next. These sensors are expected to take accurate readings of the proposed data and transfer it [7].

The data transfer technique proposed is via Internet over a GPRS network [8]. This was chosen because GPRS network is easily accessible in even the most remote areas of the country and is quite affordable for everyone and is reliable, safe, and fast. Detailed explanation is given next (Fig. 1).

2 Data Acquisition

Various sensors [9] will be used to determine the needed parameters. By deploying sensors and mapping fields, farmers can begin to understand their crops at a microscale, conserve resources, and reduce impacts on the environment. These sensors include:

Location Sensors: Use signals from GPS satellites to determine latitude, longitude, and altitude to within feet. Three satellites minimum are required to triangulate a position. Precise positioning is the cornerstone of precision agriculture. GPS integrated circuits like the NJR NJG1157PCD-TE1 are a good example of location sensors.

Electrochemical Sensors: provide key information required in precision agriculture: pH and soil nutrient levels. Sensor electrodes work by detecting specific ions in the soil. Currently, sensors mounted to specially designed "sleds" help gather, process, and map soil chemical data.

Mechanical Sensors: measure soil compaction or "mechanical resistance." The sensors use a probe that penetrates the soil and records resistive forces through the use of load cells or strain gauges. For example, EC-5 determines volumetric water content (VWC) by measuring the dielectric constant of the media using capacitance/frequency domain technology [10].

A similar form of this technology is used on large tractors to predict pulling requirements for ground engaging equipment. Tensiometers, like Honeywell FSG15N1A, detect the force used by the roots in water absorption and are very useful for irrigation interventions.

Airflow Sensors: measure soil air permeability. Measurements can be made at singular locations or dynamically while in motion. The desired output is the pressure required to push a predetermined amount of air into the ground at a prescribed depth. Various types of soil properties, including compaction, structure, soil type, and moisture level, produce unique identifying signatures.

Agricultural Weather Stations: are self-contained units that are placed at various locations throughout growing fields. These stations have a combination of sensors appropriate for the local crops and climate. Information such as air temperature, soil temperature at various depths, rainfall, leaf wetness, chlorophyll, wind speed, dew point temperature, wind direction, relative humidity, solar radiation, and atmospheric pressure are measured and recorded at predetermined intervals. This data is compiled and sent wirelessly to a central data logger at programmed intervals. Their portability and decreasing prices make weather stations attractive for farms of all sizes.

Optical Sensors: use light to measure soil properties. The sensors measure different frequencies of light reflectance in near-infrared, mid-infrared, and polarized light spectrums. Sensors can be placed on vehicles or aerial platforms such as drones or even satellites. Soil reflectance and plant color data are just two variables from optical sensors that can be aggregated and processed. Optical sensors have been developed to determine clay, organic matter, and moisture content of the soil. Vishay, for example, offers hundreds of photodetectors and photodiodes, a basic building block for optical sensors (shown in Fig. 2).

3 Data Communication System

The proposed transfer mechanism (Fig. 3) consists of three main parts:

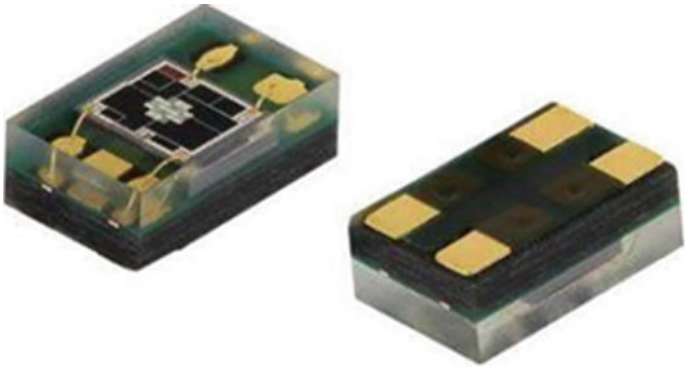


Fig. 2 Basic building block of optical sensor

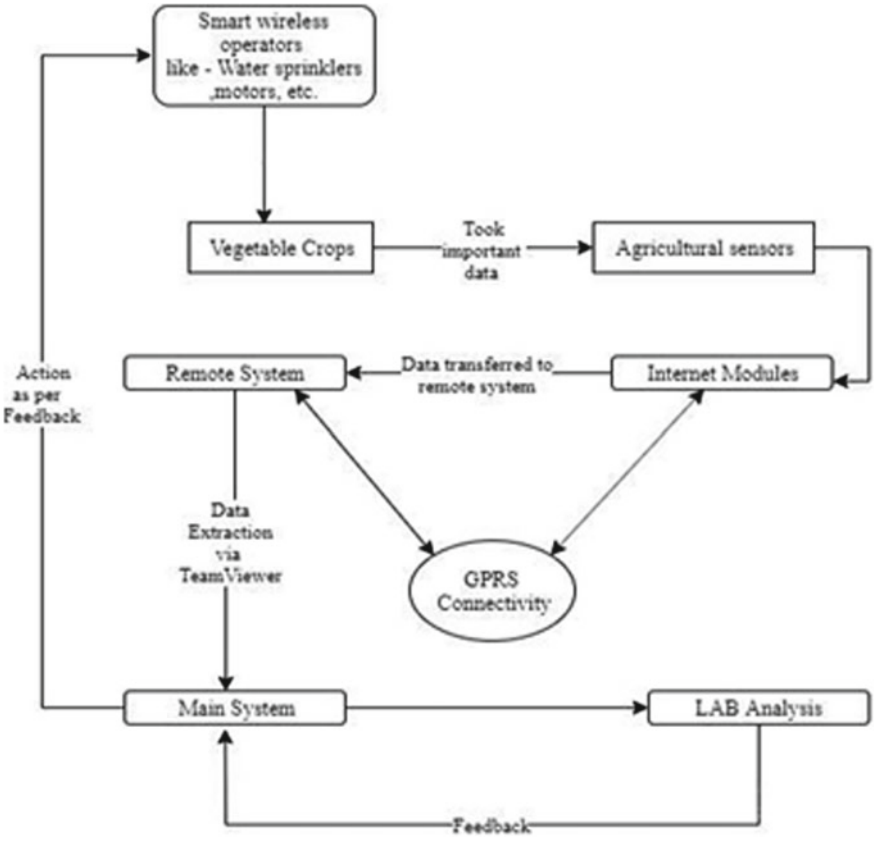


Fig. 3 Designed flowchart for proposed telemetry system

(e) **Remote System**: It is the main system situated near the place from where data has to be collected. This system consists of a computer which is connected to a wired or a wireless Internet network (via GPRS) and has the application TeamViewer installed on it. Data from the various sensors is collected here in a tabular form (preferably) at different time intervals which can be set according to the requirement.

(f) **Remote telemetry units(RTUs)**

An RTU monitors the status of equipment and the environment at your site, reporting an alarm to you when something out of the ordinary occurs. RTUs typically set three types of alarms [11]:

1. Discrete Alarms
2. Analog Alarms
3. Ping Alarms.

(a) **Communication**: Any two computers can communicate with each other through Internet; in order to provide direct access from one computer to another, they should be assigned public IP addresses [12]; using public IP addresses requires to rent a public IP address and extra equipment. However, the application we are using (telemetry for agricultural applications) does not require high-speed connection or continuous connection [13].

TeamViewer software [14] is an application that allows direct computer access between two computers with private IP addresses on the Internet. It can be used to remotely login to authorized computers to access information via the Internet and is a fast, safe, and reliable way to access remote information.

(b) **Main System**: This is the main computer where the transferred data will be evaluated, and the results will be published. All the data collected by the Remote system will be stored here for laboratory evaluation by experts and correspondingly, crop monitoring can be done, farmers can be informed of better practices, and can be advised on the type of farming they should follow based on the results.

This will result in better utilization of available resources, better communication with the farmers and would overall result in an increase in food production as farming would be monitored by experts both an off the field.

4 Need for Reform in Agricultural Marketing

1. Empower producers with knowledge, information and capability to undertake market-driven production.
2. Provide multiple choice and competitive marketing channels to farmers.
3. Attract large-scale investments needed for building.
4. Post-harvest infrastructure [15].

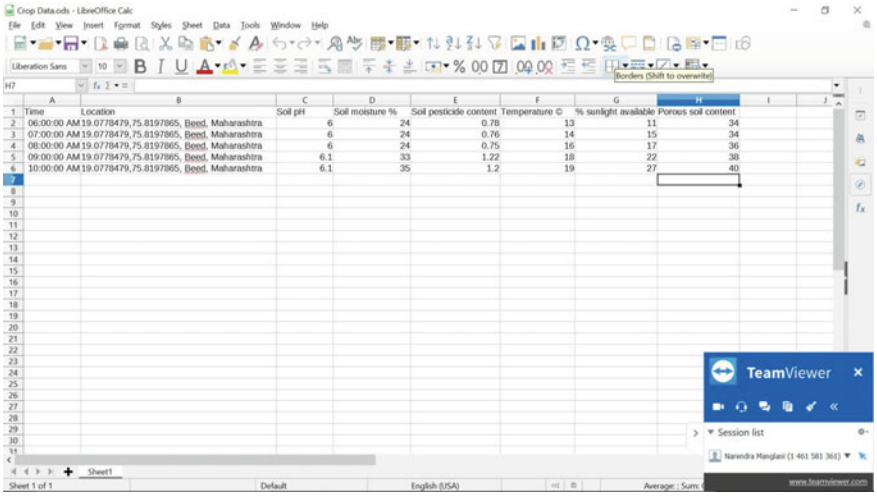


Fig. 4 Data stored in a tabular form in the remote system

5 Software Implementation

Using the TeamViewer software mentioned above, the implementation of the project was done. We took two different systems, one acting as a remote system where the data was being stored, one as the main system from where the data was being accessed. An excel sheet was made where real-time data of the crop was stored. A cotton field in the district Beed, Maharashtra, was chosen as the object for case study. Real-time soil-related data like soil pH, temperature, water-retaining capacity, pesticide content was stored in a tabular form and the results were transferred to the remote system using Internet modules. Given next are the results of the implementation (Figs. 4 and 5).

6 Linking Small Farmers to Market

We can try to make a universal account for all the farmers and the traders so that all the government IDs [16] and the bank account information of them can be stored within it and the farmer can manually add the details of the crop which has been produced in the period and can sell them manually at a reasonable price to the traders so that the farmers do not have to suffer more [17]. That universal account will help the farmers and traders to directly transact the money from their bank accounts without any risk.

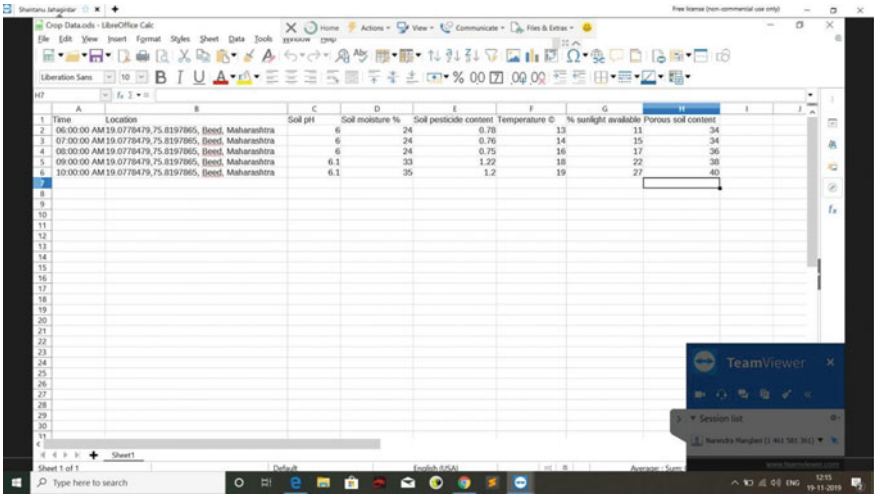


Fig. 5 Data being accessed by the remote system

7 Conclusion

This paper attempted to propose a system for monitoring and better understanding techniques to eliminate different kinds of problems faced in agriculture. It introduces the design and implementation of a flexible [18] and practical telemetry system to collect and transfer the agricultural data from rural areas to the processing and displaying destination. The implemented system has the advantage of its simplicity, low cost, and easy to adapt with the available means of communication [19].

It will also result in creation of jobs and employment and will also result in reduction in the overall budget spending on agricultural practices [20].

This system can be considered as a generic system which can be utilized for many other applications [21].

Acknowledgements We would like to thank the Vice-Chancellor, Birla Institute of Technology, Dr S. Konar, Dean Academic Programs Dr. P. Padmanabhan, Dean Student Welfare Dr Anand Kumar Sinha, HoD of the EEE Department, Dr P.R Thakura and HoD of the ECE Department, Dr. Srikanta Pal for giving us this opportunity to research and prepare a low-cost data communication system to help the agricultural sector of our country, and for encouraging research and giving the right guidance to us. Without their help, we couldn't have completed this research. This would also help us in our future studies.

References

1. Rythu Bazars (AP), http://market.ap.nic.in/rythu_bazars.html
2. http://spiremeteringtechnology.blogspot.com/2013/09/what-are-advantages-of-wireless.html

3. Jabir S.A., Osama A.H.: Design and Implementation of a Telemetry System for Environmental Applications, pp. 2–3 (2009)
4. Norbert, F., Manuel, S.: Telemetry based Flexible Embedded System (Agriculture and Aquaculture), pp. 1–4. Fraunhofer Institute of Optronics IOSB (2014)
5. Governmental Agricultural Schemes, www.indiaagronet.com
6. Soil Properties: www.sciencelearn.org.nz
7. Telemetry monitoring: Advantages and Technologies https://www.dpstele.com/network-monitoring/telemetry/protection-technologies.php
8. Halit, E. (2005) Telemetry Systems, p. 510 (2005)
9. GPRS-explained—searchmobilecomputing.techtarget.com
10. An agricultural telemetry system implemented using Arduino: https://ieeexplore.ieee.org/document/7016199
11. Benefits of Telemetry http://www.ferret.com.au/c/jo-com-telemetry/benefits-of-telemetry-n713059
12. Mouser Electronics-Smart Agricultural Sensors. mouser.in/applications/smart-agriculture-sensors
13. https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/telemetry
14. Public IP addresses-cloud.docs.oracle.com
15. TeamViewer software: www.teamviewer.com
16. Hanumanth Rao, C.H.: Reform Agenda for Agriculture, pp. 615–620 (2003)
17. Aadhaar Card: uidai.gov.in
18. Soil moisture sensor: https://www.metergroup.com/environment/products/ec-5-soil-moisture-sensor/
19. Products and services-Farming data systems http://www.farmdatasystems.com/products
20. Climatic conditions for agriculture http://www.fao.org/climate-smart-agriculture/en
21. Basics of Telemetry system: https://stackify.com/telemetry-tutorial/

Correction to: Segregation of Speech and Music Signals for Aiding the Hearing Impaired

Ankit Ranjan, Shubhankar Sinha, Anuranjana, and Deepti Mehrotra

Correction to:
Chapter "Segregation of Speech and Music Signals for Aidingthe Hearing Impaired" in: V. Nath and J. K. Mandal (eds.),
***Nanoelectronics, Circuits and Communication Systems*,**
Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_57

The original version of the book was published with incorrect author name "Ankita Ranjan" has been corrected as "Ankit Ranjan" in the chapter "Segregation of Speech and Music Signals for Aiding the Hearing Impaired" . The chapter and book have been updated with the change.

The updated version of this chapter can be found at
https://doi.org/10.1007/978-981-15-7486-3_57

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3_71

Author Index

V. Nath and J. K. Mandal (eds.), *Nanoelectronics, Circuits and Communication Systems*, Lecture Notes in Electrical Engineering 692,
https://doi.org/10.1007/978-981-15-7486-3