Debashis Dutta
Haranath Kar
Chiranjeev Kumar
Vijaya Bhadauria   *Editors*

Advances in VLSI, Communication, and Signal Processing

Select Proceedings of VCAS 2018

Springer

Lecture Notes in Electrical Engineering

Volume 587

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

Aninda Bose, Senior Editor (aninda.bose@springer.com)

Japan

Takeyuki Yonezawa, Editorial Director (takeyuki.yonezawa@springer.com)

South Korea

Smith (Ahram) Chae, Editor (smith.chae@springer.com)

Southeast Asia

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

USA, Canada:

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

All other Countries:

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

** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink **

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

Debashis Dutta · Haranath Kar ·
Chiranjeev Kumar · Vijaya Bhadauria
Editors

Advances in VLSI, Communication, and Signal Processing

Select Proceedings of VCAS 2018

Springer

Editors
Debashis Dutta
Ministry of Electronics and Information
Technology
New Delhi, Delhi, India

Chiranjeev Kumar
Indian Institute of Technology (ISM)
Dhanbad, Jharkhand, India

Haranath Kar
Motilal Nehru National Institute
of Technology Allahabad
Allahabad, Uttar Pradesh, India

Vijaya Bhadauria
Motilal Nehru National Institute
of Technology Allahabad
Allahabad, Uttar Pradesh, India

VCAS 2018 Committees

Chief Patron

Prof. Rajeev Tripathi, Director, MNNIT Allahabad

Patron

Prof. Sudarshan Tiwari, MNNIT Allahabad

General Chair

Prof. Vijaya Bhadauria, MNNIT Allahabad

Program Chair

Prof. Haranath Kar, MNNIT Allahabad

Organizing Chair

Prof. V. K. Srivastava, MNNIT Allahabad

Keynote Chair

Prof. Amit Dhawan, MNNIT Allahabad

Tutorial Chair

Prof. R. K. Nagaria, MNNIT Allahabad

Workshop Chair

Prof. V. S. Tripathi, MNNIT Allahabad

Public Relation Chair

Prof. R. A. Mishra, MNNIT Allahabad

Publication Chair

Prof. Richa Negi, MNNIT Allahabad

Finance Chair

Dr. Arvind Kumar, MNNIT Allahabad

Program Co-Chair

Dr. V. Krishna Rao K., MNNIT Allahabad

Event Management Chair

Er. Asim Mukherji, MNNIT Allahabad

Publicity Chair

Dr. Basant Kumar, MNNIT Allahabad

Industrial Chair

Dr. Manish Tiwari, MNNIT Allahabad

Hospitality Chair

Dr. Sanjeev Rai, MNNIT Allahabad

Registration Chair

Dr. S. K. Gupta, MNNIT Allahabad

Resource Generation Chair

Dr. Y. K. Prajapati, MNNIT Allahabad

Website Chair

Dr. Divya Kumar, MNNIT Allahabad

Workshop Co-Chair

Er. Rajeev Gupta, MNNIT Allahabad

Hospitality Co-Chair

Dr. Shweta Tripathi, MNNIT Allahabad

Tutorial Co-Chair

Dr. P. Karuppanan, MNNIT Allahabad

Registration Co-Chair

Dr. Bhuktare Swapnil Sopanrao, MNNIT Allahabad

Organizing Secretary

Dr. Arun Prakash, MNNIT Allahabad

Advisory Committee

Prof. M. Jagadesh Kumar, JNU Delhi
Prof. A. K. Chaturvedi, IIT Roorkee
Prof. N. C. Shivaprakash, IISc Bangalore
Prof. Sagar Naik, University of Waterloo, Canada
Prof. David Harvey, Liverpool John Moores University, UK
Prof. S. C. Dutta Roy, IIT Delhi
Prof. S. S. Pathak, IIT Kharagpur
Prof. Asok De, DTU Delhi
Prof. Raj Senani, NSIT Delhi
Prof. Rajendra Prasad, IIT Roorkee
Prof. A. K. Misra, SPMIT Allahabad
Dr. Debashis Dutta, MeitY GoI

Prof. Shekhar Verma, IIIT Allahabad
Prof. Suneeta Agarwal, MNNIT Allahabad
Prof. M. M. Gore, MNNIT Allahabad
Prof. R. S. Yadav, MNNIT Allahabad
Prof. Neeraj Tyagi, MNNIT Allahabad
Prof. Vineeta Agrawal, MNNIT Allahabad
Prof. R. K. Singh, MNNIT Allahabad
Prof. Subhi Purwar, MNNIT Allahabad
Prof. R. K. Tripathi, MNNIT Allahabad
Prof. Sudarshan Tiwari, MNNIT Allahabad
Prof. Rajeev Tripathi, MNNIT Allahabad
Prof. Haranath Kar, MNNIT Allahabad
Prof. V. K. Srivastava, MNNIT Allahabad
Prof. Vijaya Bhadauria, MNNIT Allahabad
Prof. Amit Dhawan, MNNIT Allahabad
Prof. R. K. Nagaria, MNNIT Allahabad
Prof. V. S. Tripathi, MNNIT Allahabad
Prof. R. A. Mishra, MNNIT Allahabad

Technical Program Committee

Prof. Naveen Chilamkurti, La Trobe University Australia
Prof. Raveendra Rao, Western University, London, Ontario, Canada
Prof. Brahmjit Singh, NIT Kurukshetra
Prof. Ganapati Panda, IIT Bhuawaneshwar
Prof. Shrish Verma, NIT Raipur
Prof. K. D. Kulat, VNIT Nagpur
Prof. Anupam Shukla, ABV-IIITM Gwalior
Prof. Chiranjeev Kumar, IIT (ISM) Dhanbad
Prof. Manav Bhatnagar, IIT Delhi
Prof. Y. N. Singh, IIT Kanpur
Prof. Aditya Trivedi, ABV-IIITM Gwalior
Prof. Nagendra P. Pathak, IIT Roorkee
Prof. K. K. Mohapatra, NIT Rourkela
Prof. D. R. Bhaskar, Jamia Millia Islamia Delhi
Prof. M. P. S. Bhatia, NSIT Delhi
Prof. Chandan Kumar Sarkar, Jadavpur University
Prof. Abdul Quaiyum Ansari, Jamia Millia Islamia Delhi
Prof. Kavita Khare, MANIT Bhopal
Prof. R. K. Baghel, MANIT Bhopal
Prof. M. M. Tripathi, DTU Delhi
Dr. Ming-Fong Tsai, National United University, Taiwan
Dr. Dhananjay Singh, Hankuk University of Foreign Studies, South Korea

Dr. Tamer Mostafa Abdelkader, Ain Shams University, Egypt
Dr. Sudeb Dasgupta, IIT Roorkee
Dr. Tarek Khalifa, Antalya Bilim University, Turkey
Dr. Ram Bilas Pachori, IIT Indore
Dr. Arvind Sharma, NIT Kurushetra
Dr. Nitesh Purohit, IIIT Allahabad
Dr. Rajiv Tripathi, NIT Delhi
Dr. B. K. Mohanti, JUET Guna
Dr. Tarun Kumar Gupta, MANIT Bhopal
Dr. Gargi Khanna, NIT Hamirpur
Dr. Rajat Kumar Singh, IIIT Allahabad
Dr. Manish Kumar, IIIT Allahabad
Dr. K. P. Singh, IIIT Allahabad
Dr. Satinder Sharma, IIT Mandi
Dr. Vinay Kumar, VNIT Nagpur
Dr. Pawan Kumar Verma, NIT Jalandhar
Dr. A. K. Shankhwar, HBTU Kanpur
Dr. Sarsij Tripathi, NIT Raipur
Dr. A. S. Raghuwanshi, NIT Raipur
Dr. Krishan Kumar, NIT Hamirpur
Dr. Nishu Gupta, VCE Warangal
Er. Asim Mukherjee, MNNIT Allahabad
Dr. Arvind Kumar, MNNIT Allahabad
Dr. V. Krishna Rao K., MNNIT Allahabad
Dr. Basant Kumar, MNNIT Allahabad
Dr. Y. K. Prajapati, MNNIT Allahabad
Dr. Manish Tiwari, MNNIT Allahabad
Dr. Sanjeev Rai, MNNIT Allahabad
Dr. Santosh Kumar Gupta, MNNIT Allahabad
Dr. Arun Prakash, MNNIT Allahabad
Dr. ShwetaTripathi, MNNIT Allahabad
Dr. P. Karuppanan, MNNIT Allahabad
Er. Rajeev Gupta, MNNIT Allahabad
Dr. Bhuktare Swapnil Sopanrao, MNNIT Allahabad
Prof. Paulson Samuel, MNNIT Allahabad
Prof. Asheesh Kumar Singh, MNNIT Allahabad
Prof. Rajesh Gupta, MNNIT Allahabad
Prof. Nand Kishor, MNNIT Allahabad
Prof. Richa Negi, MNNIT Allahabad
Dr. Nitin Singh, MNNIT Allahabad
Dr. Navneet Kumar Singh, MNNIT Allahabad
Dr. Deepak Kumar, MNNIT Allahabad
Dr. Niraj Kumar Choudhary, MNNIT Allahabad
Dr. M. Venkatesh Naik, MNNIT Allahabad
Prof. D. S. Kushawaha, MNNIT Allahabad

Prof. Anil Kumar Singh, MNNIT Allahabad
Prof. D. K. Yadav, MNNIT Allahabad
Er. Rajesh Tripathi, MNNIT Allahabad
Er. Manoj Wariya, MNNIT Allahabad
Dr. Anoj Kumar, MNNIT Allahabad
Dr. Mayank Pandey, MNNIT Allahabad
Dr. Krishn Kumar Mishra, MNNIT Allahabad
Dr. Ranvijay, MNNIT Allahabad
Dr. Rupesh Kumar Dewang, MNNIT Allahabad
Dr. Pragya Dwivedi, MNNIT Allahabad
Dr. Divya Kumar, MNNIT Allahabad
Dr. Dushyant Kumar Singh, MNNIT Allahabad
Dr. Dinesh Singh, MNNIT Allahabad
Dr. Shashank Er. vastava, MNNIT Allahabad
Dr. Shashwati Banerjea, MNNIT Allahabad
Dr. Rajitha B., MNNIT Allahabad

Organizer

Department of Electronics and Communication Engineering, Motilal Nehru
National Institute of Technology Allahabad, Prayagraj, India

Preface

We are pleased to introduce the proceedings of the International Conference on VLSI, Communications, and Signal Processing (VCAS 2018) organized by the Department of Electronics and Communication Engineering, MNNIT Allahabad, Prayagraj, India. This conference focuses on recent research works on VLSI, communication systems, signal processing, and related technology. The conference provides a platform for the exchange and dissemination of technical knowledge among the researchers, academicians, and practicing engineers in the field.

VCAS 2018 received a total of 137 submissions. The members of technical program committee took tremendous efforts to review the submitted papers. It was a challenging task to conduct the peer-review process in a very short time. We sincerely thank all the reviewers who spared their valuable time to review the submitted papers. This volume contains 90 papers which are selected from the submitted papers. This conference includes 3 tracks (VLSI, communication, and signal processing), 18 technical sessions, and 2 workshop sessions.

We are grateful to Springer for their technical co-sponsorship and the sponsorship of MNNIT TEQIP-III, Council of Scientific and Industrial Research (CSIR), Tektronix Ltd., and Edutech Ltd., India.

We would like to thank Prof. Rajeev Tripathi, Director, MNNIT Allahabad, for his constant support and guidance for organizing this conference. We gratefully acknowledge the contributions of technical program committee members, track chairs, all other committee members, volunteers, and those persons involved in background activities to make it a success.

Last but not least, we are extremely thankful to all the invited speakers, authors, and members of the advisory committee, technical program committee, and organizing committee; without their support, this conference could not have been successful.

New Delhi, India   Debashis Dutta
Allahabad, India   Haranath Kar
Allahabad, India   Vijaya Bhadauria
Dhanbad, India   Chiranjeev Kumar

Contents

VLSI

Editors and Contributors

About the Editors

Dr. Debashis Dutta obtained his Masters from the Electrical Engineering department of Indian Institute of Technology, Delhi and his PhD in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur. He worked as Scientist G and Group Coordinator (R&D in Electronics Group) in the Ministry of Electronics and Information Technology (MeitY), Government of India. Dr. Dutta has also worked as Programme Director for Special Manpower Development Program (SMDP) in VLSI Design. He was the Director in the board of NICSI, a society under National Informatics Centre and Governing Council member in several R&D institutions. Currently, he works as a consultant for various academic institutions and electronics manufacturing industries. His research interests include low-power analog circuit design, biomedical electronics and neural networks.

Dr. Haranath Kar received his Bachelors from Bengal Engineering College, Shibpur (now IIEST) in 1989, his Masters from Banaras Hindu University, Varanasi in 1992 and his PhD from the University of Allahabad in 2000. After spending a brief period at the Defence Research and Development Organization as Scientist B, he joined Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India, as a Lecturer in 1991, where he became an Assistant Professor in 2001, Associate Professor in2006 and Professor in 2007. He spent two years at the Atilim University, Turkey (2002-2004) as an Assistant Professor. He was Head of Electronics and Communication Engineering Department at MNNIT during 2013-2015. His current research interests include digital signal processing, nonlinear dynamical systems, delayed systems, robust stability, guaranteed cost control, and multidimensional systems. He is a recipient of the 2002-2003 IEE Heaviside

Premium Award. He was conferred with the D.N. Agrawal Award of excellence and the Bharat Vikas Award in 2005 and 2017, respectively. He is also a member of editorial board of the Mathematical Problems in Engineering.

Dr. Chiranjeev Kumar received his Bachelors in Telecommunication Engineering from Banglore University, his Masters in Computer Science, and his PhD from MNNIT Allahabad. Currently, he is a professor in the Department of Computer Science & Engineering, Indian Institute of Technology (IIT-ISM), Dhanbad, India. His research interests include wireless networks, software engineering, mobility management, and cognitive radio networks. He has published more than 100 articles in international journals and conferences proceedings. Dr. Kumar has also contributed several books and book chapters on wireless communication and networking, and he has worked as an ad hoc reviewer of many international journals.

Dr. Vijaya Bhadauria received her Bachelors (Electronics) and Masters (Control & Instrumentation), and her PhD in Electronics Engineering from MNNIT Allahabad. At present, she is Professor and Head in the Department of Electronics & Communication Engineering, MNNIT Allahabad. Her major research interests include VLSI circuits and system, digital integrated circuit design, advanced analog integrated circuit design, VLSI technology and Semiconductor Device and Modeling. She has published many papers in international journals and proceedings, and has also acted as an ad hoc reviewer of many international journals.

Contributors

Abhishek Agarwal Department of Electronics and Communication Engineering, Bundelkhand Institute of Engineering and Technology, Jhansi, India

Lucky Agarwal Department of Electronics & Communication Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India

Komal Agrawal Electrical Engineering Department, MNNIT, Allahabad, India

Mudit Agrawal IIIT-Allahabad, Allahabad, India

Neelesh Agrawal Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

T. Akhil Chandran Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, India

Repala Akhil Department of ECE, National Institute of Technology, Andhra Pradesh, Tadepalligudem, AP, India

Brij Bansh Nath Anchal Department of Ceramic Engineering, Indian Institute of Technology (BHU), Varanasi, India

J. A. Ansari Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

Prateek Asthana Department of Electronics and Communication Engineering, National Institute of Technology, Hamirpur, India

Rachna Asthana Department of Electronics Engineering HBTU, Kanpur, Uttar Pradesh, India

Shubhanker Banerjee Jaypee Institute of Technology, Noida, India

Vijaya Bhaduaria Department of Electronics and Communication Engineering, MNNIT Allahabad, Allahabad, India

Dinesh Bhatia Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Prachin Bhoyar Visvesvaraya National Institute of Technology, Nagpur, India

P. Chakrabarti Department of Electronics & Communication Engineering, Indian Institute of Engineering Sciences and Technolgy, Shibpur, West Bengal, India

Prajjwal Singh Chauhan Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

R. K. Chauhan Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India

Niraj Kumar Choudhary Department of Electrical Engineering, MNNIT Allahabad, Prayagraj, India

Sahil Dalal University School of Information Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India

R. B. Deshmukh Visvesvaraya National Institute of Technology, Nagpur, MH, India

Raghavendra Deshmukh Visvesvaraya National Institute of Technology, Nagpur, India

Amit Dhawan Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India

Sanjay B. Dhok Center of VLSI and Nanotechnology, Visvesvaraya National Institute of Technology, Nagpur, India

Saurabh Dixit Central Institute of Plastics Engineering and Technology, Lucknow, India

Tanmay Dubey Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Apoorva Dwivedi Department of Electronics and Communication Engineering, National Institute of Technology, Hamirpur, India

Som Pal Gangwar Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur, UP, India

Anil Kumar Gautam NERIST Deemed to be University, Nirjuli, Arunachal Pradesh, India

Prateek Raj Gautam Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Shashank Gavel Department of Electronics and Telecommunication, National Institute of Technology, Raipur, India

Girish Chandra Ghivela Electronics and Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur, India

Sougata Ghosh Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Allahabad, Uttar Pradesh, India

S. P. Ghrera Department of CSE, JUIT, Solan, HP, India

Manish Goswami Department of Electronics and Communication Engineering, Indian Institute of Information Technology Allahabad, Allahabad, Uttar Pradesh, India

Akshay Goyal National Institute of Technology, Hamirpur, Hamirpur, India

Mohd. Gulman Siddiqui Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

Abhinav Gupta Electronics Engineering Department, Rajkiya Engineering College Sonbhadra, Sonbhadra, India

Anuradha Gupta Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur, UP, India

Mahesh Kumar Gupta RailTel Corporation of India, Gurugram, India

Mridul Gupta Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

Namrata Gupta National Institute of Technology, Raipur, India

Pushpendra Kumar Gupta Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Sanjay Gupta Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India

Santosh Kumar Gupta Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Deo Chandra Jaiswal Department of Electronics Engineering HBTU, Kanpur, Uttar Pradesh, India

R. K. Jaiswal Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Rashid Jamal Department of Electronics Engg, Institute of Engineering and Technology, Lucknow, India;
Dr. APJ Abdul Kalam Technical University, Lucknow, India

Ravi Kumar Jatoth Department of ECE, National Institute of Technology, Warangal, Warangal, Telangana, India

Rajesh Kumar Jha Department of Electronics and Communication Engineering, Indian Institute of Information Technology Allahabad, Allahabad, Uttar Pradesh, India

Rajib Kumar Jha Indian Institute of Technology Patna, Patna, Bihar, India

Sristi Jha Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Nandkishor Joshi IIIT-Allahabad, Allahabad, India

Pallavi Joshi Department of Electronics and Telecommunication, NIT Raipur, Raipur, India

Bhupendra Jothe Electronics and Telecommunication Engineering Department, Government Polytechnic College, Itarsi, India

V. Krishna Rao Kandanvli Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Mayank Kansal Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

P. Karuppanan Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Kavindra Kumar Kavi Department of ECED, MNNIT Allahabad, Allahabad, India

Gargi Khanna Department of Electronics and Communication Engineering, National Institute of Technology, Hamirpur, India

Priyank Khare Department of Electronics & Communication Engineering, MNNIT, Allahabad, India

Shahneela Jamal Kidwai Institute of Engineering and Technology, Lucknow, India

Amrish Kumar Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, India

Anurag Kumar Bharat Sanchar Nigam Limited, Ballygunge, Kolkata, India

Arvind Kumar Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Ayush Kumar Indian Institute of Technology Patna, Patna, Bihar, India

Basant Kumar Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Prayagraj, U.P., India

Basant Kumar Department of ECE, MNNIT Allahabad, Allahabad, UP, India

Krishan Kumar National Institute of Technology, Hamirpur, Hamirpur, India

Manish Kumar Indian Institute of Information Technology, Allahabad, India

Mukesh Kumar Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

Prashant Kumar Department of Electronics & Communication Engineering, MNNIT, Allahabad, India

Prem Kumar Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Prince Kumar Electronics and Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur, India

Rupesh Kumar IIT, Kanpur, India

Sandeep Kumar Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India

Sumit Kumar Indian Institute of Technology Patna, Patna, Bihar, India

Sunil Kumar Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Sunil Kumar Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Vinay Kumar Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

Ritesh Kumar Kushwaha Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Radhey Lal Council of Science & Technology, Lucknow, UP, India

Lal Chand Mangal Defence Electronics Application Laboratory (DRDO), Dehradun, Uttarakhand, India

G. Manikanta Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

J. B. Maurya Department of Electronics and Communication Engineering, Bundelkhand Institute of Engineering and Technology, Jhansi, Uttar Pradesh, India

Digvijay Singh Mehta Motilal Nehru National Institute of Technology, Allahabad, UP, India

G. R. Mishra Department of Electronics & Communication Engineering, Amity School of Engineering & Technology (ASET), Amity University, Lucknow, Uttar Pradesh, India

R. A. Mishra Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Rishabh Bhooshan Mishra Smart Sensor Area, CSIR-Central Electronics Engineering Research Institute (CEERI), Pilani, Rajasthan, India

Shalini Mishra Qualcomm, Bengaluru, India

Varun Mishra Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Yogesh Kumar Mishra Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur, UP, India

Rajkrishna Mondal Department of Mathematics, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Ravindra Mukhiya Smart Sensor Area, CSIR-Central Electronics Engineering Research Institute (CEERI), Pilani, Rajasthan, India

Vadthiya Narendar Electronics and Communication Engineering Department, National Institute of Technology Warangal, Telangana, India

Alok Naugarhiya National Institute of Technology, Raipur, India

Richa Negi Electrical Engineering Department, MNNIT, Allahabad, India

Ayush Kumar Nigam Department of Electronics & Communication Engineering, MNNIT, Allahabad, India

Sunil Kumar Ojha Department of Electronics & Communication Engineering, Amity School of Engineering & Technology (ASET), Amity University, Lucknow, Uttar Pradesh, India

Raghavendra Pal Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Sarika Pal Department of Electronics Engineering, NIT Uttarakhand, Srinagar, Garhwal, Uttarakhand, India

Hemant Kumar Pande Defence Electronics Application Laboratory (DRDO), Dehradun, Uttarakhand, India

Ashutosh Kumar Pandey Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology, Prayagraj, Uttar Pradesh, India

Richa Parihar Electronics and Communication Engineering Department, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India

Sumit Patra Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Sudip Paul Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Y. K. Prajapati Departement of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, Uttar Pradesh, India

Arun Prakash Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

A. Laxmi Prasanna Visvesvaraya National Institute of Technology, Nagpur, India

Neetesh Purohit IIIT-Allahabad, Allahabad, India

Ram Pyare Department of Ceramic Engineering, Indian Institute of Technology (BHU), Varanasi, India

Ajay Singh Raghuvanshi Department of Electronics and Telecommunication, National Institute of Technology, Raipur, India

Sanjeev Rai Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Meenu Rani Visvesvaraya National Institute of Technology, Nagpur, MH, India

Ranvijay Computer Science and Engineering Department, MNNIT Allahabad, Allahabad, India

Ashish Kumar Rao Department of Electronics Engineering, Institute of Engineering and Technology, Lucknow, India;
Dr. APJ Abdul, Kalam Technical University, Lucknow, India

Tarique Rashid Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Akhilesh Kumar Ravat Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India

Akash Singh Rawat Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, India

Abhishek Ray National Institute of Technology, Raipur, India

Sonali Ray MNNIT, Allahabad, India

Vishal Singh Sachan Computer Science and Engineering Department, MNNIT Allahabad, Allahabad, India

Nitin Sahai Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Pritam Keshari Sahoo Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Sima Sahu Dr. A. P. J. Abdul, Kalam Technical University, Lucknow, Uttar Pradesh, India

J. P. Saini Netaji Subhas Institute of Technology (NSIT), New Delhi, India

S. Santosh Kumar Smart Sensor Area, CSIR-Central Electronics Engineering Research Institute (CEERI), Pilani, Rajasthan, India

A. K. Saroj Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

Joydeep Sengupta Electronics and Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur, India

Ravi Shankar Department of Revenue, Ministry of Finance, Hyderabad, India

Tripurari Sharan NERIST Deemed to be University, Nirjuli, Arunachal Pradesh, India

Meenakshi Sharma MNNIT, Allahabad, India

Shweta Sharma Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur, UP, India

Anubhav Shivhare Indian Institute of Information Technology, Allahabad, India

Prabhat Chandra Shrivastava Department of Electronics & Communication Engineering, MNNIT, Allahabad, India

Anurag Shukla National Institute of Technology, Raipur, India

Devarshi Shukla Department of Electronics and Communication Engineering, MNNIT Allahabad, Allahabad, India

Devesh Shukla Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Vandana Shukla Amity University Uttar Pradesh, Lucknow, India

Rahul Singhal Department of Electrical Engineering, MNNIT Allahabad, Prayagraj, India

A. K. Singh Department of CSE, NIT Patna, Patna, Bihar, India

Akhilendra Pratap Singh National Institute of Technology, Shillong, Meghalaya, India

Amit Kumar Singh Department of Computer Science and Engineering, NIT, Patna, India

Arun Kumar Singh Rajkiya Engineering College, Kannauj, India

B. R. Singh Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India

Brijesh Kumar Singh Department of Electronics & Communication Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India

Ekta Singh Department of Electronics Engg, Institute of Engineering and Technology, Lucknow, India;
Dr. APJ Abdul Kalam Technical University, Lucknow 226021, India

Ghanshyam Singh Department of Electronics and Telecommunication, National Institute of Technology, Raipur, India

Harsh Vikram Singh Department of Electronics, Kamla Nehru Institute of Technology (KNIT), Sultanpur, UttarPradesh, India

Indra Kumar Singh Department of Electronics Engineering, Institute of Engineering & Technology, Lucknow, India;
Dr. APJ Abdul Kalam Technical University, Lucknow, India

Khoirom Johnson Singh National Institute of Technology, Chumukedima, Dimapur, Nagaland, India

Kumari Suniti Singh Department of Electronics Engineering, KNIT Sultanpur, Sultanpur, India

Naveen Singh Department of Electronics and Communication Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India

Ningthoujam Qoonand Singh Motilal Nehru National Institute of Technology, Allahabad, Allahabad, UP, India

Nitin Singh Electrical Engineering Department, MNNIT, Allahabad, India

O. P. Singh Department of Electronics & Communication Engineering, Amity School of Engineering & Technology (ASET), Amity University, Lucknow, Uttar Pradesh, India

Prashant Singh Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India

Preetam Singh Department of Ceramic Engineering, Indian Institute of Technology (BHU), Varanasi, India

Rajesh Kumar Singh Feroze Gandhi Institute of Engineering and Technology, Raebareli, India

Rajiv Kumar Singh Department of Electronics Engineering, Institute of Engineering & Technology, Lucknow, India;
Dr. APJ Abdul Kalam Technical University, Lucknow 226021, India

Richa Singh Department of Electronics & Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India

Vivek Kumar Singh Indian Institute of Information Technology, Allahabad, India

T. K. Sinha Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Ankita Srivastava Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Atul Kumar Srivastava Jaypee Institute of Technology, Noida, India

Gaurav Srivastava Motilal Nehru National Institute of Technology, Allahabad, Prayagraj, India

Kapil Kumar Srivastava Defence Electronics Application Laboratory (DRDO), Dehradun, Uttarakhand, India

N. A. Srivastava Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India

Neelam Srivastava Rajkiya Engineering College, Kannauj, India

Rohini Srivastava Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Prayagraj, U.P., India

Vinay Kumar Srivastava Department of Electronics & Communication Engineering, MNNIT, Allahabad, India

S. Thakur Department of CSE, JUIT, Solan, HP, India

Shriram Thyagarajan Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

Ankita Tiwari Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Devesh Tiwari Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India

Manish Tiwari Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India; ITM College, Gwalior, MP, India

R. K. Tiwari Dr. Ram Manohar Lohia Avadh University, Faizabad, India

Siddharth Tiwari Jaypee Institute of Technology, Noida, India

Sudarshan Tiwari Department of Electronics and Communication, Motilal Nehru National Institute of Technology, Allahabad, India

Divya Tripathi Institute of Engineering and Technology, Lucknow, India

Raghuvendra Pratap Tripathi Department of Biomedical Engineering, School of Technology, North Eastern Hill University, Shillong, Meghalaya, India

Rajeev Tripathi Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Sarsij Tripathi National Institute of Technology, Raipur, India

Shweta Tripathi Department of Electronics & Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India

Yogesh Tripathi Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

D. K. Upadhyay Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Dwarka, New Delhi, India

P. R. Vaya Department of Electrical Engineering, Indian Institute of Technology Madras (IITM), Chennai, India

Vutukuri Venkatesh Department of ECE, National Institute of Technology, Andhra Pradesh, Tadepalligudem, AP, India

Akshay Verma Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Alka Verma Departement of Electronics Engineering, Institute of Engineering and Rural Technology, Allahabad, Uttar Pradesh, India

Prateek Kishor Verma Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Swati Verma Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Yogesh Kumar Verma Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Abhay Vidyarthi ITM College, Gwalior, MP, India

Virendra P. Vishwakarma University School of Information Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India

Subodh Wairya Institute of Engineering and Technology, Lucknow, India

Ajay Kumar Yadav Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Mainejar Yadav Computer Science and Engineering Department, MNNIT Allahabad, Allahabad, India

Pawan K. Yadav Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India

Sadanand Yadav Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

Sharad Kumar Yadav ECE Department, Motilal Nehru National Institute of Technology, Allahabad, Allahabad, UP, India

Sudesh Yadav University School of Information Communication & Technology, Guru Gobind Singh Indraprastha University, Dwarka, New Delhi, India

Balaji Yeswanth Department of ECE, National Institute of Technology, Andhra Pradesh, Tadepalligudem, AP, India

Communication Engineering

BER Performance Evaluation of Different Modulation Techniques for Underwater FSO Communication System

Sadanand Yadav, Vinay Kumar, Sanjay B. Dhok, Gaurav Srivastava, Akhilendra Pratap Singh and Mahesh Kumar Gupta

Abstract Wireless data transmission in the underwater medium is of high importance for applications related to the oil industry, military, and scientific community purposes. In the underwater environment, optical wave propagation experiences scattering effect that produces dispersion and creates inter symbol interference that minimizes the signal-to-noise ratio. This paper evaluates different modulation techniques like Phase Shift Keying (PSK), Differential Phase Shift Keying (DPSK), Pulse Amplitude Modulation (PAM), and Quadrature Amplitude Modulation (QAM) in terms of Bit Error Rate (BER) for four different types of underwater media, viz., pure sea, clean ocean, coastal ocean, and turbid harbor water by assuming the Rayleigh fading channel model and found the best suitable technique for Underwater Free Space Optical (UW-FSO) communication.

S. Yadav · V. Kumar (✉)
Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur 440010, India
e-mail: vinayrel01@gmail.com

S. Yadav
e-mail: sadanand.0501@gmail.com

S. B. Dhok
Center of VLSI and Nanotechnology, Visvesvaraya National Institute of Technology, Nagpur 440010, India
e-mail: sbdhok@ece.vnit.ac.in

G. Srivastava
Motilal Nehru National Institute of Technology, Allahabad, Prayagraj 211004, India
e-mail: gaurav4mnnit@gmail.com

A. P. Singh
National Institute of Technology, Shillong 793003, Meghalaya, India
e-mail: akhilendra.singh@nitm.ac.in

M. K. Gupta
RailTel Corporation of India, Gurugram, India

1   Introduction

Underwater wireless communication has a number of applications such as communication among divers, submarines, unmanned underwater vehicles (UUV), underwater sensors, and ships. Generally, wireless data communication underwater employs acoustic waves because its communication range is high due to low loss in water, however, it exhibits some disadvantages like narrow bandwidth of acoustic communication channel resulting in a low data bit rate. Also, the acoustic wave suffers latency due to low propagation delay of 1500 m/s. So another alternative communication technique that is possible in underwater communication is underwater free space optics (FSO). FSO signal is appropriate to overcome some of the problems related to acoustic communication, especially for coastal, shallow, and freshwater environments. With the use of green/blue light wavelengths, FSO wireless communication is admissible in an aqua medium because green/blue light bears less attenuation under water compared to other colors. The propagation speed of FSO signal under water is much higher than the acoustic wave approximately (2.255×10^8 m/s) and also provides higher bandwidth. But, in comparison with acoustic signals, one disadvantage of FSO wave is that it faces more absorption and scattering in aquatic medium. Hence, characteristics of light signal propagation in underwater medium facing the absorption and scattering effects are crucial to investigate.

2   Motivation and Contribution

Many analytical models for FSO communication under water, either based on theoretical or Monte Carlo approaches have been presented in the literature [1, 2]. These papers are basically dependent on physical models that demand complex computations which do not consider the depth-varying ambient light noise and time-varying optical channels. In [3], the Woods Hole Oceanographic Institution has conducted their optical modem trail on shallow water medium, deep water medium at night, and in daytime situation. In the initial phase of design, they conducted a few preliminary simulations of optical transmissions. In [4–10], the authors focused the prospects and problems of underwater FSO wireless communication for applications in the area of sensor networks. In [4], the author focused on optimal clustering in underwater wireless sensor network (UWSN) using EM, acoustic, and FSO communication and make the system energy efficient. In this paper, we evaluated Bit Error Rate (BER) of different modulation schemes, Phase Shift Keying (PSK), Differential Phase Shift Keying (DPSK), Pulse Amplitude Modulation (PAM), and Quadrature Amplitude Modulation (QAM) in different ocean water media like coastal ocean, turbid harbor
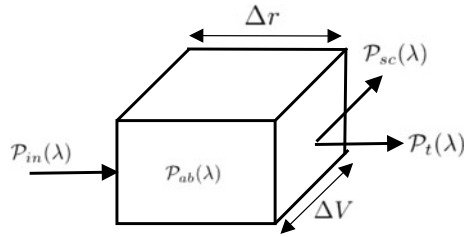
Fig. 1 Geometry of optical property [13]
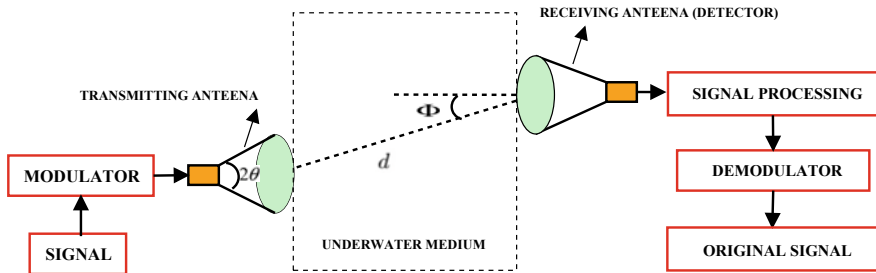
Fig. 2 Underwater optical communication system

water, pure sea, and clean ocean. The BER value of different modulation techniques depends on the signal-to-noise ratio value (SNR) of the signal [11, 12]. It is performed for four types of underwater media assuming the Rayleigh fading channel model. The optical property model in underwater is shown in Fig. 1. The underwater optical communication system with transmitter and receiver is shown in Fig. 2, where transmitter can be LED/Laser-based optical source and receiver can be COTS photo-diode detector.

3 Proposed Work

In underwater medium, absorption and scattering coefficient are the two parameters that create the loss in optical power. For the sake of knowledge of absorption and scattering coefficient, in Fig. 1, a geometrical model of the elemental volume of water ΔV with thickness Δr has been given. When an optical beam of wavelength λ with a power ($\mathcal{P}_{in}$) is transmitted in water, a small amount ($\mathcal{P}_{ab}$), of transmitted light is absorbed by the water, some fraction of power ($\mathcal{P}_{sc}$) are scattered by water and the remaining light power $\mathcal{P}_t$ is passed without any destruction through water. So, by the principle of energy conservation it can be stated as [13, 14]

$$\mathcal{P}_{in}(\lambda) = \mathcal{P}_{ab}(\lambda) + \mathcal{P}_{sc}(\lambda) + \mathcal{P}_t(\lambda) \tag{1}$$

The two terms related to underwater FSO communication is, Absorbance $\mathcal{A}$ and Scatterance $\mathcal{B}$ that can be defined as the ratio of absorbed power to the incident power and the ratio of scattered power to the absorbed power, respectively.

$$\mathcal{A}(\lambda) = \frac{\mathcal{P}_{ab}(\lambda)}{\mathcal{P}_{in}(\lambda)}, \ \mathcal{B}(\lambda) = \frac{\mathcal{P}_{sc}(\lambda)}{\mathcal{P}_{in}(\lambda)} \tag{2}$$

If we assume thickness (Δr) infinitesimally small, the absorption and scattering coefficients can be defined as

$$\alpha(\lambda) = \lim_{\Delta r \to 0} \frac{\mathcal{A}(\lambda)}{\Delta r} = \frac{d\mathcal{A}(\lambda)}{dr} \tag{3}$$

$$\beta(\lambda) = \lim_{\Delta r \to 0} \frac{\mathcal{B}(\lambda)}{\Delta r} = \frac{d\mathcal{B}(\lambda)}{dr} \tag{4}$$

The total attenuation also known as beam attenuation coefficient, due to water can be defined as the linear combination of absorption and scattering coefficients

$$\gamma(\lambda) = \alpha(\lambda) + \beta(\lambda) \tag{5}$$

where $\alpha(\lambda)$ and $\beta(\lambda)$ depend both on the wavelength of light and turbidity.

In Table 1, the value of attenuation coefficients at 520 nm wavelength for turbid harbor water, clean ocean, coastal ocean, and pure sea are given. The optical propagation loss factor, $\mathcal{L}_{olf}(\lambda, d)$ can be expressed as

$$\mathcal{L}_{olf}(\lambda, d) = e^{-\gamma(\lambda)d} \tag{6}$$

In Eq. (6), d represents the distance between the transmitter and receiver in meters and λ represents the wavelength in nanometer (nm). On the basis of the attenuation coefficient, Beer–Lambert law describes the light attenuation effects in underwater environment. They defined that under water, the transmitted optical power loss is an exponentially decaying function of path length d and can be expressed as (Fig. 2)

$$\mathcal{P}_t = \mathcal{P}_0 e^{-\gamma(\lambda)d} \tag{7}$$

where $\mathcal{P}_0$ is the power of the optical source in watt.

Table 1 Attenuation coefficient of different water conditions [13]

Water type	Attenuation coefficient (m^{-1})
Pure seawater	0.043
Clean ocean	0.141
Coastal ocean	0.398
Turbid harbor	2.190

4 Signal-to-Noise Ratio (SNR) Equation

The SNR equation of optical wave in underwater medium can be expressed as [15]

$$SNR = \frac{\mathcal{P}_0 e^{-\gamma(\lambda)d} \mathcal{D}_{rap}^2 \cos(\phi)}{\tan^2(\theta) 4d^2 \mathcal{N}_{eqP}} \tag{8}$$

where all the parameters used in the above equation, their value and meaning is given in Table 2. The $\mathcal{N}_{eqP}$ is considered as a summation of several noise terms, i.e., the solar background shot noise, signal shot noise, dark current shot noise, and preamplifier noise. The first noise terms solar background shot noise which is calculated with the help of upwelling solar radiance is given by [15]

$$\mathcal{L}_{usr} = \frac{\mathcal{E}_{di} \mathcal{R}_{rdi} \mathcal{R}_{fac} exp^{-\mathcal{K}\mathcal{H}}}{\pi} \tag{9}$$

The optical power detector can be calculated as

$$\mathcal{P}_{opd} = \frac{\pi^2 (FOV)^2 \delta \lambda \mathcal{L}_{usr} \mathcal{D}^2}{16} \tag{10}$$

Table 2 Parameters used for computation

Notation	Meaning	Value
$\mathcal{P}_0$	Optical power of the signal	0.1−10 (W)
$\mathcal{H}$	Depth	30 m
θ	Half angle transmitted beam width	5^0-30^0
ϕ	Angle between the optical axis of the receiver and the line-of-sight between the transmitter	30°
$\mathcal{R}_{rdi}$	Underwater reflectance of the downwelling irradiance	1.25%
$\mathcal{E}_{di}$	Downwelling irradiance	1440 (W/M^2)
$\mathcal{L}_{fac}$	Factor describing the directional dependence of the underwater radiance	2.9
$\mathcal{K}$	Diffuse attenuation coefficient	0.2−0.8 m^{-1}
$\mathcal{D}$	Diameter of the collecting optics	20 cm
FOV	Field of view of the system in radians	50, 100
BW	Bandwidth	50 KHz
$\mathcal{S}$	Radiant sensitivity of the detector	85 (A/W)
$\mathcal{I}_{mdc}$	Multiplied dark current	1.5 nA
$\mathcal{I}_{dc}$	Non-multiplied dark current	1.226×10^{-9} A
G_{det}	Detector current gain	1
$\mathcal{I}_{amp}$	Preamplifier current	2 nA

With the help of Eq. (10), solar background shot noise can be calculated as

$$\mathcal{P}_{sbsn} = \frac{\sqrt{2q\mathcal{S}\mathcal{P}_{opd}BW_{eq}\mathcal{F}}}{16} \tag{11}$$

where $\mathcal{F}$ excess noise factor $=1$ for photodiode, >1 for avalanche photodiodes (dimensionless), and BW_{eq} is the effective noise bandwidth, i.e., equal to $\frac{\pi BW}{2}$.

The signal shot noise power is given by the equation [15],

$$P_{ssnp} = \frac{\sqrt{2q\mathcal{S}\mathcal{P}_{ops}BW_{eq}\mathcal{F}}}{\mathcal{S}} \tag{12}$$

where P_{ops} is the optical power of the signal.

The dark current shot noise power can be calculated as

$$P_{dcsn} = \frac{\sqrt{(2q\mathcal{I}_{mdc}G_{det}^2\mathcal{F} + 2q\mathcal{I}_{dc})BW_{eq}}}{\mathcal{S}G_{det}} \tag{13}$$

The preamplifier noise power is given by

$$\mathcal{P}_{amp} = \frac{\mathcal{I}_{amp}\sqrt{BW_{eq}}}{\mathcal{S}G_{det}} \tag{14}$$

where all the parameters used in the above equation, their value and meaning has been given in Table 2. The total noise equivalent power is equal to

$$\mathcal{N}_{eqP} = \sqrt{\mathcal{P}_{ssnp}^2 + P_{sbsn}^2 + P_{dcsn}^2 + P_{amp}^2} \tag{15}$$

5   Results and Discussion

In this work, MATLAB software is used for simulating the result. Figure 3 demonstrates the variation of BER with respect to distance for different modulation schemes, i.e., PSK, DPSK, PAM, and QAM. For pure seawater communication medium, as the distance increases the BER value also increases. QAM modulation provides long communication distance of up to 60 m with the least BER value. PSK provides less communication distance in comparison with QAM but provides more as compared to DPSK. PAM provides the least communication distance among all. Similarly, for clean ocean medium, from Fig. 4, it is observed that QAM provides the maximum communication distance up to 30 m with the least BER value and PAM modulation provides the worst BER with least communication distance. From the above two plots, it can be said that there are more losses in clean ocean water as compared with
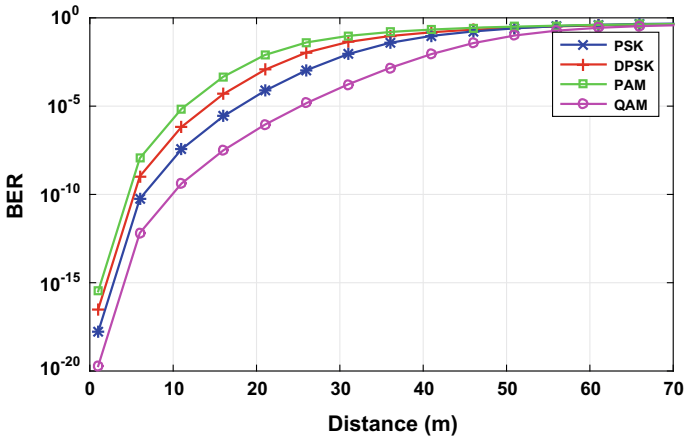
Fig. 3 BER versus distance (between transmitter and receiver) plot when medium is **Pure sea water**
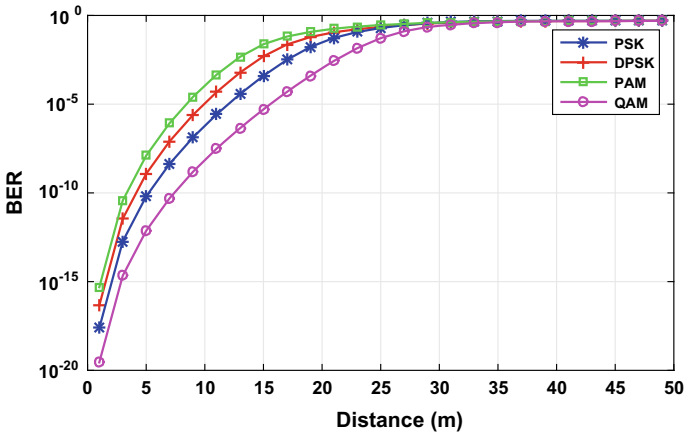
Fig. 4 BER versus distance (between transmitter and receiver) plot when medium is **Clean ocean water**

pure seawater. From Fig. 5, for coastal ocean medium, it is observed that optical wave does not travel more distance due to attenuation and the QAM modulation provides the long communication distance of up to 15 m with the least BER value and PAM provides the least communication distance among others. Lastly, for turbid harbor water medium, from Fig. 6, it is observed that the transmitting distance decreases slightly and QAM modulation technique gives the best transmission distance up to 9 m and PAM provides the worst communication distance with the worst BER. Therefore, it can be inferred that from the above discussion that we can say that the communication is not supportive in turbid harbor water and best communication distance is found in pure seawater with QAM modulation technique.
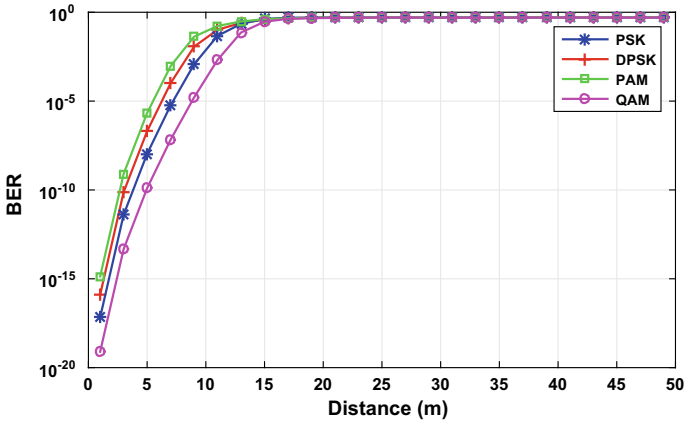
Fig. 5 BER versus distance (between transmitter and receiver) plot when medium is **Coastal ocean water**
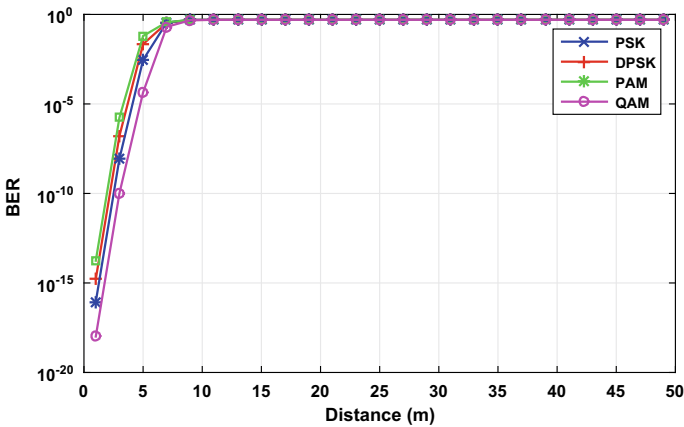
Fig. 6 BER versus distance (between transmitter and receiver) plot when medium is **Turbid harbor water**

6 Conclusion

We investigated the performance of modulation techniques (PSK, DPSK, PAM, QAM) for different sea and ocean water media. It is found that in pure seawater, optical wave communicates through a larger distance due to low attenuation of waves. For clean ocean, coastal ocean, and turbid harbor water media, optical waves attenuate more and the worst communication distance is found for turbid harbor water. The results suggest that QAM modulation technique is the most suitable technique for underwater wireless optical communication for all seawater media.

References

1. Dalgleish, F.R., Shirron, J.J., Rashkin, D., Giddings, T.E., Dalgleish, A.K.V., Cardei, I., Ouyang, B., Caimi, F.M., Cardei, M.: Physical layer simulator for undersea free-space laser communications. Opt. Eng. **53**(5), 1–14 (2014)
2. Anguita, D., Brizzolara, D., Parodi, G., Hu, Q.: Optical wireless underwater communication for AUV: preliminary simulation and experimental results. In: Proceedings of the IEEE/OES Oceans, Santander, Spain, June 2011
3. Farr, N., Chave, A.D., Freitag, L., Preisig, J., White, S.N., Yoerger, D., Sonnichsen, F.: Optical modem technology for seafloor observatories. In: Proceedings of the MTS/IEEE Oceans, Boston, MA, Sept 2006
4. Yadav, S., Kumar, V.: Optimal clustering in underwater wireless sensor networks: acoustic, EM and FSO communication compliant technique. IEEE Access **5**, 12761–12776 (2017)
5. Sandeep, D., Kumar, V.: Review on clustering, coverage and connectivity in underwater wireless sensor networks: a communication techniques perspective. IEEE Access **5**, 11176–11199 (2017)
6. Kumar, V., Bhusari, R., Dhok, S., Prakash, A., Tripathi, R., Tiwari, S.: Design of magnetic induction based energy efficient WSNs for nonconventional media using multilayer transmitter enabled novel energy model. IEEE Syst. J. (Accepted for publication) (2018)
7. Kumar, V., Kumar, V., Sandeep, D., Yadav, S., Barik, R.K., Tripathi, R., Tiwari, S.: Multi-hop communication based optimal clustering in hexagon and voronoi cell structured WSNs. AEU-Int. J. Electron. Commun. **93**, 305–3016 (2018)
8. Tambe, S., Kumar, V., Bhusari, R.: Magnetic induction based cluster optimization in non-conventional WSNs: a cross layer approach. AEU-Int. J. Electron. Commun. **93**, 53–62 (2018)
9. Kumar, V., Yadav, S., Kumar, V., Sengupta, J., Tripathi, R., Tiwari, S.: Optimal clustering in Weibull distributed WSNs based on realistic energy dissipation model. In: Progress in Computing, Analytics and Networking, pp. 61–73. Springer (2018)
10. Kumar, V., Yadav, S., Sengupta, J., Kumar, S., Barik, R.K., Tripathi, R., Tiwari, S.: TMSM-based optimal clustering in a Gaussian distributed wireless sensor network. In: TENCON 2017-IEEE Region 10 Conference, Penang, Malaysia (2017)
11. Rao, K.D.: Channel Coding Techniques for Wireless Communications. Springer (2016)
12. Simon, M.K., Alouini, M.S.: Digital Communication Over Fading Channels, vol. 95. Wiley (2005)
13. Kaushal, H., Kaddoum, G.: Underwater optical wireless communication. IEEE Access **4**, 1518–1547 (2016)
14. Han, S., Noh, Y., Liang, R., Chen, R., Cheng, Y., Gerla, M.: Evaluation of underwater optical-acoustic hybrid network. China Commun. **11**, 49–59 (2014)
15. Giles, J.W., Bankman, I.N.: Underwater optical communications systems. Part 2: Basic design considerations. In: Military Communications Conference, MILCOM IEEE, pp. 1700–1705 (2005)

Reliable Location-Aware Routing Protocol for Urban Vehicular Scenario

Check for
updates

Ankita Srivastava and Arun Prakash

Abstract Vehicular Ad hoc Network is a subcategory of Mobile Ad hoc Network where a wireless communication is established between vehicles and road side unit. However, the frequently changing topology of vehicles and the absence of any centralized administration make routing a challenging task. This paper presents a Reliable Location-Aware Routing Protocol where each vehicle is assumed to be equipped with a Global Positioning System and a digital road map. Therefore, a vehicle can get its location on the road map. If the forwarder vehicle is present at the road intersection, then it starts forwarding data through the road having the highest weight factor. This weight factor is calculated on the basis of distance, direction, and vehicular density on that road. On the other hand, when the vehicle lies between intersections, a two-level process is applied to get a reliable data transmission. This two-level process helps vehicles to forward data without getting stuck in a local optimum situation. The performance of this work has been evaluated using NS 2.34 and is compared with the existing road-aware geographic routing protocol. The simulation shows better results in terms of packet delivery ratio and throughput.

Keywords Geographical · Local optimum · Distance · Direction · Vehicular density · Routing · Reliable

1 Introduction

Vehicular Ad hoc Network (VANET) is a kind of Mobile Ad hoc Network (MANET) which corresponds to the heart of an Intelligent Transportation System (ITS) [1]. VANET allows multi-hop vehicle–to-vehicle (V2V) and vehicle to infrastructure communication (V2I). This technology makes use of IEEE 802.11p, a standard

A. Srivastava (✉) · A. Prakash
Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India
e-mail: rel1708@mnnit.ac.in

A. Prakash
e-mail: arun@mnnit.ac.in

designed specially to support highly dynamic scenarios [2]. To develop an efficient routing protocol for highly mobile vehicles in the network is one of the major challenges. The routing protocols proposed so far can be classified into three main groups: topology-based, clustering-based, and position-based routing protocols [3]. Topology-based protocols need the topological information of the network through beacon messages to make routing decisions. Prior to the data forwarding, the route is formed through broadcasting. Once a complete route is formed, then the data is forwarded through that route. This method is not so effective for VANET as the route formed before forwarding data may broke in high dynamic scenario. In cluster-based routing, a group of nearby vehicles having a similar characteristic such as speed and direction forms cluster. This improves scalability but maintaining a cluster causes excessive overheads in the network. Among the three, position-based routing is most effective for VANET scenario [4] because it allows better long-distance data forwarding and efficient progress toward the destination. They make routing decisions based on the information of the vehicles obtained through Global Positioning System (GPS) such as position, velocity, etc. However, position-based routing protocols still have some drawbacks in an urban area because there is unavailability of the optimal path and the next best hop in the network as there is more signal attenuation and frequent link disconnectivity. Authors in [5] have designed a routing protocol suitable for urban scenario by using a powerful routing metric for data forwarding. It uses distance, direction, and traffic density for selecting the road if the forwarder vehicle is present at the intersection and if it lies between intersections, then the vehicles present in the mid of the transmission range of the forwarder is selected as the next forwarder. However, the solution to the unavailability of the next best hop known as the problem of the local optimum is still a challenging task. The present work proposes a method to avoid this problem. The main contributions of the paper are

1. In the proposed protocol, when the forwarding vehicle is present at the road intersection, then to forward data it selects the road based on its calculated weight factor. The weight value is assigned on the basis of the traffic density on that road, its distance from the destination, and the direction.
2. On the other hand, when the forwarder geographic location is found between the intersections, then the neighbors of the vehicle lying within its radio range is evaluated. The neighbor node which is having the maximum number of neighbors traveling toward the destination is selected to forward the data. This two-level process is useful in avoiding the local optimum situation.
3. The performance of the proposed routing protocol is evaluated using NS 2.34 and is compared with RAGR [5]. The results show its effectiveness in packet delivery ratio and throughput.

The rest of the paper is organized as follows. In Sect. 2, the existing works based on position-based routing are reviewed. Section 3 illustrates the present work. In Sect. 4, analyses and comparison of the present work with RAGR protocol are done. Finally, Sect. 5 concludes the paper with an outline of future work.

2 Related Work

This section describes existing work done in the area of position-based routing. Greedy Perimeter Stateless Routing (GPSR) [6] uses simple greedy mode. In greedy mode, the vehicle which is geometrically close to the destination as compared to the other vehicles within the radio range of the forwarder is selected as the next forwarding vehicle. However, this protocol fails when the path contains more obstacles. Therefore, this method is not recommended in an urban environment.

Geographic Source Routing (GSR) [7] takes advantage of Dijkstra algorithm to discover the shortest path and then forwards the data using greedy mode. GeoSpray protocol [8] works on the concept of spray phase and wait mechanism where a small set of packet copies are sprayed to some fixed but different vehicles in the network. Now, these packet copies get to be forwarded further to other vehicles that can take the packets near the destination. Additionally, to control the overhead there is an upper bound limit to control the duplicate packet flooding in the network. The implementation of this protocol lacks the simulation environment. In Junction-Based Routing (JBR) [9], there is a reduction of an end to end delay. The main reason behind this is the reduction in extra beacons needed for calculating vehicles position. This process uses selective greedy forwarding and minimum angle method to forward the packet.

Road Selection Routing Protocol (RSRP) [10] selects the best routes which are connected across the junction based on their rating. If a road is selected and the forwarder vehicle predicts the occurrence of network gap while transferring the data, then the sufferer vehicle adjusts its velocity in forward or backward direction. Therefore, this protocol is not suitable for city environment because of its prediction mechanism as the VANET topology is unpredictable.

From all the aforementioned protocols, major work is done to reduce end to end delay and improve packet delivery ratio. For position-based routing protocols, data forwarding in a local optimum situation is challenging. However, to avoid a local optimum situation, majority of the protocols use *carry and forward method* which increases the delay. As transmission through wireless media is comparatively much faster than vehicular speed, it has therefore been concluded that the *local optimum* problem in geographic routing is still a challenging task. It occurs when the source vehicle does not find any neighbor vehicle closer to the destination than itself as shown in Fig. 1.

3 Protocol Overview

In this paper, a Reliable Location-Aware Routing Protocol for Urban Vehicular Scenario (RLARP) is proposed to solve the problem of the local optimum in urban areas. This protocol works in two modes: The first is when the vehicle is found in the middle of the intersection and the second is when the vehicle is found at the intersection. In

Fig. 1 Local optimum
problem
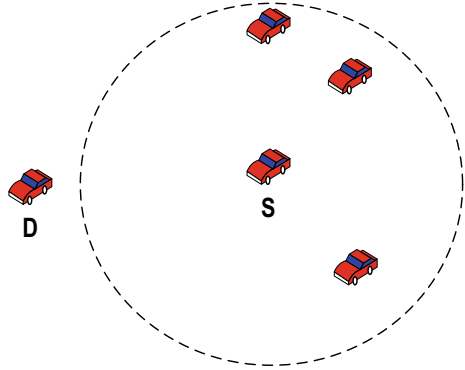
the middle of the intersection, the forwarder selection process is executed and at the intersection, the road selection process is done.

Besides this, some assumptions are made, such as each vehicle must be equipped with a GPS for computing the location of the source, destination, and the neighbors up to two hops. Beacon messages assist this location information to propagate in the network. Additionally, each vehicle must contain a digital map of the street with the help of which every vehicle gets the location of the intersection [5].

3.1 Forwarder Selection Process: A Two-Level Process

The forwarder selection process takes the help of direction metrics and location information of the neighbor up to two hops (level) to find the next forwarder. In a simple greedy approach [6], the vehicle which is closer to the destination than the source itself is selected as a forwarder. This method has a disadvantage that generally for highly mobile vehicles, the forwarder located at the edge of transmission range gets out of range till the data packet reaches it. Whereas, when the vehicle with minimum distance is selected as a forwarder, it leads to high interference of packets in the network. Road-Aware Geographic Routing (RAGR) [5] selects mid area forwarder, but it fails to solve the local optimum problem completely. In RAGR, when a source vehicle is not successful to track any mid area vehicle, then that vehicle waits until it finds any forwarder. This method could not resolve the local optimum situation completely because it could lead to an unbearable delay in low-density areas.

RLARP addresses this issue by finding the neighbors of the vehicles which are in the range of the current vehicle. This two-level process helps in the prediction of the network topology in advance. In the forwarder selection process, the below-mentioned steps are followed:

Step (1) The neighbors of the source vehicle lying within the semicircle of transmission range and traveling toward the destination or the next intersection is taken.
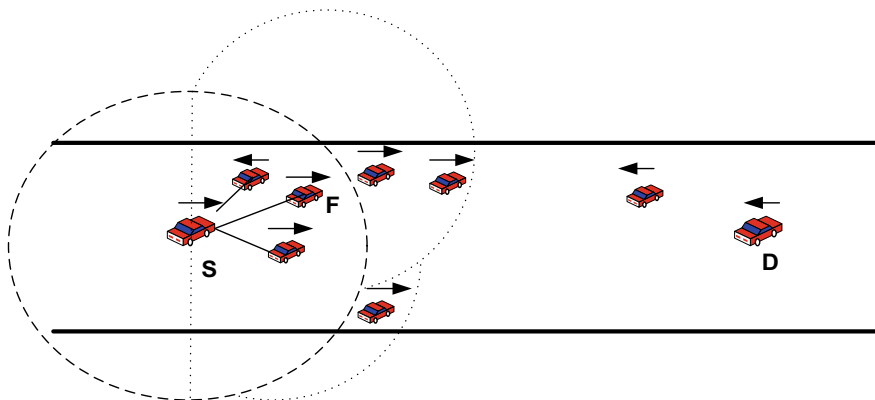
Fig. 2 Forwarder selection process when the source and destination are on the same path

The semicircle is toward the area which is closest to the destination or the neighbor intersection as shown in Fig. 2.

Step (2) Second, neighbors of those above found vehicles are examined. Among them, the vehicle which is having maximum neighbors moving toward the destination (when the destination lies on the same road) or neighbor intersection (when the destination lies on the different road) within 300 m of its range is selected. The flow chart of RLARP is shown in Fig. 3.

3.2 Road Selection Process

As a data forwarding vehicle reaches the intersection, the protocol selects the road on the basis of distance, direction, and traffic density. Based on these three parameters, the weight factor is assigned to each vehicle lying between the current intersection and the neighbor intersections. Through a digital road map, the candidate vehicle is capable of detecting the position of neighbor intersections. The vehicle with the highest weight value is then elected as the next forwarder. When this forwarder on any of the neighbor routes is selected, then the protocol moves to the first mode.

First, the distance between the geographical position of each neighbor vehicle and the destination is calculated. The candidate vehicle having minimum calculated distance must be given the highest weight value.

$$WF_{dis} = \frac{1}{dis(i, d) + 0.001} \tag{1}$$

In Eq. (1), $dis(i, d)$ represents the Euclidean distance between each neighbor vehicle and the destination. Second, the weight value according to the direction is given. The data must be forwarded to the vehicle which travels toward the next
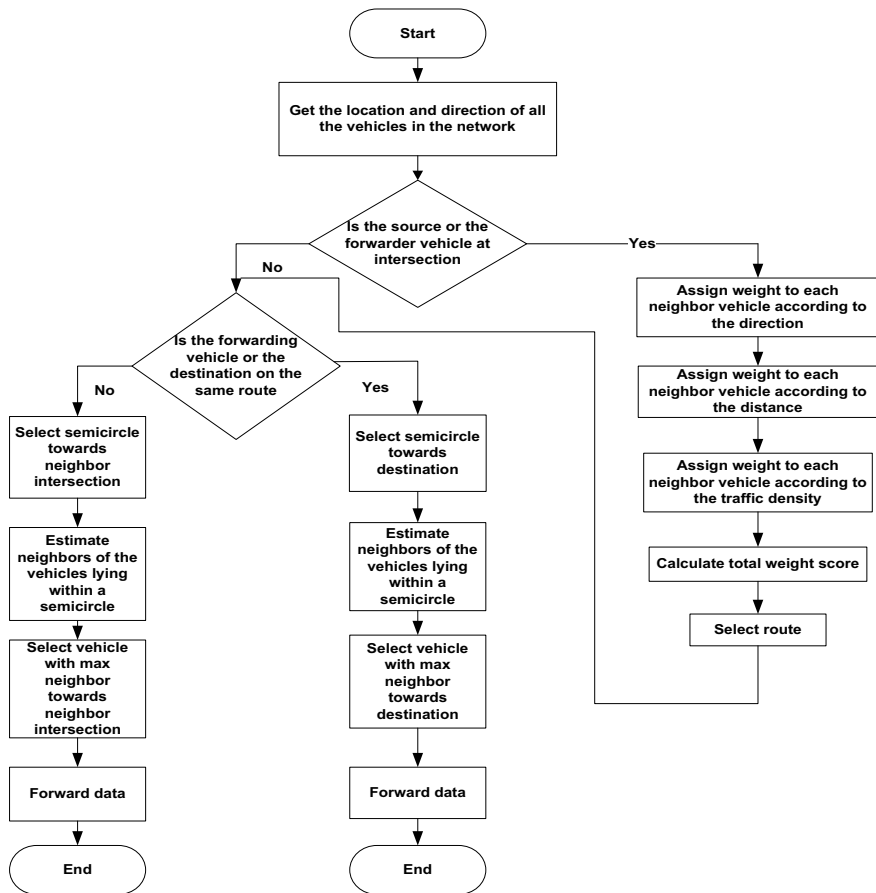
Fig. 3 Flowchart of RLARP

intersection. If WF_{dir} denotes the direction weight factor, then,

$$WF_{dir} = \left\{ \begin{array}{l} 1 \text{ node towards the next intersection} \\ 0 \text{ node opposite to the next intersection} \end{array} \right\} \tag{2}$$

At last, the real-time traffic density is estimated for each road linked with the current intersection. In an event, if a path contains no vehicles, then that path would be given the lowest priority.

$$WF_{td} = \frac{N_{road}}{N_{avg}} \tag{3}$$

Here, WF_{td} denotes the weighting value according to the vehicular density. N_{road} represents the density of the vehicle between the current intersection and neighbor intersection on which the vehicle lies. N_{avg} represents the average of vehicles lying between the current intersection and all the neighbor intersections. On the basis of the three parameters calculated above for each vehicle, the total score for each vehicle is calculated by Eq. (4). The vehicle with the highest value is the next forwarder.

$$Score = WF_{dis} * WF_{dir} * WF_{td} \qquad (4)$$

4 Simulation Results

The simulation parameters and the environment in which this protocol has been implemented are demonstrated here. The part of road scenario is shown in Fig. 4. There are nine intersections and 12 interconnecting roads. Performance of the proposed protocol is analyzed by using network simulator (NS) 2.34. Table 1 shows other important parameters used in the simulation.

4.1 Packet Delivery Ratio (PDR)

It is defined as the ratio of data received to the total data sent. A number of experiments are carried out on the proposed protocol by varying vehicles in the network from 50 to 120. In RLARP, the two-level clarity of the network guaranteed more reliable communication as compared to RAGR. Therefore, as shown in Fig. 5, the packet delivery ratio improves since the method prevents the loss of packet in the case of local optimum and value of PDR is almost linear even when the number of vehicles in the network changes.

Fig. 4 Part of road scenario
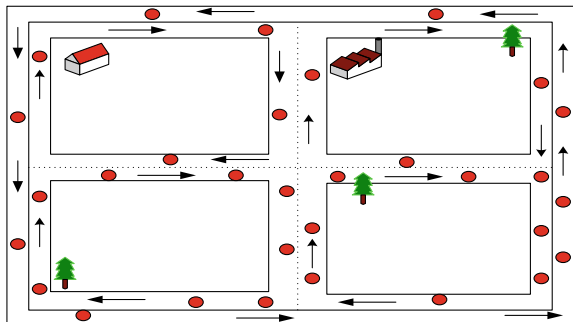
Table 1 Simulation parameters

Parameter	Settings
Simulator name	Network simulator 2.34
Packet sending rate	1 packet/s
Map size	$2000 \times 2000 \text{ m}^2$
Transmission range	300 m
Number of vehicles	50–120
MAC 1	802.11p
Data packet size	511 bytes
Simulation time	50 s
Speed	30–60 m/s
Propagation model	Nakagami radio propagation model

4.2 Throughput

Throughput is defined as the rate at which the data is transmitted in the network. It depends mainly on the packet delivery ratio and the simulation time. From Fig. 6, as the vehicles in the network increase from 60 to 120, it is better when compared with RAGR.

Fig. 5 Packet delivery ratio of RAGR and RLARP
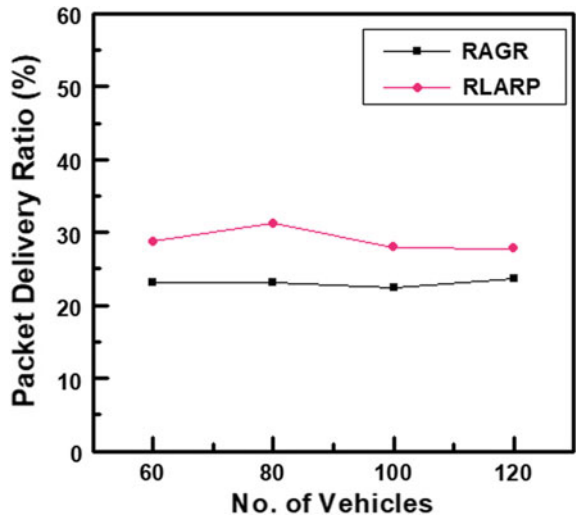
Fig. 6 Throughput of
RAGR and RLARP
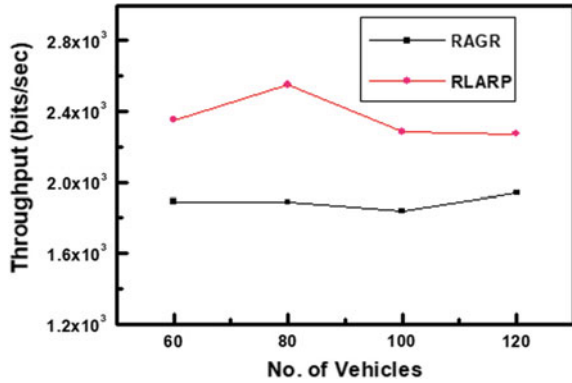
4.3 End to End Delay

As shown in Fig. 7, when there is a smaller vehicular density on the road, the delay in the case of RAGR is more. This is because RAGR protocol waits for a random time until it gets any forwarder. This increases the delay in the local optimum situation which generally arises when there is lesser vehicular density. When the number of vehicles on the road increases, then delay in RLARP starts increasing. However, this difference in the delay is much less and is because it checks the network status up to two levels to provide more reliable communication.

Fig. 7 End to end delay of
RAGR and RLARP
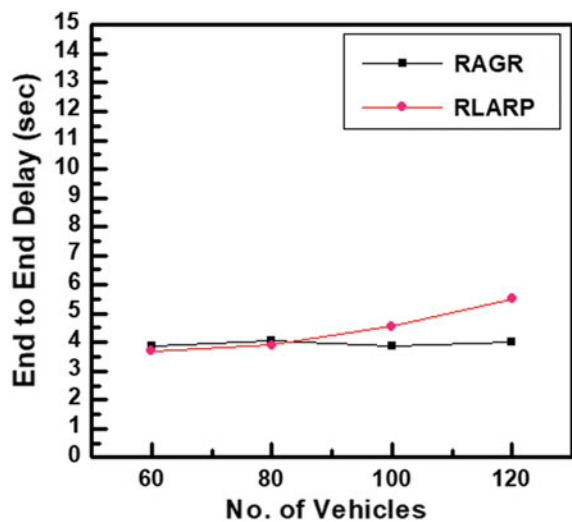
5 Conclusion

In this paper, a Reliable Location-Aware Routing Protocol for Urban Vehicular Scenario (RLARP) has been proposed which mainly aims to reduce the local optimum problem that is one of the major problems of position-based routing when the vehicular density is uneven or low. The process of assuring the neighbor of the vehicles which are in the range of the current vehicle before forwarding data results into a trustworthy communication. It is an effective step taken at the earlier stage before getting stuck into a situation which causes packet loss. Based on the simulation results, RLARP significantly improves the packet delivery ratio and throughput. The obtained result justifies that the proposed protocol is better than RAGR in the realistic vehicular scenario. In future, delay can also be reduced more by applying some optimization technique.

References

1. Singh, S., Agrawal, S.: VANET routing protocols: issues and challenges. In: 2014 Recent Advances, Engineering and Computational Sciences (RAECS), pp. 1–5 (2014)
2. Kenney, J.B.: Dedicated short-range communications (DSRC) standards in the United States. In: Proceedings of the IEEE, pp. 1162–1182 (2011)
3. Dua, A., Kumar, N., Bawa, S.: A systematic review on routing protocols for vehicular ad hoc networks. Veh. Commun. **1**(1), 33–52 (2014)
4. Boussoufa-Lahlah, S., Semchedine, F., Bouallouche-Medjkoune, L.: Geographic routing protocols for vehicular adhoc networks (VANETs): a survey. Veh. Commun. (2018)
5. Qureshi, K.N., Abdullah, A.H., Altameem, A.: Road aware geographical routing protocol coupled with distance, direction and traffic density metrics for urban vehicular adhoc networks. Wirel. Pers. Commun. **92**(3), 1251–1270 (2017)
6. Karp, B., Kung, H.T.: GPSR: greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 243–254 (2000)
7. Lochert, C., Hartenstein, H., Tian, J., Fussler, H., Hermann, D., Mauve, M.: A routing strategy for vehicular ad hoc networks in city environments. In: Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 156–161 (2003)
8. Soares, V.N.G.J., Rodrigues, J.J.P.C., Farahmand, F.: GeoSpray: a geographic routing protocol for vehicular delay-tolerant networks. Inf. Fusion **15**, 102–113 (2014)
9. Tsiachris, S., Koltsidas, G., Pavlidou, F.N.: Junction based geographic routing algorithm for vehicular ad hoc networks. Wirel. Pers. Commun. **71**(2), 955–973 (2013)
10. Bhoi, S.K., Khilar, P.M.: A road selection based routing protocol for vehicular ad hoc network. Wirel. 0Pers. Commun. **83**(4), 2463–2483 (2015)

DFT Precoder Technique Combined with μ-Law Companding for PAPR Reduction in OFDM System

Ajay Kumar Yadav, Pritam Keshari Sahoo and Y. K. Prajapati

Abstract Transmitting high data rate is one of the benefits of the orthogonal frequency division multiplexing (OFDM). However, it undergoes high Peak-to-Average Power Ratio (PAPR) which degrades the system performance. To overcome high PAPR of the OFDM system, the μ-law companding technique is applied. This technique expands only small signals and is not able to compress the high peaks effectively. As a result, average power increases. In this paper, Discrete Fourier Transform (DFT) precoding technique is employed to the μ-law-based OFDM system to reduce further PAPR by converting a multicarrier OFDM into a single-carrier OFDM system. Moreover, it not only reduces PAPR but also minimizes the system complexity because it is a linear technique. Simulation results show how the DFT precoder-based μ-law companding technique realizes reduced PAPR and improved Bit Error Rate (BER) performance in comparison to Discrete Hartley Transform (DHT) precoding with a μ-law companded-based OFDM system.

1 Introduction

Orthogonal frequency division multiplexing (OFDM) is well-known for high-speed digital communications such as digital audio broadcasting (DAB), high definition television (HDTV), digital video broadcasting (DVB), 4G wireless communication,

A. K. Yadav (✉) · P. K. Sahoo · Y. K. Prajapati
Electronics and Communication Engineering Department,
Motilal Nehru National Institute of Technology Allahabad, Allahabad, India
e-mail: er.ajayyadavec@gmail.com

P. K. Sahoo
e-mail: sahoo.pritamkeshari@gmail.com

Y. K. Prajapati
e-mail: yogendrapra@gmail.com

and asymmetric digital subscriber line (ADSL) because of robustness to severe multipath fading and narrowband interference [1].

However, typical OFDM signals still contain very high Peak-to-Average Power Ratio (PAPR) which needs high power amplifier (HPA) at transmitter side to operate nonlinear regions. As a result, nonlinear signal distortion occurs and leads to poor system performance [2]. In order to reduce PAPR of OFDM signals, many techniques have been already proposed in the past, such as partial transmit sequence (PTS) technique [3], selective mapping (SLM) technique [4], clipping and filtering [5], precoding [6, 7], and companding technique [1, 8, 9]. Furthermore, the μ-law companding technique [9] can exhibit less PAPR than the clipping scheme but compressed signals still exhibit nonuniform distribution [1] in comparison to original OFDM signals.

Companding may cause in-band distortion due to its nonlinear process which may degrade BER performance. Some authors have proposed a hybrid method that combined precoding and companding [10–12]. In Ref. [12], authors proposed DHT precoding and rooted-based μ-law companding of an OFDM system for PAPR reduction and maintaining BER.

In this paper, a technique proposed based on the DFT precoder which combined with μ-law companded OFDM signals. The μ-law companding reduces PAPR but not up to a significant level. Also, it is responsible for amplitude distortion. When the DFT precoder is combined with μ-law, PAPR of OFDM signals reduces up to a significant level with low complexity by the converting multicarrier OFDM into single-carrier OFDM systems which have already low PAPR. Moreover, BER is maintained because the DFT precoder is responsible for phase distortion.

This paper is prepared as follows: Problem formulation is described in Sect. 2. Section 3 presents proposed PAPR reduction and system model. Simulation results and discussion are provided in Sect. 4. Ultimately, Sect. 5 describes the conclusion of this paper.

2 Problem Formulation

The baseband OFDM symbol $x(n)$ in discrete form can be obtained with N number of subcarriers by

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi \frac{kn}{LN}}, \quad 0 \le n \le LN - 1 \tag{1}$$

where X_k and L denote modulated data symbols of kth subcarrier and oversampling factor, respectively. Meanwhile, at baseband OFDM signals, the oversampling factor $(L \ge 4)$ gives approximately the true PAPR [8]. In this paper, oversampling has been done by zero padding, i.e., inserting $(L - 1) . N$ zeros to the middle of X to make $\{X_0, X_1, X_2, \ldots 0, 0, \ldots, 0, X_{LN-1}\}$ and LN size of IFFT. The input modulated data

symbols X_k are assumed to be statistically independent and identically distributed random variables. Assume that $Re\{x(n)\}$ and $Im\{x(n)\}$ are real and imaginary part of the OFDM symbols $x(n)$. Therefore, the expression of the amplitude of $x(n)$ can be written as

$$|x(n)| = \sqrt{Re^2\{x(n)\} + Im^2\{x(n)\}}. \tag{2}$$

and according to central limit theorem for large subcarriers (generally $N \geq 64$ [13]), these real as well as imaginary parts of $x(n)$ follow the Gaussian random variables with zero mean and the same variance $\sigma^2 = E[|X_k|^2]/2$, where $E[.]$ is nothing but mathematical expectation and $|.|$ represents modulus. Furthermore, in Ref. [14], the amplitude $|x(n)|$ can be defined using the cumulative distribution function (CDF) of a Rayleigh distribution. So, the CDF of random variable of $|x(n)|$ is given by

$$\begin{aligned} F_{|x(n)|}(x) &= Prob[|x(n)| \leq x] \\ &= \int_0^x \frac{2z}{\sigma^2} e^{\left(\frac{-z^2}{\sigma^2}\right)} dz \\ &= 1 - e^{\left(\frac{-x^2}{\sigma^2}\right)}, \ x \geq 0. \end{aligned} \tag{3}$$

The power of OFDM signals $x(n)$ can be evaluated by

$$|x(n)|^2 = \frac{1}{N} \sum_{J=0}^{N-1} \sum_{k=0}^{N-1} X_j X_k e^{\left(\frac{j2\pi(j-k)n}{LN}\right)} \tag{4}$$

Therefore, PAPR of the OFDM signals $x(n)$ is defined as

$$PAPR = 10 \ \log_{10} \frac{Max[|x(n)|^2]}{E[|x(n)|^2]} \text{dB} \tag{5}$$

The phase of N number of modulated symbols is the same. When it is added, peaks power arises. Therefore, OFDM signals are known to be of high PAPR. Generally, complementary cumulative distribution function (CCDF) measures the performance of PAPR and it can be defined as the probability that the PAPR of the signals greater than the predefined threshold. Let $PAPR_0$ be the threshold value, then the expression of CCDF is as follows:

$$CCDF = \Pr(PAPR > PAPR_0) \tag{6}$$

3 Proposed PAPR Reduction and System Model

3.1 Discrete Fourier Transform (DFT) and DFT Matrix

For a baseband OFDM system, each OFDM symbol contains N number of subcarriers which are orthogonal to each other. Consider a block of N data symbols $X = \{X_k, k = 0, 1, \ldots, N-1\}$. Figure 1 displays the block diagram of the proposed system model. The precoding matrix S of size $(N \times N)$ is implemented before IFFT block and the precoding matrix S is created based on Discrete Fourier Transform (DFT) [15] and it can be written as

$$S = \frac{1}{\sqrt{N}} \begin{bmatrix} S_{0,0} & S_{0,1} & \ldots & S_{0,(N-1)} \\ S_{1,0} & S_{1,1} & \ldots & S_{1,(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ S_{(N-1),0} & S_{(N-1),1} & \ldots & S_{(N-1),(N-1)} \end{bmatrix} \tag{7}$$

Discrete Fourier Transform $X(k)$ can be defined for input signal x_n as

$$X(k) = \sum_{n=0}^{N-1} x_n \cdot e^{-j2\pi nk/N}, \quad k = 0, 1, 2, \ldots, N-1 \tag{8}$$
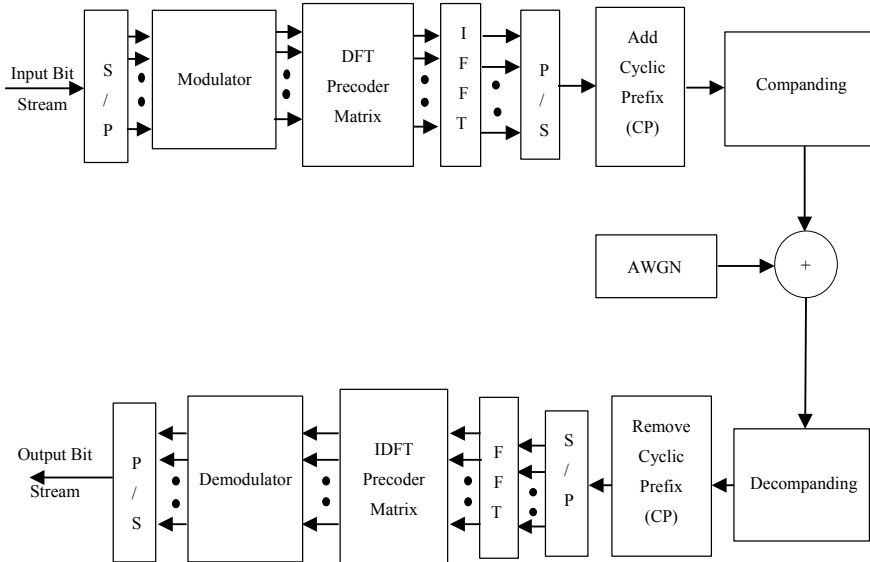
Fig. 1 The block diagram of proposed system model

Inverse DFT x_n can also be written as

$$x_n = \frac{1}{N} \sum_{n=0}^{N-1} X(k) \cdot e^{j2\pi nk/N}, k = 0, 1, 2, \ldots, N-1 \tag{9}$$

The above DFT precoding matrix S can be created by the following equation:

$$S_{l,k} = e^{-j2\pi lk/N} \tag{10}$$

$S_{l,k}$ represents lth row and kth column of DFT precoding matrix.

3.2 DFT Precoder Matrix Based OFDM System

In this section, DFT precoding matrix is implemented to the constellation symbols, i.e., $X_k = \begin{bmatrix} X_0, X_1, X_2, \ldots, X_{N-1} \end{bmatrix}^T$. The length of the DFT precoder matrix is considered the same as the length of the IFFT block which is used for spreading code. Therefore, the original OFDM becomes a single-carrier system and it has low PAPR in comparison to a multicarrier OFDM system. In the DFT precoding matrix based OFDM systems, input data stream is passed through a serial to parallel (S/P) convertor. Then, the parallel data symbol is modulated by subcarriers to generate data symbols of size N which can be written as

$$X_k = \begin{bmatrix} X_0, X_1, X_2, \ldots, X_{N-1} \end{bmatrix}^T \tag{11}$$

The DFT precoding matrix S of size $(N \times N)$ which is created by Eq. (10) is applied to these data symbols to get new data symbols of the same size N that can be expressed as

$$Y\begin{bmatrix} Y_0, Y_1, Y_2, \ldots, Y_{N-1} \end{bmatrix}^T = SX \tag{12}$$

Therefore, the DFT precoded symbols Y_l can be written as

$$Y_l = \sum_{k=0}^{N-1} S_{l,k} \cdot X_k \quad l = 0, 1, \ldots, N-1 \tag{13}$$

Let $S(k)$ is taken as the frequency domain of DFT precoded symbols with oversampled and can be written as:

$$S(k) = [Y_0, Y_1, \ldots Y_{LN-1}] \tag{14}$$

Using the DFT precoder matrix, a set of these frequency domain symbols $S(k)$ will be spread in phase. Therefore, PAPR of an OFDM signal is turned down. At the transmitter side, these frequency domain symbols can be realized into discrete time-domain signals by using Inverse Fourier Transform (IFFT). Hence, OFDM signals can be written as in discrete time domain:

$$S_p(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S(k) e^{j2\pi \frac{kn}{LN}}, \quad \forall n, k = 0, 1, \ldots, LN - 1 \tag{15}$$

3.3 Companded OFDM Signals

From Eq. (15), the signal $S_p(n)$ represents precoded OFDM signals, and PAPR of these signals is minimized up to the single-carrier system because of spreading input data symbols [16]. Recall that the phase of the signals is distorted by the DFT precoder. Without increasing the system complexity, the precoding technique efficiently reduces PAPR of an OFDM system because it is a simple linear technique. However, the amplitude is distorted by companding. When the DFT precoding technique is combined with companding, PAPR of OFDM signals reduces effectively. In this section, we discuss the companding technique which is used for further PAPR reduction. Therefore, after the cyclic prefix (CP) is added, the resulting DFT precoded OFDM signal $S_p(n)$ is companded using μ-law companding transform [9] at the transmitter side and it can be expressed as

$$S_c(n) = F\{S_p(n)\} = A * \text{Sgn}\big[S_p(n)\big] \frac{\log_e\Big[1 + \mu \Big| \frac{S_p(n)}{A} \Big|\Big]}{\log_e[1 + \mu]},$$
$$\forall n = 0, 1, 2, \ldots, LN - 1 \tag{16}$$

where F{.} represents companding transform that is done to the time-domain precoded OFDM signals. However, μ and A are the degree of companding and peak amplitude of the OFDM signals, respectively.

3.4 Receiver Section

From Fig. 1, in order to minimize inter symbol interference (ISI), cyclic prefix (CP) is used. In this case, the companded signals $S_c(n)$ are being sent into a wireless channel. Then, the received signals $y(n)$ can be written as

$$y(n) = S_c(n) + w(n), \forall n = 0, 1, 2, \ldots, LN - 1 \tag{17}$$

where $w(n)$ is supposed to be additive white Gaussian noise (AWGN). At the receiver side, the companded signals are expanded by using μ-law decompanding function which is shown by following expression:

$$r(n) = \mathrm{F}^{-1}\{y(n)\} = \mathrm{A} * \mathrm{Sgn}\big[y(n)\big] * \frac{1}{\mu}\left[e^{\frac{|y(n)\cdot\log(1+\mu)|}{A}} - 1\right]$$

$$\forall n = 0, 1, 2, \ldots, LN - 1 \tag{18}$$

where F^{-1} denotes the decompanding transform. After removing cyclic prefix (CP), the FFT operation is performed to get signal back into Fourier-frequency domain. Then, frequency domain signals are down-sampled. Next, to get back the original transmitted symbols, we implement the inverse DFT matrix.

4 Simulation Results and Discussion

In this section, we executed MATLAB simulation to evaluate the performance of the proposed system model which is based on DFT precoding and μ-law companding. To show the PAPR and BER analysis of the proposed system model, random bits are generated. Then modulated by different types of modulators such as QPSK, 16-QAM, and 64-QAM. The simulation results are compared with the original OFDM, precoded OFDM, μ-law companded OFDM, and combined precoded and μ-law companded OFDM system and all the taken parameters are shown for simulation in Table 1.

Figure 2 illustrates the envelope of the proposed (DFT precoder with μ-law companding, $\mu = 30$) OFDM signals, the original OFDM signals, compressed signals, and precoded based OFDM signals. As we can observed that our proposed technique has reduced the peaks of the power in comparison to the original OFDM signals, μ-law companded signals, precoded based OFDM signals, and DHT precoded with companded signals as results PAPR reduces.

The performances of PAPR versus CCDF are shown in Figs. 3, 4, and 5 for QPSK, 16-QAM, and 64-QAM modulator, respectively. In Fig. 3, the proposed technique is

Table 1 Simulation parameters

Parameter	Details
Simulation tool	MATLAB
Sub-carries	256
Modulator	QPSK, 16-QAM, 64-QAM
Number of cyclic prefix (CP)	32
Oversampling factor (L)	4
FFT, size	1024
Channel	AWGN channel
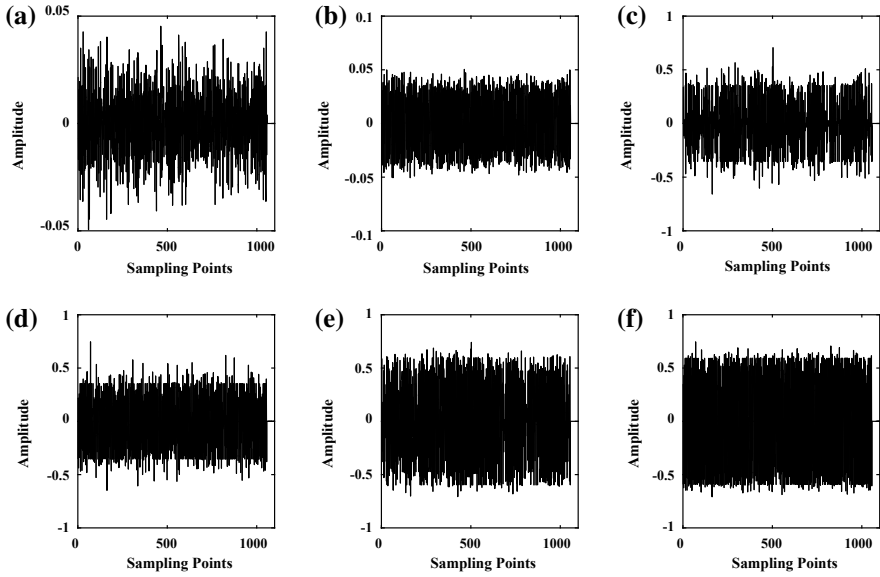
Fig. 2 Envelope waveforms of the OFDM signals. **a** Original OFDM signals. **b** Compressed by μ-law. **c** DHT-OFDM signals. **d** DFT-OFDM signals. **e** DHT with μ-law ($\mu = 30$) companded OFDM signals. **f** Proposed OFDM signals
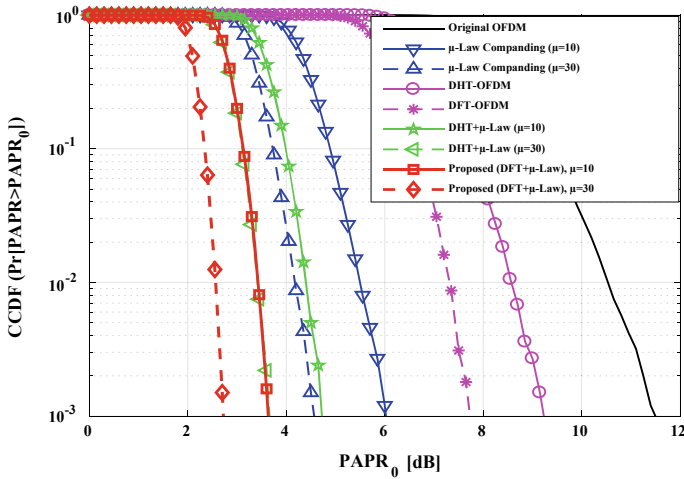
Fig. 3 The CCDFs of proposed OFDM signals, precoded signals, companded signals, precoded with companded signals and original OFDM signals (QPSK)
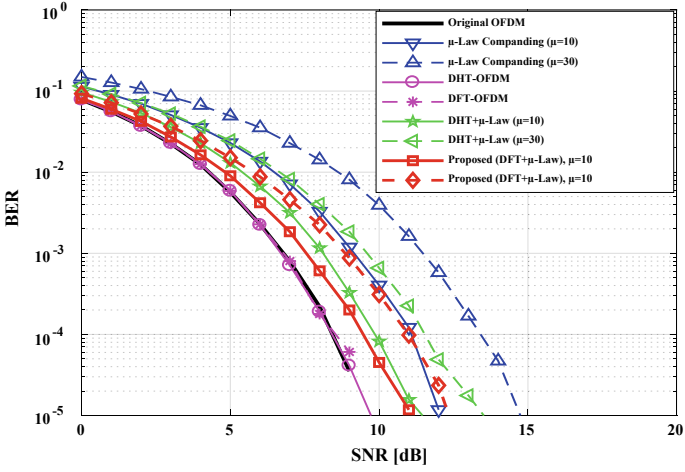
Fig. 4 The BER of proposed OFDM signals, precoded signals, companded signals, precoded with companded signals and original OFDM signals in AWGN channel (QPSK)
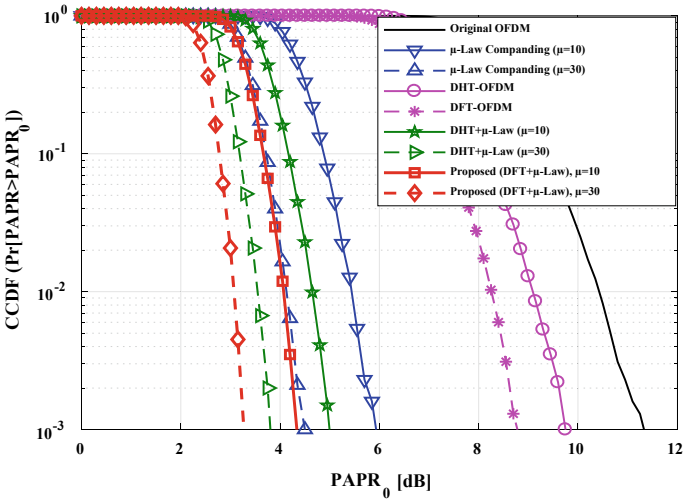
Fig. 5 The CCDFs of proposed OFDM signals, precoded signals, companded signals, precoded with companded signals and original OFDM signals (16-QAM)

compared with other PAPR reduction techniques such as μ-law companding, DHT-OFDM, DFT-OFDM, and DHT + μ-law. At CCDF = 10^{-3}, clearly, the proposed technique ($\mu = 30$) is better than μ-law, DHT-OFDM, DFT-OFDM, DHT + μ-law ($\mu = 30$), and the original OFDM by 1.8 dB, 6.51 dB, 5.02 dB, 0.95, and 8.8 dB, respectively.

Figure 4 illustrates the performance of BER in comparison to the proposed OFDM signals, precoded signals, compressed signals by μ-law, and precoded with compressed signals for QPSK. At BER = 10^{-5}, the proposed technique is improved over μ-law ($\mu = 30$) and DHT-μ-law ($\mu = 30$) in terms of BER performance.

In Fig. 5, PAPR of the proposed technique ($\mu = 30$) is 0.54 dB superior to DHT-μ-law ($\mu = 30$) for 16-QAM. Also, BER performance of the proposed technique ($\mu = 30$) is improved over μ-law and DHT-μ-law ($\mu = 30$) which is illustrated in Fig. 6.

Figure 7 shows the performance of PAPR in comparison to the proposed technique, precoded OFDM signals, compressed signals by μ-law, and precoded with compressed signals for 64-QAM. As we can see, the proposed technique ($\mu = 30$) is 0.45 dB better than DHT-μ-law ($\mu = 30$). The comparison of CCDF and BER of the proposed technique with other techniques is given in Table 2.
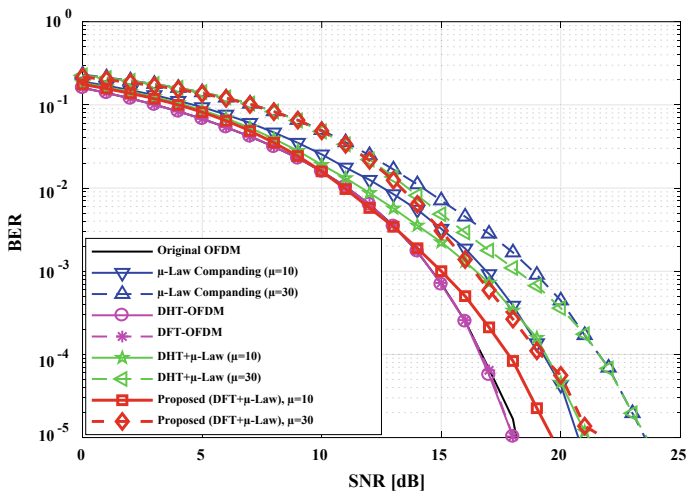
Fig. 6 The BER of proposed OFDM signals, precoded signals, companded signals, precoded with companded signals and original OFDM signals in AWGN channel (16-QAM)
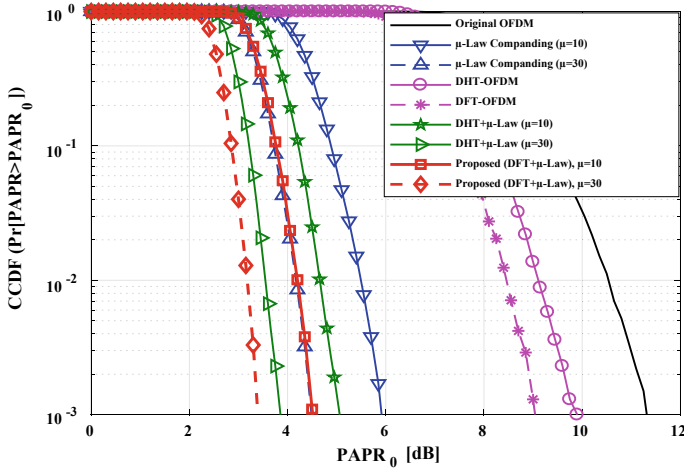
Fig. 7 The CCDFs of proposed OFDM signals, precoded signals, companded signals, precoded with companded signals and original OFDM signals (64-QAM)

5   Conclusion

As far as PAPR reduction is concerned, the proposed technique is based on a combination of the DFT precoder and μ-law companding in OFDM system. By employing this technique (DFT precoder with μ-law), the performance of BER is reduced in comparison to μ-law, and the DHT precoder with μ-law schemes for AWGN channel. Also, the computational complexity decreases due to linearity of the DFT precoder. By the DFT precoder approach, a μ-law companded OFDM system converts into signals carrier system which has low PAPR. We can see the simulation results; the proposed technique provides improved PAPR and BER performance in comparison to other PAPR reduction techniques.

Table 2 Comparison table of PAPR and BER performance

Modulator	Transmitter side $PAPR_0$ (dB) at CCDF $= 10^{-3}$						Receiver side SNR [dB] at BER $= 10^{-5}$					
	Proposed (DFT + μ-Law, μ = 30)	DHT + μ-Law, μ = 30	μ-Law, μ = 30	DFT-OFDM	DHT-OFDM	Original OFDM	Proposed (DFT + μ-Law, μ = 30)	DHT + μ-Law, μ = 30	μ-Law, μ = 30	DFT-OFDM	DHT-OFDM	Original OFDM
QPSK	2.7	3.65	4.53	7.72	9.21	11.5	>10	>10	>10	<10	<10	<10
16-QAM	3.27	3.81	4.5	8.75	9.75	11.27	>18	>18	>18	<18	<18	18
64-QAM	3.4	3.85	4.49	9.1	9.9	11.32						

References

1. Jiang, T., Zhu, G.: Nonlinear companding transform for reducing peak-to-average power ratio of OFDM signals. IEEE Trans. Broadcast. **50**(3), 342–346 (2004)
2. Nee, R.V., Prasad R.: OFDM for Wireless Multimedia Communications. Artech House, Inc. (2000)
3. Chung, K.C., Chen, H., Yang, T.Y.: Low complexity PTS algorithms with error correction capability in OFDM systems. In: Seventh International Conference on Ubiquitous and Future Networks (ICUFN), pp. 254–256. IEEE, July (2015)
4. Chen, H., Liang, H.: Combined selective mapping and binary cyclic codes for PAPR reduction in OFDM systems. IEEE Trans. Wirel. Commun. **6**(10) (2007)
5. Deng, S.K., Lin, M.C.: Recursive clipping and filtering with bounded distortion for PAPR reduction. IEEE Trans. Commun. **55**(1), 227–230 (2007)
6. Hao, M.J., Lai, C.H.: Precoding for PAPR reduction of OFDM signals with minimum error probability. IEEE Trans. Broadcast. **56**(1), 120–128 (2010)
7. Ali, I., Pollok, A., Luo, L., Davis, L.: A DHT precoded OFDM system with full diversity and low PAPR. In: 2012 IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 2383–2388. IEEE (2012)
8. Wang, Y., Ge, J., Wang, L., Li, J., Ai, B.: Nonlinear companding transform for reduction of peak-to-average power ratio in OFDM systems. IEEE Trans. Broadcast. **59**(2), 369–375 (2013)
9. Wang, X., Tjhung, T.T., Ng, C.S.: Reduction of peak-to-average power ratio of OFDM system using a companding technique. IEEE Trans. Broadcast. **45**(3), 303–307 (1999)
10. Elmaroud, B., Faqihi, A., Abbad, M., Aboutajdine, D.: PAPR reduction of FBMC signals by combining exponential companding and hadamard transforms. In: 2014 International Symposium on Networks, Computers and Communications, pp. 1–4. IEEE (2014)
11. Xiao, J., Yu, J., Li, X., Tang, Q., Chen, H., Li, F., Chen, L.: Hadamard transform combined with companding transform technique for PAPR reduction in an optical direct-detection OFDM system. J Opt. Commun. Netw. **4**(10), 709–714 (2012)
12. Anoh, K., Adebisi, B., Rabie, K.M., Tanriover, C.: Root-based nonlinear companding technique for reducing PAPR of precoded OFDM signals. IEEE Access **6**, 4618–4629 (2018)
13. Jiang, T., Guizani, M., Chen, H.H., Xiang, W., Wu, Y.: Derivation of PAPR distribution for OFDM wireless systems based on extreme value theory. IEEE Trans. Wirel. Commun. **7**(4), 1298–1305 (2008)
14. Rahmatallah, Y., Bouaynaya, N., Mohan, S.: On the performance of linear and nonlinear companding transforms in OFDM systems. In: Proceedings of Wireless Telecommunications Symposium (WTS), pp. 1–5 (2011)
15. Aboul-Dahab, M.A., Esam, A.A.A., Elhaseeb, A.A.: PAPR reduction based on DFT precoding for OFDM signals. Int. J. Future Comput. Commun. **2**(4), 325 (2013)
16. Ciochina, C., Sari, H.: A review of OFDMA and single-carrier FDMA and some recent results. Adv. Electron. Telecommun. **1**(1), 35–40 (2010)

Power Sector Reforms, Strategies, and Contribution of Private Sector

Rajesh Kumar Singh and Arun Kumar Singh

Abstract Power is a fundamental necessity for all walks of life and has been perceived as a fundamental human need. Indian power area is seeing real changes. Nonetheless, the interest for control has been overwhelming the development of accessibility. This is because of deficiencies in Generation, Transmission & Distribution and in addition wasteful utilization of power and shortage of financing. High level of specialized and business misfortunes and absence of business approach in the administration of utilities has prompted unsustainable monetary activities. According to Census 2001, around 44% of the families do not approach power. Power Act, 2003 gives an empowering structure to quickened and more effective improvement of the power part. The paper has been organized as follows. In the introductory section, we provide a brief background of the key issues facing the Indian power sector. In the section on the global experience in power sector reform, we discuss the key initiatives in the areas of reforms in other developing countries and what broad lessons can be learnt. Some of the key issues of the debate about privatization and its potential benefits and drawbacks are dealt in the section on way forward for the Indian power sector. In the concluding section, we try and identify some of the "Best Practices" available from the global experience in power sector reform to draft innovative solutions for the Indian power sector.

Keywords Power sector · Unbundling of the power sector · Rural electrification · Transmission & Distribution losses · Power generation sector · Privatization of power sector

R. K. Singh (✉)
Feroze Gandhi Institute of Engineering and Technology, Raebareli, India
e-mail: rajeshasinghme@gmail.com

A. K. Singh
Rajkiya Engineering College, Kannauj, India
e-mail: arun@reck.ac.in

1 Key Issues Facing the Indian Power Sector

1.1 Financial Viability of the SEBs

The M. S. Ahuluwalia Committee for restructuring of SEBs estimates that the annual aggregate loss of the SEBs in India is approx. Rs. 26,000 crore (2000–2001) [1].

It is estimated that the SEBs owe approx. Rs. 23,000 crore to the Central PSUs like NTPC, NHPC, and fuel suppliers like Coal India Limited.

SEBs also carries large pension liabilities.

1.2 Inadequate Investment in the Sector

Total energy shortage during the year 2000–2001 was 39,816 million units, i.e., 7.8% about 100,000 MW needs to be added by the year 2012.

Only 53.7% of the capacity addition target was achieved during the Eighth plan, and during the Ninth plan only half the target is likely to be achieved.

Investment in the growth of Transmission & Distribution systems has not matched the increase in generation capacity.

About 20% of the installed generation is more than 25 years old reaching their design limit.

1.3 T&D Losses

Government statistics estimate T&D losses at about 30%, but maybe at present more than 40%.

1.4 Inefficient Tariff Structure

The tariff for agriculture supply and domestic supply is much less than the cost of the service.

To cross subsidize the low agriculture and domestic tariffs, industrial tariffs are kept very high. This has led to many industries setting up their own captive generation units because they are less expensive and more reliable than power from the grid.

2 Key Lessons from Latin America and East Asian Experience

All the nations in Latin America unbundled their capacity utilities, set up a free regulatory and from that point privatized the utilities.

For Example

- In Chile, electricity scope rates have expanded from 70 to 97% and in Peru from 48 to 70%.
- In Argentina, the vitality costs have fallen altogether in the spot showcase by around half and these have been converted into bring down retail taxes also.
- In Chile, vitality misfortunes are down to 8% from 24%, in Argentina to 12% from 21% and in Peru, to 12% from 22%.

  Almost all the companies, which were making losses, turned profitable.

3 Power Sector Reforms

An ongoing report by the World Bank has cautioned India that except for the present changes in its capacity area, it could lose a large amount number of dollars in the income (Table 1).

To pivot the money-related soundness of the influence division, the Government has taken up changes in the influence area for continuous disposal of misfortunes.

Table 1 Countries in each region taking key reform steps (in Percent) [9]

Steps	East Asia and Pacific	Europe and Central Asia	Latin America and Caribbean	Middle East and North Africa	South Asia
Corporation	44	63	61	25	31
Independent regulatory body	11	41	83	0	40
IPPs allowed	78	33	83	13	100
Restructuring	44	52	72	38	40
Generation assets divested	22	37	39	13	40
Distribution assets divested	11	30	44	13	20
Privatization	33	41	78	13	40
Reform indicator	41	45	71	17	50

Power division stops to be a weight to the state's financial plan and in the end, turns into a net generator of money-related assets.

- Electricity will be provided under the most proficient conditions as far as cost and quality are concerned to help the monetary improvement of the state.
- Restructuring and unbundling of State Electricity Board while isolating Power Generation, Transmission & Distribution capacities into self-ruling and independently responsible substances, through exchange of advantages, liabilities, and work force.
- Corporatization and commercialization of new developing substances in a staged way.
- Establishing a free regulatory body.
- Promotion of private area support in control age. Privatize appropriation business in stages.
- Tariff change with the target to defend duty for full cost recuperation and limit cross appropriation.
- To activate private division assets for control creating limit expansion.

3.1 Steps to Be Taken for Power Sector Reform in a Developing Country like India

- The unbundling of the power part into Generation, Transmission & Distribution, and supply is practical.
- In the new unbundled structure, conceivably focused regions must be isolated from common restraining infrastructures and rivalry must be presented wherever conceivable.
- While presenting rivalry, passage obstructions ought to be evacuated and a level playing field be made for all players.
- An unsurprising, straightforward, and unambiguous administrative system ought to be set up.
- Privatization works are seen not exclusively to enhance the budgetary practicality of the segment, yet in addition give advantages to purchasers. Both rivalry and direction are more powerful under private possession over the long haul.
- The sequencing of the changes is basic. Having a settled administrative system, industry demonstrates preceding privatization itself accelerates the procedure and decreases the danger of disappointment.
- Government duty as budgetary help, getting purchase—in of all the partners and imparting the critical need of change is crucial for the accomplishment of the changes. The administration should draft a guide of the change pate to be embraced in the power segment to achieve more straightforwardness in the process and increment financial specialist certainty.

4 Policies of Power Sector

The policies are adopted by the power sector to create an environment which will ultimately increase the efficiency and will create revenues for the sector.

Around 56% of rustic families have not yet been jolted despite the fact that a large number of these families will pay for power. Rustic Electrification Distribution Backbone (REDB) is with no less than one 33/11 kV (or 66/11 kV) substation in each Block.

Non-customary wellsprings of vitality could be used even where the lattice network exists given that it is observed to be financially savvy (Table 2).

4.1 Generation

Despite redesigning the general availability of acquainted capacity with 85%, a turning store of no under 5%, at national dimension, would be made to ensure arrange security and quality and relentless nature of power supply. There is no need for approving for age.

To give openness of those in excess of 1000 units of per capita power by the year 2012, it had been evaluated that need-based breaking point development of more than 1,00,000 MW would be required in the midst of the period 2002–12 [2].

According to the General Review 2017 of Central Electricity Authority (CEA), India's top Government body for the planning of power sector, the consumption of power is showing an increasing trend. The nation's per capita power utilization has achieved 1075 kilowatt-hour (kWh) in 2015–16 while in 2014–15 it was 1010 kWh and in 2013–14 it was 957 kWh [3].

Per capita power utilization crossing 1,000 units a year is surely a landmark, but one-fourth of the family units in the nation still have no entrance to power. Government of India considers a village electrified if all the public buildings and

Table 2 Comparing India with Latin America countries [5]

Comparative criteria	Brazil	Argentina	Chile	Peru	India
Total foreign direct investment (Million dollar)	33.0	24	9.2	2.0	2.2
GDP at market prices	752	283	68	52	447
GDP per capita	4.350	7.550	4.630	2.130	440
Overall budget deficit	NA	NA	−1.47	−2.09	−4.22
Population density (per sq km)	19.68	13.37	20.06	19.71	335.5
Total population (M)	168	37	15	25	998
Surface area (Sq km)	8.547	2.780	756	1.285	3.287
Urban population (%)	80.72	89.6	85.44	72.42	28.08

at least 10% of the households in that village get power and by this definition, all Indian villages have now been electrified but 100% family units having 24×7 quality supply of power is still a challenge.

India is still among the bottom in the world as far as per capita power utilization is concerned, far behind China averaging 4,000 kWh, and developed nations 15,000 kWh. Around 200 million individuals lack access to power in the nation. In examination, China has per capita utilization of with created countries averaging around per capita.

4.2 Thermal Generation

Government flammable gas represents around 10% of the aggregate limit. Power division expends around 40% of the aggregate gas in the nation which is accessible at sensible costs.

Imported coal-based warm power stations, especially at seaside areas, would be supported in light of their financial suitability. Utilization of low fiery debris content coal would likewise help in diminishing the issue of fly slag outflows.

Critical lignite assets in the nation are situated in Tamil Nadu, Gujarat, and Rajasthan.

4.3 Power

Open-area ventures to make atomic age limit should be ventured up. Private division organization would likewise be encouraged to see that objectives are accomplished as well as surpassed.

4.4 Nonconventional Energy Sources

The possible capability of non-regular vitality assets, primarily a little hydro, wind and bio-mass would likewise should be Hydro Generation.

The 50,000 MW hydro activities have been now propelled and are as a rule vivaciously sought-after with DPRs for tasks of 33,000 MW limits effectively under planning.

Atomic energy misused completely to make extra power age limit.

4.5 Renovation & Modernization (R&M)

Remodeling and modernization for accomplishing higher proficiency levels should be sought after energetically and all current age limits ought to be conveyed to least worthy benchmarks.

4.6 Transmission

Keeping in view the huge increment arranged in age and furthermore for improvement of intensity showcase, there is a requirement for satisfactorily increasing transmission limit.

Open access in transmission has been acquainted with advance rivalry among the creating organizations which would now be able to pitch to various dissemination licensees the nation over. This should prompt accessibility of less-expensive power. To encourage deliberate development and advancement of the power division and furthermore for the secure and dependable activity of the matrix, sufficient edges in transmission framework ought to be made.

4.7 Distribution

Appropriation is the most basic fragment of the power business chain. The genuine test of changes in the power segment lies in productive administration of the dispersion segment. For accomplishing productivity increases, the appropriate rebuilding of conveyance utilities is fundamental.

Multi-Year Tariff (MYT) structure is a vital basic impetus to limit dangers for utilities and buyers, advance productivity, and quicken lessening of framework misfortunes. It would serve open enthusiasm through financial productivity and enhanced administration quality.

The Electricity Act, 2003 empowers contending producing organizations and exchanging licensees, other than the territory dissemination licensees, to pitch power to purchasers when open access in dispersion is presented by the State Electricity Regulatory Commissions.

SCADA and information administration frameworks are helpful for proficient working of Distribution systems. High Voltage Distribution System is a compelling technique for lessening of specialized misfortunes, avoiding of robbery, enhancing voltage profile, and bettering shopper benefit. It ought to be elevated to diminish LT/HT proportion keeping in view the techno-financial contemplations.

4.8   Technology Development and R&D

Compelling use of every single accessible asset for age, transmission, and conveyance of power utilizing productive and savvy advances is of central significance.

4.9   Transmission & Distribution Losses

It would need to be obviously perceived that the power sector will stay unviable until the point that T&D misfortunes are cut down altogether and quickly. A substantial number of states have been detailing misfortunes of more than 40% in the ongoing years.

4.10   Energy Conservation

There is a noteworthy capability of vitality reserve funds through vitality proficiency and request-side administration measures. Department of Energy Efficiency (BEE) will start acting in such manner.

In the horticulture area, the pump sets and the water conveyance framework built for high effectiveness would be advanced.

4.11   Environmental Issues

Setting up of metropolitan strong waste vitality extends in urban zones and recuperation of vitality from mechanical effluents will likewise be urged with a view to decreasing ecological contamination separated from creating extra vitality.

4.12   Protection of Consumer Interests and Quality Standards

Fitting Commission ought to control utilities in view of predecided lists on the nature of intensity supply. Parameters ought to incorporate, among others, recurrence and term of intrusion, voltage parameters, music, transformer disappointment rates, sitting tight time for the rebuilding of supply, rate flawed meters, and holding up rundown of new associations.

4.13 Fuel Usage

The new fluid fuel strategy is relied upon to take into account the hostage fuel needs of such plants. Free power makers will be allowed to utilize overwhelming oil stock (HPS) [4], low sulfur substantial stock (LSHS), substantial heater oil (HFO), heater oil (FO), and flammable gas as essential fuel for the power ventures.

5 Private Sector Participation

5.1 Introduction

India is diving headlong into a vitality emergency that may injure the nation's prospering economy if the administration does not act quickly. The opening up of the vitality area to private financial specialists as a major aspect of the Indian monetary change design may well be the jolt of energy this industry urgently needs. With private power extends probably not going to be operational for in any event, an additional two years, endeavors to enhance the effectiveness of the current plants, appear to be the main answer for India's approaching dark out.

5.2 Need for Privatization

For the certainty trap to succeed, the private power organizations and their partners have needed to encourage open trust in them and all the more essentially, enroll confided in social establishments to shore up confidence in these wobbly guarantees and enable individuals to set aside their basic resources [5]. They have possessed the capacity to select network pioneers, instructive organizations, the media, some natural gatherings, and specialists and experts, to help put forth the defense for private control of power.

Is it important to privatize, or should comparative conditions (for example, motivators and political freedom) be reproduced under government possession with the goal that proficiency might figure it out.

Setting up the earth for privatization comprises three foundations:

1. An entrenched and unambiguous institutional structure, including the administrative specialists and clear standards and strategy rules for the division is basic. A portion of the perspectives incorporates the following.

   - Independence of the administrative specialists a straightforwardness in their procedures.
   - Independence of the transmission administrator from the Market members.

Table 3 The following table shows the impact of privatization

Impact of privatization	Possible sources of increase in cost burden for poor	Mitigating factors and welfare gains
Improved collection	Revenue collection will become stricter leading to lesser informal connections	Better financial health of the utility will help reduce tariff, improve enable better quality of service and targeting of subsidies
Tariff-level adjustment	Average tariff levels tend to increase due to reduction in subsidies	Efficiency improvement, reduction in T& D losses, increase in competition, and price caps tend to decrease average tariff. Further, the government can continue to provide subsidy to the "really" needy consumer classes through an improved targeting scheme
Access cost	Cost of obtaining a connection to the service is likely to increase	The government, if necessary, can produce the actual subsidy; in any case having access at a cost is better than having no access

- Determining an ideal duty—setting component for every one of the players, in this manner the framework guarantees the decrease in cost.

2. It has been seen that if privatization is to be exceptionally successful, it is fundamental to make an aggressive domain with suitable market rules.

Some of the safeguards include
Restrictions on the level on vertical joining, cross-proprietorship crosswise over age, transmission, dispersion, and retail-supply fragment.

- Restriction on a piece of the pie. A couple of players overwhelming the market can prompt a non focused market and extraction of imposing business model rents from different fragments of the esteem chain.
- Competition in the age showcase.

3. The role that the government plays during the privatization process is critical to the success of the process (Table 3).

5.3 Financial Viability of the SEBs

- Effective correspondence with all partners in the privatization procedure including customers and representatives.

- The government additionally needs to proceed with its money-related promise to the part in the post-privatization situation.
- The government will likewise need to give a capital endowment to those private administrators who will put resources into generally ugly territories like provincial districts and urban ghettos.
- Incentives can likewise be offered to the speculators, for example, incomplete hazard certifications or halfway credit ensures.
- The development of the modern economy of India relies upon the quick improvement of the seriously slacking power age area. Despite the fact that India has bounteous regular assets for age of both Hydro and Thermal power, infrastructural weaknesses may push the power business to the brink of collapse.
- By the end of 1995, India's aggregate introduced control age limit was 81,164 MW [6]. Of that, hydropower (hydel) represents almost nearly 26%, yet hydel represents under 13% of the aggregate power produced in the initial three long periods of the present eighth arrangement.
- India has colossal stores of coal, a segment which has been halfway privatized, yet the administration has not done what is needed to pull in speculators to build up the mining business. Additionally, the idea of creating hostage mines has not yielded much advantage for the power organizations because of the issues identified with economies of scale [7].
- Another related region where this contention is expanded is the vehicle segment. Once the fundamental transport and asset use (mining, oil refining), foundation begins growing apace with the requirements of the power age industry, the warm power units will begin to be practically effective, and the necessities of Indian industry will be better met.
- Among the undertakings on the iron block is an arrangement to privatize the fuel supply framework. In most proposed control ventures, fuel spillage and fuel hazard duty have been among the principal hindrances. The state as the fuel provider is unwilling to sign any certifications and the private power organizations are unwilling to chance coming up short on fuel.
- Between 1980 and 1993, the GDP expanded by one percent, while the flexibility of power age and utilization expanded by 1.65 and 1.61%, individually. The versatility has declined from more than 3% in the primary and second 5-year wants to almost 1.5% in the Seventh Plan.

6 Ensuring Long-Term Commitment of the Private Operators

One of the issues which faces government privatizing fundamental administration utilities like power and telecom is guaranteeing the long haul responsibility of the new administrators. One of the approaches to guarantee duty is through punitive authoritative activity.

Another method for guaranteeing long haul duty is by permitting just those players to take an interest in the privatization procedure, which has the imperative specialized and business ability and a solid money-related position.

Another technique for guaranteeing responsibility is by setting up ideal conditions for the administrator in the underlying time frame. These may incorporate constrained rivalry provisions or multi-year tax administering guaranteeing a great profit for value for the administrator.

The investor understanding normally contains punishment provisions about the level of persistence, reasonability, and consideration required with respect to the new administrator as great business rehearse. It might likewise incorporate a "secure" period for the venture of the private administrator before which the financial specialist would have no alternative, however, to stay focused on the part, since he does not have the privilege to arrange off his holding without the assent of the legislature.

6.1   Enron-Led Dabhol Power Project

The primary errand indicating at any progress is the 2.5 billion dollar Enron-driven Dabhol control wander. In August 1995, things looked dismal for Enron when the wander was dropped when a preservationist government came to control in the area of Maharashtra. Later the course of action was renegotiated with insignificant changes.

The State Electricity Boards (SEBs), with the help of private accessories, need to redesign the present power plants by upgrading development.

6.2   Losses

- For The state-run power sheets, losing an expected 20 billion dollars (Rs.7,000 crore) every year.
- In 1994, the World Bank cut $750 million in credits to the power segment in light of the fact that the states declined to quit giving without end power, for the most part to politically ground-breaking agriculturists.
- An absence of proficient observing implies that an enormous measure of power is stolen in about all states; principally Delhi, Mumbai, Calcutta, Bihar, U.P., and Gujarat.
- The report recommends an infusion of 10 billion dollars throughout the following 5 years into the power division, without which India will confront 8–10 h control cuts, every day [8] (Table 4).

Table 4 Some of the problems faced by different countries [10]

Parameters	Argentina	Brazil	Chile	Peru	India
Poor financial situation	✓	✓	✓	✓	✓
High T&D losses	✓	✓	✓	✓	✓
Large investment requirement	✓	✓	✓	✓	✓
Complex structure	✓	✓	x	x	✓
Highly subsidized tariff	✓	✓	✓	✓	✓
Large unmetered supply	✓	✓	✓	x	✓
Low electricity coverage	x	x	✓	✓	✓
Low employee productivity	✓	✓	✓	✓	✓
Poor quality of power	✓	x	✓	✓	✓

7 Conclusion

Power area is seeing a basic stage. State Electricity Boards (SEBs) are in charge of giving power to the general population. The vast majority of the SEB share is desperate. They are not in any case ready to win a base Rate of Return (RoR) of 3% on their net settled resources in the benefit in the wake of accommodating deterioration and intrigue accuses in agreement of Section 59 of the Electricity (Supply) Act, 1948. The power area in the nation has aggregated a gigantic shortage, duty to Central Power Generating Companies in light of the disintegrating monetary execution.

Financial changes, presented in 1991, are gradually changing India from a monetary slowpoke into an essential developing business sector. Be that as it may, governmental issues, a tangled organization and defilement have dismissed numerous potential speculators. Privatization can prompt India to be a noteworthy power creating nation, yet it ought to be kept free from political impacts and impacts which will not just ked to vain outcomes, yet will be more awkward to break down.

References

1. www.uppcl.org
2. http://www.livemint.com/…/Indias-per-capita-electricity-consumption-touches-1010
3. http://www.cea.nic.in/reports/others/planning/pdm/2017.pdf
4. www.ieo.org
5. Arizu, B., Tenenbaum, B.: Regulating Transmission: Why System Operators Must be Truly Independent (2001)
6. www.ntpc.com
7. www.recl.com
8. Abdala, A., Chamboulevron, A.: Transmission Investment in Competitive Power System-Decentralizing Decisions in Argentina (1999)

9. Albouy, Y., Bousha, R.: The Impact of IPPs in Developing Countries—Out of the Crisis & into the Future (1998)
10. Arthur Andersen.: Fundamentals of the Global Power Industry, Energy Crisis in the Western U.S, Power Thinking_Global Electricity Market

Performance Evaluation of IEEE 802.11p Physical Layer for Efficient Vehicular Communication

Devesh Shukla, Vinay Kumar and Arun Prakash

Abstract Intelligent transportation systems have gained huge interest worldwide for providing safer, secure, and efficient communication systems for emergency, time management, and fuel-saving applications. Vehicular Ad-hoc Network technology provides a set of services and applications for the above requirements much effectively than other technologies. IEEE 802.11p standard is developed for handling the requirements and issues of vehicular communication that includes communication between Vehicle-to-Vehicle and Vehicle-to-Infrastructure. IEEE 802.11p employs orthogonal frequency division multiplexing technology for vehicular communication. This work evaluates the performance of 802.11p physical layer under extreme channel conditions. The quality of transmission is measured for different standard modulation techniques in terms of bit error rate with respect to varying signal-to-noise ratio.

Keywords Vehicular Ad-hoc Network · DSRC · WAVE · IEEE 802.11p · Intelligent transportation systems

1 Introduction

With the recent advancements and increased transport on roads, there is a huge demand for a safer and secure transportation system. Vehicular Ad-hoc Network (VANET) provides both Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication systems and makes vehicular communications more intelligent. Hazardous obstacles and traffic congestions around can be avoided with the use of VANET technology. VANET application includes traffic management, driv-

D. Shukla (✉) · A. Prakash
Department of Electronics and Communication Engineering,
Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India
e-mail: erdevesh89@hotmail.com

V. Kumar
Department of Electronics and Communication Engineering,
Visvesvaraya National Institute of Technology, Nagpur, India

ing assistance, navigation, electronic payments, and commercial applications. These applications depend on the transfer of data between vehicles and other roadside units where data rate can vary from 3 to 27 Mbps for maximum velocity of vehicle up to 260 kmph and works within a range of approximately 300 m [1].

IEEE 802.11p standard known as Wireless Access in Vehicular Environments (WAVE) is a modification to IEEE 802.11a protocol that is specially developed to support Intelligent Transportation Systems. It provides PHY and MAC amendments to IEEE 802.11a as per VANET requirements. A comparison between IEEE 802.11a and 802.11p standard with the modification made in the existing IEEE 802.11a in order to handle the VANET requirement and channel variation accurately is shown in Table 1 [2, 3]. It mainly involves twofold of all the timing OFDM parameters to overcome the channel fading caused by wireless fading channel.

IEEE 802.11p employs a mechanism to operate in the Dedicated Short Range Communication (DSRC) band to provide V2V and V2I communication. DSRC operates in the 5.9 GHz frequency band having a bandwidth of 75 MHz (5.85–5.925 GHz) and communication range typically between 100–300m. DSRC spectrum is structured into seven 10 MHz wide channels (172–184) and a guard band of 5 MHz channel 172 and 184 is dedicated as safety channels while 178 is control channel used for link establishment and control of transmission broadcast. The other channels are used as service channels for bidirectional communication between the vehicles. Two 10 MHz channels can sometimes be combined together to be used as a single 20 MHz channel. DSRC uses Orthogonal Frequency Division Multiplexing (OFDM) modulation scheme for data transmission that uses the spectrum more effectively. OFDM employs multicarrier for data transmission that subsequently divides the available frequency bands into multiple sub-bands also called subcarriers with a fix frequency separation *f*. The high data rate streams are divided into *k* number of parallel subcarriers which are modulated by low data rate streams. The subcarriers are spaced orthogonally to one another to prevent the interference from adjacent sub-bands. Thus, OFDM proves to be a reliable method for efficient transmission of data over a radio channel.

Table 1 Comparison of IEEE 802.11a and 802.11p PHY layer parameters

Parameters	IEEE 802.11a	IEEE 802.11p
Bit rate (Mbps)	6, 9, 12, 18, 24, 36, 48, 54	3, 4.5, 6, 9, 12, 18, 24, 27
Modulation mode	BPSK, QPSK, 16-QAM, 64-QAM	BPSK, QPSK, 16-QAM, 64-QAM
Code rate	1/2, 2/3, 3/4	1/2, 2/3, 3/4
No. of subcarriers	52	52
Symbol duration (μs)	4	8
Guard time (μs)	0.8	1.6
FFT period (μs)	3.2	6.4
Preamble duration (μs)	16	32
Subcarrier spacing (MHz)	0.3125	0.15625

The main contribution of this work is the performance evaluation of IEEE 802.11p physical layer model based on OFDM technology for the transmission. The parameter used for analyzing the performance of the system is bit error rate (BER) for different standard modulation schemes (BPSK, QPSK, 16-QAM, 64-QAM) for varying Signal-to-Noise Ratio (SNR). The simulation analysis is carried out for severe channel conditions.

The rest of the paper is organized as follows. Section 2 covers the literature survey of amendments of IEEE 802.11p PHY proposed by researchers. Section 3 illustrates the detailed transmission process of the IEEE 802.11p standard. It covers the steps involved in the formation of the complete OFDM symbol. Section 4 deals with the results and discussions of this work which includes performance analysis of the PHY layer on the basis of BER. Section 5 concludes the paper with an outline of future work.

2   Related Works

VANET eases the transportation system by providing a huge set of applications specially designed to reduce accidents and traffic by providing coordination among vehicles and roadside scenario. Study of VANET and its physical layer is an important step before its implementation in real-time scenario. An overview of DSRC technology and IEEE 802.11p physical layer is given in [1, 4]. The PHY specifications, components, performance, and challenges are discussed and analyzed in the same. PHY defined in IEEE 802.11p consists of two sublayers namely Physical Medium Dependent (PMD) sublayer which acts as an interface to the physical transmission medium; and Physical Layer Convergence Protocol (PLCP) sublayer which communicates with the MAC layer.

The technologies used in the WAVE standard, its applications and limitations have been discussed in [5]. In [6], OFDM transmission performance in vehicular communication is implemented for different scenarios using different modulation schemes. The BER versus SNR of all modulation schemes with different coding rates has been evaluated for performance analysis of the communication. There are several parameters that affect the transmission reliability such as channel type, coding rate, speed, SNR variance, and their dependence on system performance has been provided.

The design parameters of OFDM technology impact the performance of the IEEE 802.11p physical layer. The selection of OFDM design parameters proves to be an important area for the desired result. The detailed analysis of OFDM system with various design parameters like constellation mapping schemes, symbol rate, and other factors like mobility, RSU antenna height have been studied in [7, 8]. All factors and OFDM design parameters should be considered while designing VANET physical layer. The minimum necessary amendments that should be made in the parent IEEE 802.11a PHY and MAC layer are discussed in [9]. These amendments are necessary at PHY level to support the communication among moving vehicles and overcome

the effects caused by the vehicular scenario. IEEE 802.11p uses 10 MHz of channel bandwidth instead of 20 MHz that is used in the earlier IEEE 802.11a standard. The main reason behind this modification is to make the system more compatible with RMS delay spread.

The channel modeling for the VANET scenario is suggested in [10]. Transmission Channel in case of VANET is radio channel which plays an important role and affects the successful communication in the vehicular environment. In a vehicular communication network in addition to Line-of-Sight (LOS) and multipath propagation, signal is affected by doppler shift introduced by moving vehicles. Therefore, propagation channel is modeled as a combination of AWGN, block fading or frequency selective channel, and time-variant channel.

Channel estimation is an important aspect in IEEE 802.11p due to extremely time-variant characteristics of wireless channels and thus proper estimation scheme enhances the system performance and accuracy. The basic technique used commonly for the channel estimation is Least Square (LS) estimation with very low complexity level. Since vehicular environment channel is time-variant and LS estimation has limitations, system performance gets affected. In [11], a survey on the current channel estimation techniques such as for the IEEE 802.11p standard has been presented.

VANET physical layer employs OFDM technology for data processing and communication. OFDM deals efficiently according to the requirement of the VANET environment. BER analysis compares the system performance which depends upon several factors that include mainly coding rate, modulation schemes, channel variations, and velocity of the vehicle. Channel estimation and equalization increase the efficiency and accuracy of the receiver by reducing the error probabilities caused due to channel variations or multipath fading and time-variant channel.

3   IEEE 802.11p Transmission Process

The detailed steps involved in the transmitter and receiver of the VANET physical layer is explained in Fig. 1 [4]. First of all, the data received from the upper layer is scrambled in order to avoid errors that arise due to long bit sequence. It simplifies the task of a timing recovery circuit and makes the system more immune toward PAPR
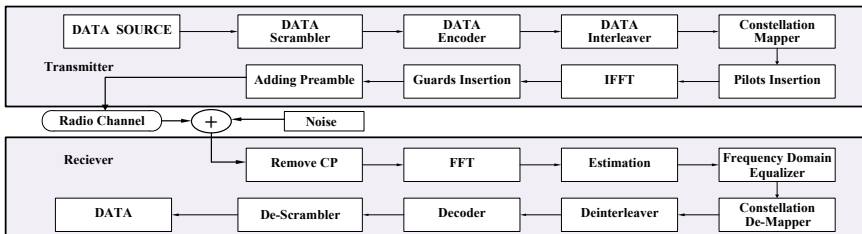
Fig. 1 An IEEE 802.11p transceiver

requirement. Scrambler generates a 127-bit sequence using a generator polynomial $S(x)$ for initial state 1011101 [5].

$$S(x) = x^7 + x^4 + 1$$

Scrambled data is encoded which introduces some redundancy in the incoming stream through the use of linear shift registers. It is used for error detection and correction and also to make the data more compatible against channel interference. Convolutional encoding with coding rate 1/2 is used which can be varied to other rates 2/3 and 3/4 with the help of puncturing. Further data is interleaved to make it more secure against correlated noise like burst errors and fading. Interleaver basically is a two-step permutation process which ensures the correct mapping of data. IEEE 802.11p PHY specifically describes four mapping schemes BPSK, QPSK, 16-QAM, and 64-QAM. Interleaved data is then modulated using different modulation schemes which can result in data rate ranging from 3 Mbps (with BPSK and 1/2 code rate) to 27 Mbps (with 64-QAM and 3/4 code rate) with the help of various modulation schemes and code rates. The data rate for various modulation schemes for different code rates is based on Table 2 [12]. Coded serial bits are converted into complex values according to Gray-coded constellation mapping. In IEEE 802.11p, there are a total of 64 subcarriers available for transmission. Pilot signals use 4 subcarriers out of a total of 64 for carrying the fixed pattern of signal to make coherent detection robust and to ensure accurate frequency and time synchronization.

Modulation and demodulation in OFDM are implemented using IFFT and FFT, respectively. Complex data stream is transformed into an analog signal waveform in time domain. The IFFT output is the summation of all 64 sinusoids that results in a single OFDM symbol s[n].

$$s[\text{n}] = \frac{1}{K} \sum_{k=0}^{k-1} S[\text{k}] e^{j2\pi \frac{kn}{K}}$$

After successful transformation in time domain, each OFDM symbol is added with cyclic prefix. Cyclic prefix is prepending the copy of the last samples of the symbol to the beginning of the same symbol. This is done to avoid the Inter Channel Interference and Inter Symbol Interference problems caused by multipath propagation and fading. Final OFDM symbol consists of 10 Short training preambles of 1.6 μs each used for

Table 2 IEEE 802.11p modulation schemes and data rate

Modulation type	BPSK		QPSK		16-QAM		64-QAM	
Coding rate	1/2	3/4	1/2	3/4	1/2	3/4	2/3	3/4
Coded bit rate in Mbps	6		12		24		36	
Data rate in Mbps	3	4.5	6	9	12	18	24	27
Data bits per OFDM symbol	24	36	48	72	96	144	192	216
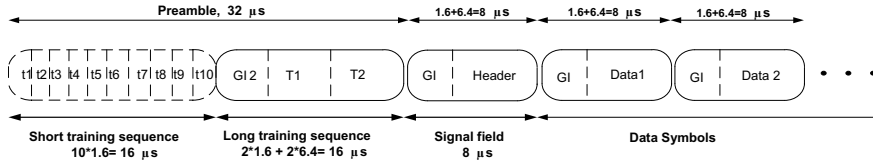
Fig. 2 IEEE 802.11p PPDU frame structure

signal detection, diversity selection and 2 long training symbols of 3.2 μs each which are used for the synchronization and channel estimation in receiver, signal field which carries information about coding rate, modulation type, etc., and various numbers of data symbols. The Protocol Packet Data Unit (PPDU), i.e., the final frame structure is composed of preambles, signal field, and the payload containing the significant data is shown in Fig. 2 [11].

Finally, the OFDM signal is transmitted over a wireless channel that is generally a frequency selective channel with additive noise. OFDM technology converts the frequency selective channel into parallel flat fading channel by using the concept of cyclic prefix. The channel is simply AWGN but due to the presence of both direct and reflected links between the sender and receiver, the channel is also considered as Rayleigh and Rician fading radio channel. The receiver performs the exact reverse process of transmitter for data extraction with additional step of estimation and equalization. Channel estimation is done with the help of the long training sequence preambles by Least Square Estimation Method and also the frequency domain equalizer is added for equalization.

4 Results and Discussion

In order to investigate the performance of IEEE 802.11p physical layer, simulation is done using MATLAB–Simulink software [13]. The study of PHY layer shows that the quality of transmission depends on multiple factors including modulation schemes, velocity of vehicles, propagation channel, frame size, coding rate, proper channel estimation, etc. The serial data stream is converted to parallel ones using the OFDM technique. It modulates these parallel complex data onto orthogonal subcarriers using IFFT. As mentioned earlier also, IEEE 802.11p employs OFDM with 64 subcarriers (indexed from -32 to 31) for signal transmission. Out of the total subcarriers, available 48 are used for data information, 4 for pilot signals and rest 12 as zero subcarriers as shown in Fig. 3. Pilot signals are positioned onto the subcarriers $-21, -7, 7$, and 21 with pilot sequence $\{1, 1, 1, -1\}$. It modulates these parallel complex data using IFFT which converts it to time domain from frequency domain that is appropriate for the transmission of the processed data over the wireless channel. Finally, guard interval is inserted as cyclic prefix whose duration is equal to GI = IFFT/4 (typically 1.6 μs).
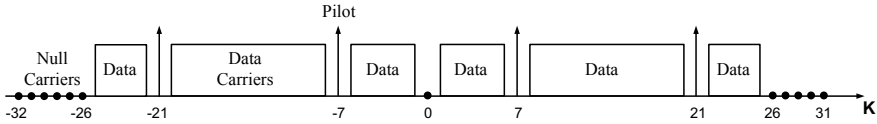
Fig. 3 Carrier allocation for OFDM signal

The encoded and modulated data is added with the signal field and preambles to form the complete PPDU frame as the complete OFDM symbol as shown in Fig. 2. The PLCP preamble is a combination of 10 Short training sequences and 2 long training sequences. A short training symbol has a duration of 1.6 μs and is given by the sequence:

$$
\begin{aligned}
S = \sqrt{(13/6)}\{&0, 0, 0, 0, 0, 0, 0, 0, +1+j, 0, 0, 0, -1-j, 0, 0, 0, \\
& +1+j, 0, 0, 0, -1-j, 0, 0, 0, -1-j, 0, 0, 0, \\
& +1+j, 0, 0, 0, 0, 0, 0, 0, -1-j, 0, 0, 0, -1-j, 0, 0, 0, \\
& +1+j, 0, 0, 0, +1+j, 0, 0, 0, +1+j, 0, 0, 0, \\
& +1+j, 0, 0, 0, 0, 0, 0, 0\}
\end{aligned}
$$

The long training symbols used for channel estimation have the duration of 6.4 μs and are given by sequence:

$$
\begin{aligned}
L = \{&0, 0, 0, 0, 0, 0, +1, +1, -1, -1, +1, +1, -1, +1, -1, +1, \\
& +1, +1, +1, +1, +1, -1, -1, +1, +1, -1, +1, -1, +1, +1, \\
& +1, +1, 0, +1, -1, -1, +1, +1, -1, +1, -1, +1, -1, -1, -1, \\
& -1, -1, +1, +1, -1, -1, +1, -1, +1, -1, +1, \\
& +1, +1, +1, 0, 0, 0, 0, 0\}
\end{aligned}
$$

This work mainly focuses on the determination of BER with respect to varying SNR for performance analysis using the available mapping schemes under extreme channel conditions. A comparison is made between each modulation technique depending upon the simulated results. Several propagation channel models are considered to examine and implement the VANET physical layer. The propagation distance in simulation is taken as 300 m and speed of the mobile station as 90 km/hr and urban environment is chosen as the scenario. Simulation is carried out using different modulation schemes and varying SNR. BER is evaluated for all available combinations of modulation schemes (BPSK, QPSK, 16-QAM, 64-QAM) and fixed coding rate for an AWGN channel and a Rayleigh fading channel. The BER versus SNR curve is plotted from the results obtained from the simulation.

The spectrum of 10 MHz channel for the transmitted OFDM signal in the frequency range of 5.9 GHz band is shown in Fig. 4. Figure 5 shows the BER versus SNR curve of various modulation schemes over an AWGN channel. Since the radio

channel in vehicular communication is a composition of both Line-of-Sight (LOS) and Non Line-of-Sight (N-LOS) path; the LOS path is modeled as Rician fading channel and N-LOS path as Rayleigh fading channel [14]. In Fig. 6, the BER versus SNR curve over a fading channel with AWGN noise where the Doppler spread and RMS delay spread play an important role in the behavior of the system. The parameter Doppler spread is chosen as 200 Hz and RMS delay spread as 100 ns in the simulation. The RMS delay spread is less than the cyclic prefix to overcome the distortion and interference. The simulation result shows that the best performance is obtained for BPSK and QPSK schemes in comparison to 16-QAM and 64-QAM. This is because in 16-QAM and 64-QAM symbols are densely spaced to each other in comparison to BPSK and QPSK. There is a trade-off between the larger spectral efficiency of 16-QAM and 64-QAM and BER. Also with the increase in the Doppler shift, the performance of the system reduces. In cases where BER is bigger than 10^{-2}, received signal suffers from frequency selective fading and several measures such as error-correcting coding and estimation techniques can be used to enhance the system performance.

5   Conclusions

VANET is emerging as an important research area that proves to be better than the technology currently in use in terms of contribution toward safer and secure transportation systems. PHY is a key aspect for achieving the applications of VANET. This paper presents the overview and specifications of PHY of IEEE 802.11p stan-
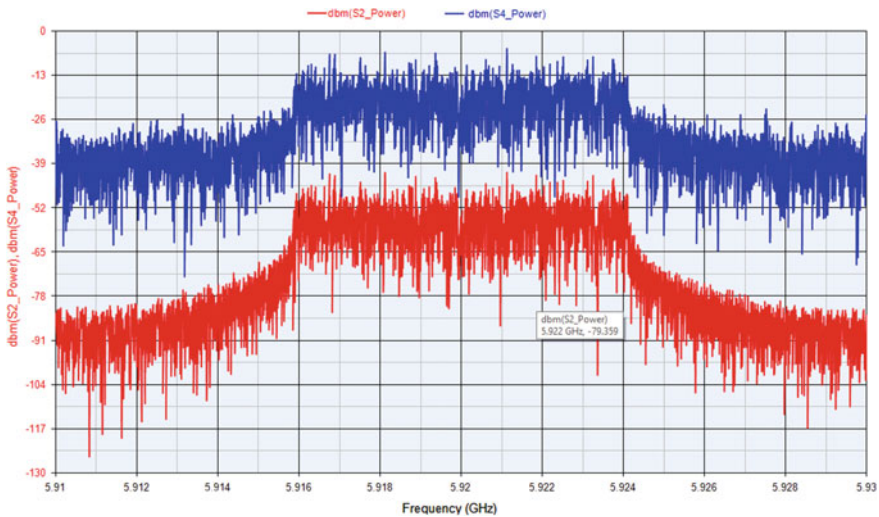
Fig. 4 Spectrum of the transmitted OFDM signal

Fig. 5 BER versus SNR over AWGN channel
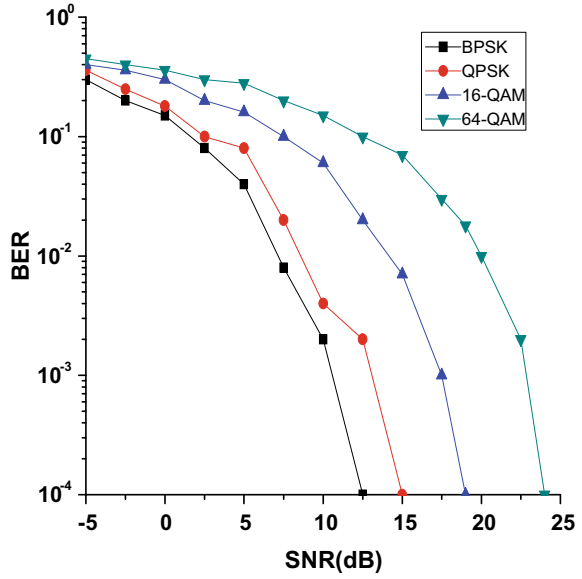
Fig. 6 Comparison of different modulation schemes over a fading channel
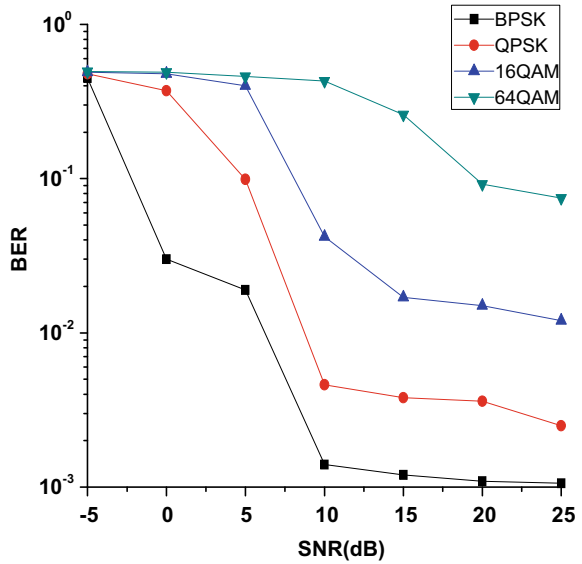
dard. Moreover, the simulation of the PHY model has been done. The results show that better performance is achieved with modulation schemes which use low coding rate and where the symbols are less densely spaced. BPSK with code rate 1/2 modulation proves to be the most robust scheme for the transmission process. Also, BER analysis can be done for more available channels based on real-time environments and different code rates which can be obtained with the help of puncturing. Channel estimation and equalization is an important issue in the designing of receiver for handling channel variations. Therefore, further sophisticated channel estimation schemes can be developed for more accurate and precise results.

References

1. Kenney, J.B.: Dedicated short-range communications (DSRC) standards in the United States. In: Proceedings of the IEEE, vol. 99, pp. 1162–1182. IEEE (2011). https://doi.org/10.1109/jproc.2011.2132790
2. Draft amendment to standard for information technology telecommunications and information exchange between systems local and metropolitan area networks specific requirements. Part 11: wireless LAN medium access control(MAC) and physical layer specifications Amendment 7: wireless access in vehicular environment, 2007
3. IEEE Std 802.11-2007 Part 11.: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications high-speed physical layer in the 5 GHz band
4. Abdelgader, A.M., Wu, L.: The physical layer of the IEEE 802.11p WAVE communication standard: the specifications and challenges. In: Proceedings of the World Congress on Engineering and Computer Science, vol. 2, p. 71 (2014)
5. Li, B., et al.: Wireless Access for Vehicular Environments, p. 205. Internet: http://www.Mehrpouyan.Info/Projects/Group (2010)
6. Sassi, A., Charfi, F., Kamoun, L., Elhillali, Y., Rivenq, A.: OFDM transmission performance evaluation in V2X communication. IJCSI Int. J. Comput. Sci. Issues **9**(2), No 3 (2012)
7. Arslan, S., Saritas, M.: The effects of OFDM design parameters on the V2X communication performance: a survey. Veh. Commun. **7**, 1–6 (2017)
8. Mohamed, M.A., Samarah, A.S., Fath Allah, M.I.: Study of performance parameters effects on OFDM systems. Int. J. Comput. Sci. Issues (IJCSI) **9**(3), 410 (2012)
9. Jiang, D., Delgrossi, L.: IEEE 802.11p: towards an international standard for wireless access in vehicular environments. In: Vehicular Technology Conference, 2008. VTC Spring 2008, pp. 2036–2040. IEEE (2008)
10. Hafeez, K.A., Zhao, L., Liao, Z., Ma, B.N.W.: The optimal radio propagation model in VANET. In: Fourth International Conference on Systems and Networks Communications, 2009. ICSNC'09, pp. 6–11. IEEE (2009)
11. Zhao, Z., Cheng, X., Wen, M., Jiao, B., Wang, C.-X.: Channel estimation schemes for IEEE 802.11p standard. IEEE Intell. Transp. Syst. Mag. **5**(4), 38–49 (2013)
12. Doufexi, A., Armour, S., Butler, M., Nix, A., Bull, D., McGeehan, J.: A comparison of the HIPERLAN/2 and IEEE 802.11 a wireless LAN standards. IEEE Commun. Mag. **40**(5), 172–180 (2002)
13. Matlab and Simulink. http://www.mathworks.com/products/simulink/
14. Kiokes, G., Amditis, A., Uzunoglu, N.K.: Simulation-based performance analysis and improvement of orthogonal frequency division multiplexing–802.11p system for vehicular communications. IET Intel. Transport Syst. **3**(4), 429–436 (2009)

A Robust Energy-Efficient Cluster-Based Routing Protocol for Mobile Wireless Sensor Network

Yogesh Tripathi, Vinay Kumar and Arun Prakash

Abstract The mobile wireless sensor network is a promising technology having a wide number of applications. The sensor nodes are mobile and able to communicate with each other in an ad hoc manner. Due to mobility, it outperforms the static wireless sensor network as MWSN increases the throughput, network lifetime, and reduces energy consumption. Mobile sensor network has better ability to monitor the target area than static sensor network. However, the routing protocol in mobile environment is complex in resource constraints MWSN. So, it is required to develop an energy-efficient routing protocol to improve network performance. In this paper, a robust energy-efficient cluster-based routing protocol is proposed. The energy-rich node is selected as cluster head with minimum velocity for maximum connectivity among the cluster members. Selection of reliable forwarder improves network performance. Extensive simulation study is carried out to evaluate the performance of the proposed routing protocol with respect to delay, throughput, PDR, and total energy consumption.

Keywords Mobile Wireless Sensor Network · Clustering · Energy efficiency · Reliability

1   Introduction

A Mobile Wireless Sensor Network (MWSN) is used in the modern world for monitoring physical or biological phenomena. It has the ability to move within where it is deployed [1]. The sensor nodes are small in size which consist of three basic components: sensing subsystem, processing subsystem for data processing, and storage

Y. Tripathi (✉) · A. Prakash
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India
e-mail: rel1706@mnnit.ac.in

V. Kumar
Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology Nagpur, Nagpur, India

and wireless communication subsystem for data transmission [2, 3]. The existing literature suggests that wireless sensor network with mobility is advantageous [4–7]. Nodes are able to move in the network with the help of mobilizer [8] or self propel via springs [9] or mounted on vehicles, animals, and robots [10]. The majority of researchers of sensor networks show their interest toward static networks but the current research is trending toward mobile sensor networks due to the following advantages [11]

- Required numbers of sensor nodes are less due to the sparse architecture of the network.
- Improved sensing of the target area.
- Due to mobility, network partition is avoided so that it improves the data fidelity as well as network lifetime.
- Communication distance decreases due to mobility which reduces energy consumption during data transmission.

Although the mobility has several advantages in sensor networks routing, it has some limitations. There exist many applications which require mobile sensor node. The major application of MWSN are ecology monitoring of migrating birds, collecting the information from mobile workers or vehicles which are equipped with the sensors about agriculture production, e-voting, traffic system, atmospheric conditions as well as health condition of workers inside coal mines and firefighters moving in burning building [1, 12, 13]. One of the crucial applications of mobile sensor networks is military application. In military application, it is used for supporting and assisting the soldiers in battlefield assessment and for searching survivors in disaster areas [11].

In this paper, a robust energy-aware cluster-based routing protocol is designed for frequently changing mobile sensor network topology. It uses energy, velocity of the node, and energy ratio for selection of cluster head (CH). Cluster head receives data from cluster members and forwards to the destination with the help of multi-hop mechanism.

The rest of the paper is organized as follows: Sect. 2 provides details of the current routing algorithm of MWSNs. Section 3 describes the proposed routing protocol which includes the details of cluster head selection and data forwarding mechanism. This is followed by the performance evaluation of the proposed routing protocol in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Work

In the literature, several energy-efficient routing protocols have been proposed for WSNs [4–6].

In E^2R^2 [14], energy-efficient and reliable routing protocol is proposed. The main contribution of this paper is to route the data packet efficiently in spite of node mobility. CH and deputy CH are responsible for the reduction of energy consumption.

In [15], energy efficiency along with connectivity among the nodes is the major designing objective. The proposed routing protocol does not depend on mobility while forwarding the data.

There are some hierarchical routing protocols that exist in the literature for MWSNs. LEACH-M [16], based on LEACH and nodes randomly moving between clusters improve the network connectivity. LEACH-ME [17] is an improvement of LEACH-M, wherein the cluster head is selected based on the mobility of the nodes. The network becomes more stable if low-velocity node is selected as the cluster head.

In [18], authors have proposed a robust ad hoc sensor network routing (RASeR) protocol. In this routing protocol, hop count is selected as gradient metric for the data transmission along with global time division multiple access scheme (GTDMA). The main feature of GTDMA is that it eliminates the collision due to slots assigned to each node, and at a time only one node can transmit data.

Thus, none of the existing routing protocols have addressed the following issues at the same time:

(1) Selection of CH based on node mobility and energy;
(2) Selection of forwarding nodes based on mobility, residual energy, and reliability of the medium;
(3) Minimizing energy consumption with the help of reduction in message overhead.

In this paper, cluster-based routing protocol for mobile WSN having the sink node is stationary. The minimum velocity node is selected as CH and forwarder node to improve the network performance.

3 Proposed Protocol

In this section, clustering and routing algorithm are discussed along with network model. The proposed algorithm minimizes energy consumption and thereby improving the network lifetime of MWSN. The different parameters like residual energy, velocity, and reliability are taken into consideration for designing of clustering and routing algorithms.

3.1 Network Model

In this paper, sensor nodes are mobile and each node has different velocities. One of the sensor nodes is selected as CH. CH is responsible for further data transmission. The network model of this proposed protocol is given in Fig. 1.

Fig. 1 Network topology
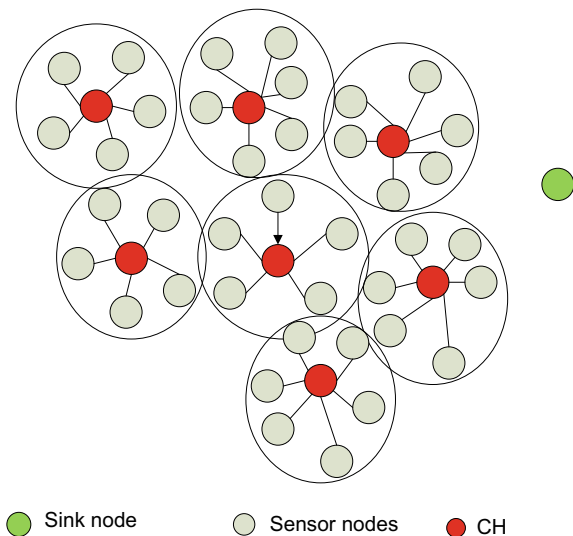
3.2 Cluster Head Selection and Cluster Formation

CH should be such that it can handle the data of its cluster member for efficient data transmission. So, CH is selected based on the velocity, residual energy, and ratio of remaining energy and initial energy. CH is selected from Eq. (1)

$$T_{CH} = 1 - \frac{E_{res} \times k}{v_i} \tag{1}$$

where T_{CH} is the timer for sensor nodes and the node which has minimum timer value will work as CH, E_{res} is residual energy of nodes, v_i is velocity of node i, and k is ratio of remaining energy and initial energy. In Eq. (1), minimum velocity of node and higher value of residual energy and energy ratio are responsible for the calculation of minimum value of timer. The minimum velocity node is selected to become CH due to connectivity among the nodes with CH. So, strong connectivity of nodes with CH improves the packet loss. To become a CH, each node calculates their timer according to Eq. (1). The node which has minimum timer value will broadcast an advertisement packet to inform the other nodes to act as CH. Once the nodes receive an advertisement packet, they stop their timer and become cluster members. After selection of CH, each node transmits a request message to CH to join the cluster and CH approves the request and responds to each node to work as cluster member.

3.3   Proposed Routing Algorithm

In the proposed algorithm, the sink node broadcast a periodical message which will help in the selection of the next hop forwarder. The next hop forwarder may be CH or any sensor node. If the next forwarder is CH, then data is directly transmitted to CH otherwise the corresponding CH transmits an RREQ message, and the node which responds is selected as the next hop forwarder. If the forwarder node is sensor node for the CH, its RSSI value should be greater than CH. This shows that the selected forwarder is nearer to sink in comparison with CH. It will reduce energy consumption that will prolong the network lifetime. The algorithm of the proposed routing protocol is given below.

Algorithm 1: Proposed Routing Algorithm

---

1. CH broadcast RREQ message with its ID and RSSI value

2. **If** next CH is in range with RSSI greater than received RSSI in RREQ message

3. Selected CH broadcast reply message for data transmission

4. Data transmission takes place

5. **else** next hop forwarder is selected from equation (2) with RSSI greater than received RSSI in RREQ message

6. Data is transmitted on selected next hop forwarder

7. **end**

In this protocol, once the CH is selected as the next hop forwarder, it will broadcast to other nodes to stop their timer for being the next hop forwarder. The selection of sensor nodes as the next hop forwarder is determined by Eq. (2) which is given below

$$t_i = \frac{1}{E_{res} \times v_i \times \eta} \tag{2}$$

where t_i is the timer for sensor nodes to become the next hop forwarder, E_{res} is residual energy of nodes, v_i is velocity of node i, and η is reliability of medium which varies between 0 and 1. From the above equation, the node which has the smallest timer value will be selected as the next hop forwarder and data transmission takes place.

4   Performance Evaluation

Simulation of the proposed routing protocol is carried on network simulator (ns-2.35) [19] and ethe performance is evaluated. In performance evaluation, different routing metrics like energy consumption, network lifetime, throughput, etc., are taken

into consideration for comparison with the existing routing protocol. The proposed routing protocol has improved results over the existing routing algorithm. It is due to clustering which selects energy-rich node as CH. The parameters used in the simulation are given in Table 1.

The results of the proposed algorithm are compared with energy-efficient and reliable routing (E^2R^2) for mobile WSN. It shows the improvement over the existing routing protocol. Figure 2 shows the variation of delay with the number of nodes. Delay shows the time taken by a packet to reach the destination from the source. In the proposed routing protocol, initially cluster formation and selection of CH takes time. Once CH is selected, data is transmitted from source node to destination node. As the number of nodes increases, due to increment in packet generation, delay also increases due to congestion. The RSSI value of the node and selection of reliable forwarder improves the delay as it decreases the retransmission.

Throughput is the measure of the number of bits processed over the given bandwidth. As the number of nodes increases, throughput decreases due to more number of packets attempting for transmission over the fixed bandwidth. The minimum velocity node as CH and reliable node as the next hop forwarder helps in the reduction in packet drop. The connectivity of CH among cluster members is high which improves

Table 1 Simulation parameters

Parameter	Value
Number of nodes (n)	20–100
Simulation time (s)	200
Simulation area (m^2)	5000 × 5000
Mac protocol	802.15.4
Traffic types	CBR
Packet transmission rate	Random
Velocity (m/s)	(0–4)

Fig. 2 Delay versus number of nodes
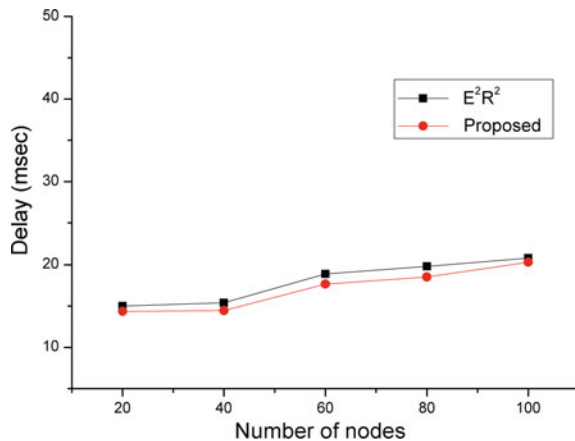
Fig. 3 Throughput versus
number of nodes
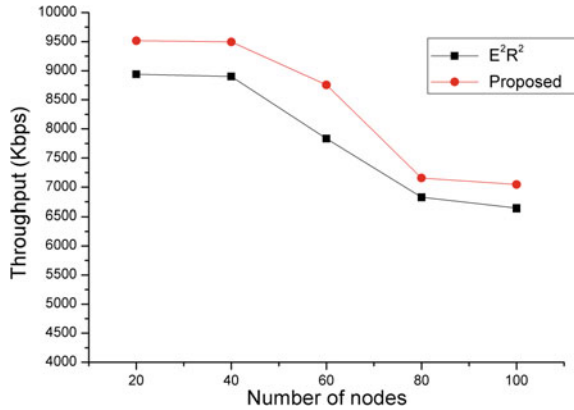
the packet reception rate as shown in Fig. 3. So, the proposed routing protocol shows
the improvement over the existing routing protocol.

Packet delivery ratio (PDR) represents how many packets are received in com-
parison of the generated packet. In Fig. 4, as the number of nodes increases, PDR
decreases due to congestion. In the proposed routing protocol, connectivity between
sensor nodes and CH is high due to minimum velocity of CH. Due to strong con-
nectivity among the nodes, it reduces the packet drop which improves the PDR. In
Fig. 5, energy consumption of the network is increased due to more number of packet
transmissions. In the proposed routing algorithm, selection of energy-efficient CH
and reliable next hop forwarder improves the packet delivery ratio which reduces the
packet retransmission attempt. So, it reduces the energy consumption.

Fig. 4 PDR versus number
of nodes
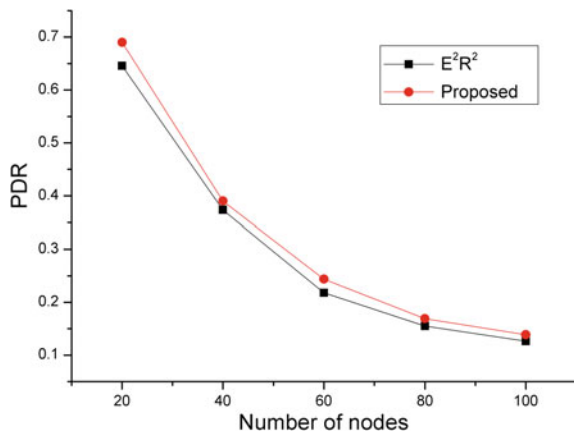
Fig. 5 Total energy consumption versus number of nodes
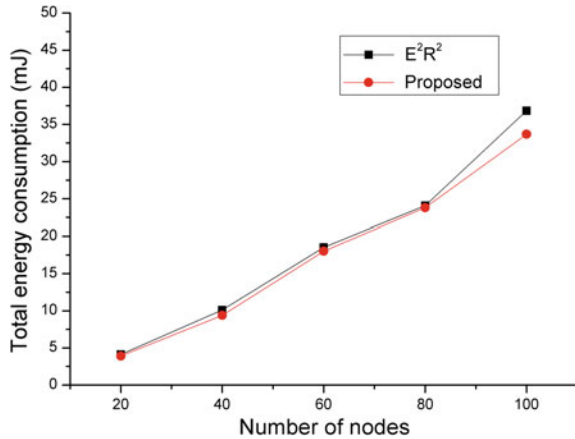
5   Conclusion and Future Work

This paper represents robust energy-efficient cluster-based routing protocol for mobile wireless sensor networks. The CH is selected based on timer expiration, which is the function of velocity, residual energy, and energy ratio. The forwarder node is selected based on reliability, velocity, and residual energy. The minimum velocity nodes help in the reduction in packet drop which reduces the retransmission attempt. So, the average energy consumption reduces which improves the network lifetime. The extensive simulation studies are carried on ns-2.35 for the performance evaluation. It is envisaged that this work can be extended to routing algorithm for mobile sink.

References

1. Munir, S.A., Biaoren,W.J., Wang, B., Xie, D., Ma, J.: Mobile wireless sensor network: architecture and enabling technologies for ubiquitous computing. In: Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshop (AINAW'07) (2007)
2. Anastasi, G., Conti, M., Di Francesco, M., Passarella, A.: Energy conservation in wireless sensor networks: a survey. Ad Hoc Netw. **7**, 537–568 (2009)
3. Guan, X., Guan, L., Wang, X.G., Ohtsuki, T.: A new load balancing and data collection algorithm for energy saving in wireless sensor networks. Telecommun. Syst. **45**, 313–322 (2010). https://doi.org/10.1007/s11235-009-9269-3
4. Liu, B., Brass, P., Dousse, O., Nain, P., Towsley, D.: Mobility improve coverage of sensor networks. In: Proceedings of ACM MobiHoc (2005)

5. Weiser, M.: In The computer for the twenty-first century. Sci. Am. (1991)
6. Rahimi, M., Shah, H., Sukhatme, G.S., Heideman, J., Estrin, D.: Studying the feasibility of energy harvesting in a mobile sensor network. In: Proceedings of the 2003 IEEE International Conference on Robotics and Automation, Taipei, Taiwan (2003)
7. Yarvis, M., Kushalnagar, N., Singh, H., Rangarajan, A., Liu, Y., Singh, S.: Exploiting heterogeneity in sensor networks. In: Proceedings of IEEE INFOCOM'2005, Miami, FL (2005)
8. Chellappan, S., Bai, X., Ma, B., Xuan, D., Xu, C.: Mobility limited flip-based sensor networks deployment. IEEE Trans. Parallel Distrib. Syst. **18**(2), 199–211 (2007)
9. Lee, U., Magistretti, E.O., Zhou, B.O., Gerla, M., Bellavista, P., Corradi, A.: Efficient data harvesting in mobile sensor platforms. In: Proceedings of PerCom Workshops (pp. 352–356) (2006)
10. Sara, Getsy S., Sridharan, D.: Routing in mobile wireless sensor network: a survey. Telecommun. Syst. **57**(1), 51–70 (2014)
11. Chen, C., Ma, J., Yu, K.: Designing energy efficient wireless sensor networks with mobile sinks. In: Proceedings of WSW'06 at SenSys'06, Colorado, USA, 31 October 2006 (2006)
12. Hameed Mir, Z., Ko, Y.-B.: A quadtree-based hierarchical data dissemination for mobile sensor networks. Telecommun. Syst. **36**, 117–128 (2007). https://doi.org/10.1007/s11235-007-9062-0
13. Felemban, E., Lee, C.-G., Ekici, E.: MMSPEED: multipath multispeed protocol for QOS guarantee of reliability and timeliness in wireless sensor networks. IEEE Trans. Mobile Comput. **5**(6), 738–754 (2006)
14. Sarma, H.K.D., Mall, R., Kar, A.: E^2R^2: energy-efficient and reliable routing for mobile wireless sensor networks. IEEE Syst. J. 1–11 (2014)
15. Kim, D., Chung, Y.: Self-organization routing protocol supporting mobile nodes for wireless sensor network. In: Proceedings of 1st International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06), June (pp. 622–626) (2006)
16. Kumar, G.S., Vinu, M.V., Athithan, P.G., Jacob, K.P.: Routing protocol enhancement for handling node mobility in wireless sensor networks. In: Proceedings of IEEE Region 10 Conference (TENCON) (pp. 1–6) (2008)
17. Awwad, S.A.B., Ng, C.K., Noordin, N.K., Rasid, M.F.A.: Cluster based routing protocol for mobile nodes in wireless sensor network. In: Proceedings of International Symposium on Collaborative Technologies and Systems (CTS'09) (pp. 233–241) (2009)
18. Hayes, T., Ali, Falah H.: Robust ad-hoc sensor routing (RASeR) protocol for mobile wireless sensor networks. Ad Hoc Netw. **50**, 128–144 (2016)
19. The Network Simulator NS-2. http://www.isi.edu/nsnam/ns/index.html

A Resource Allocation Protocol to Meet QoS for Mobile Ad-hoc Network (MANET) in Tactical Scenario

Hemant Kumar Pande, Kapil Kumar Srivastava and Lal Chand Mangal

Abstract In this work, we propose a new Resource Allocation Protocol to meet Quality of Service (QoS) parameters of Transmission Control Protocol/Internet Protocol (TCP/IP) based on mobile ad-hoc networks (MANET). Dynamic slot assignment method for Frequency Hopping (FH) based Time Division Multiple Access (TDMA) environment is proposed. Control slots are allocated to each node to take care of exposed and hidden node problems. The slot allocation protocol is designed for the performance optimization of TCP/IP data. The proposed protocol is implemented on a set up using V/UHF radios and tested for performance of TCP/IP communication for various services. Implementation results clearly show the improvement in the data rate and jitter.

1 Introduction

Networking and mobility are required and essential for anti-insurgency, disaster management, and battlefield operations. Fixed communications infrastructure does not help in this type of networking. This networking should be independent of a fixed infrastructure. A group of mobile nodes without any fixed infrastructure forms a wireless mobile ad hoc network (MANET). In case of a tactical scenario, MANET becomes an obvious choice. Frequency hopping (FH) must be an integral part of MANET because of harsh jamming environment of tactical radios.

In TDMA, one of the most important requirements is the identification of slot boundaries. Timing synchronization is required for the same. It is not possible to build a network using time division multiple access techniques without clock synchronization. In the 802.11 WLAN standard [1], a timing synchronization function

H. K. Pande (✉) · K. K. Srivastava · L. C. Mangal
Defence Electronics Application Laboratory (DRDO), Dehradun, Uttarakhand, India
e-mail: hemant.deal@gmail.com

K. K. Srivastava
e-mail: kapil_ks11@yahoo.com

(TSF) is provided to meet synchronization of clocks in a mobile ad-hoc network. The contention increases with increasing number of mobile hosts. As a result, scalability problem occurs. To solve this problem, an Adaptive Timing Synchronization Procedure (ATSP) was proposed [2]. Tiered Adaptive Timing Synchronization Procedure (TATSP) [3] and Adaptive Bidirectional Timing Synchronization Function (ABTSF) [4] were proposed to solve convergence time and scalability problems. Automatic self-time-correcting procedure (ASP) [5] was proposed for a multi-hop MANET synchronization.

Transmission of adjacent nodes should happen without interference in TDMA. This is done by assigning different time slots to nodes so that collision does not occur at the receivers. Unifying Slot Assignment Protocol (USAP) [6] proposed the same. A Resource Allocation Protocol was proposed in [7] for a tactical scenario in frequency hopping environment. That protocol used ASP as the timing synchronization method.

In [8], a typical battlefield scenario and the services used are given with the required parameters.

A typical tactical scenario presents a difficult problem of waveform medium access control (MAC) layer design which can support multiple services with very low latencies. The QoS requirements of IP data are not fulfilled in the protocol [7].

In this paper, we have proposed some modification in [7] to meet the QoS requirement of IP data. To improve TCP/IP performance, the delay in packet delivery needs to be minimized. In this protocol, slot assignment is done to optimize the latency and jitter of packet transfer. ASP is used for time synchronization as in [7].

The paper is organized as follows. In Sect. 2, an overview of the typical scenario is given. Section 3 gives the details of the proposed protocol design. Section 4 gives the channel allocation method. The implementation and results are given in Sect. 5. Conclusions are given in Sect. 6.

2 Scenario Description

We consider an anti-militant operation in our design. Only the low power portable radios are considered, i.e., Manpack and Handheld. The Leader distributes the soldiers into three units and every unit consists of five soldiers each. So a total of 16 team members along with the leaders are present.

Fig. 1 describes a typical movement of soldiers with single RF channel radios.

3 Waveform Design

The channel access method used in the proposed protocol is TDMA. It is based on the protocol proposed in [7]. The slot allocation protocol is modified for optimization for performance in case of TCP/IP data. The design takes care of hidden and exposed terminal problems. It provides prioritized voice in tactical environment and also
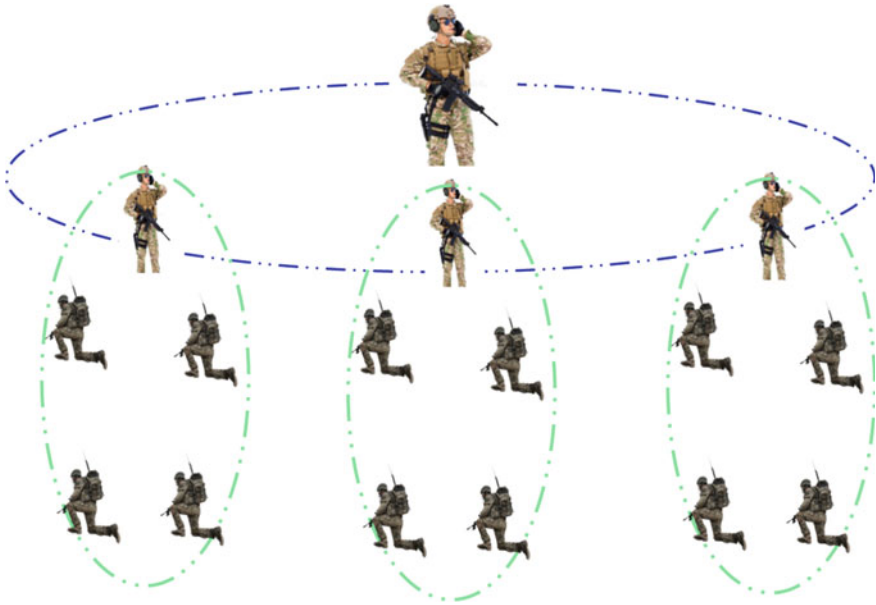
Fig. 1 Typical scenario

enables fast network entry through assignment of control slots to new nodes without contention. This Protocol supports QoS-based allocation of bandwidth for IP data with less delay in packet delivery. The frequency hopping network is designed for 16 nodes with 500 Hops per second.

Each TDMA frame is of 160 ms, which consists of 80 slots. Duration of each slot is 2 ms. A super frame of 2.56 s is formed by combining 16 TDMA frames. The TDMA frame and super-frame structures are shown in Fig. 2.

The structure of each TDMA slot is given in Fig. 3.

Each slot starts with a preamble, which consists of guard time (used for PA ramp up, frequency settling and clock drift), UW (63-bit) for bit synchronization, and
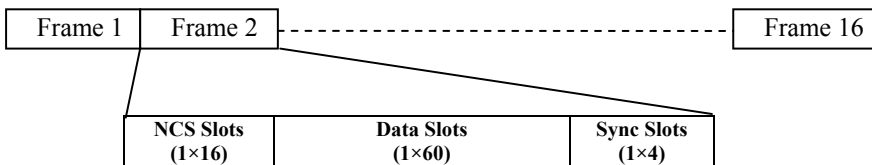
Fig. 2 TDMA frame format

Fig. 3 TDMA slot structure

16 bits for AGC estimation. At the end of packet PA, ramp-down guard time and propagation guard time are given. Remaining time is left for the transmission of the payload. The payload field is divided into three equal blocks of size 55 bytes each. Each block can carry any type of data. Slots are divided into three types which are the following.

Network Control Slot

These slots are used for control information transmission related to slot allocation. Sixteen control slots in every frame are preassigned on the basis of their MAC ID to each node in the network. One block of these slots is used for transmission of the network control packet. Remaining two blocks can be used for transmission of any other type of packet.

Traffic Data Slots

These slots are used for transmission of any type of traffic data (including voice). Each TDMA frame has 60 data slots. The allocation of these slots will be done as per the method described in the subsequent section. All three blocks can be used for transmission of traffic data.

Sync Slots

These slots are used for transmission of FH and timing synchronization information. There are four slots of this type in each TDMA frame. They are allocated in groups of four to a single node in each TDMA frame. To give each node a chance to transmit this information, a complete super frame of 16 TDMA frames is required. One block of this slot is used for FH/Timing synchronization, others two blocks are used for transmission of network routing information.

As described earlier, each slot payload can be divided into three blocks of equal size. The packets transmitted in each block can be divided into four categories: Network Control Packet, Traffic Data Packet, Network Synchronization Packet, and Network Information Packet.

3.1 Network Control Packet (NCP)

Network control packets are used for the transmission of control information related to slot allocation. Sixteen control slots in every frame are preassigned on the basis of their MAC ID to each node in the network. The format of control slot is given in Fig. 4.

The network control packet description is given below .

MAC ID	NTi	NRi	STi	LCi	Ack	CRC

Fig. 4 Network control packet

- MAC ID—Own MAC ID (8 bits).
- NTi—Neighbor's transmitting information.
- NRi—Neighbor's receiving information.
- STi—Slot request for transmitting.
- LCi—Last Collision.
- Ack—Acknowledgement of correct reception in the previous frame.
- Header CRC—32 bit CRC

3.2 Traffic Data Packet (TDP)

The structure of a traffic data packet is given in Fig. 5.
  The field description of traffic data packet slot is given below.

- MAC ID—Own MAC ID (8 bits).
- Port Num—Port Number based on service type (8 bits).
- Next Hop ID—Destination MAC Id for immediate routing (8 bits).
- DST ID—Final Destination MAC Id for which payload has been generated (8 bits).
- TTL—Time to live for a packet (8 bits)
- Service Packet ID—Sequence Number for current service (16 bits).
- Fragment Id—Packet Id used for re-segmentation (8 bits).
- Len—Payload length (8 bits)
- Payload—Data Payload
- CRC—32 bit CRC

3.3 Network Synchronization Packet (NSP)

NSP is being used for transmission of FH/Timing Synchronization. The protocol used in this packet for timing synchronization is based on ASP. Synchronization packets will provide timing information for frequency hopping and TDMA slot synchronization for new nodes entering into the network. Each node will get its dedicated slots in a super frame. Therefore, new node can join the network by listening to the transmitted synchronization slot information through Synchronization Scheme.

MAC Id	Port Num	Next Hop ID	DST ID	TTL	Service ID	Fragment ID	Len	Payload	CRC

Fig. 5 Traffic data packet

Source Id (8 bits)	FH SYNC Info	CRC (32 bits)

Fig. 6 Sync slot format

SRC Id	Network Info Node 1	Network Info Node 2	- - - - - - -	Network Info Node 16	CRC

Fig. 7 Network information packet

The packet structure of a synchronization slot is given in the Fig. 6.

3.4 Network Information Packet (NIP)

One network information packet per frame is allotted to a particular node. In a super frame of 2.56 s, each of the nodes gets a chance to transmit network information packet. Network information packet provides the complete picture of the network to each node. Network information packet structure is shown in Fig. 7.

Every network information packet has a 48-bit field corresponding to each node. Out of 48 bits, 32 bits (2 bits for each link) are for neighbor information, 8 bits for TTL, and 8 bits for packet ID. There will be a total of 512 (32×16) bits. 32 bit CRC will be used for complete packet.

4 Channel Allocation Method

The problem of channel allocation can be divided into two parts:

1. Get the busy slots of other radios.
2. Prepare list of self transmission slots to minimize collision such that QoS requirements can be met.

To achieve this, all the available slots are equally distributed into the number of nodes present. The number of nodes can be extracted from the NIP packet. Each node gets the number of nodes present in the network and its slot position based on its MAC ID. Distribution of slots using this information for a 3-node network is given in Fig. 8.

In case of specific QoS requirement (e.g., video transmission), the user needs to allocate desired capacity using HMI of the radio. Slot reservation service will calculate and broadcast the desired slot capacity needs to be reserved at each node. This

N1	N2	N3	N1	N2	N3	- - - - - - - - - - - - - - - - -	N1	N2	N3

Fig. 8 Slot distribution for 3-node network

information is broadcasted periodically to cater to the changing network topology. On receipt of this information, the nodes will remove the number of reserved slots for the available slots and distribute remaining slots evenly between them. The example of this distribution is given in Fig. 9.

5 Implementation and Results

The proposed protocol is implemented on V/UHF radios. The modem used is π/4 DQPSK with 2 MHz bandwidth. The forward error-correcting codes used is turbo product code (TPC).The system has a frequency hopping rate of 500 hops per seconds. The channel allocation protocol is implemented on ARM 9 32-bit and Analog Devices Blackfin processors. The implementation of ARM 9 uses RTOS as the operating system. Blackfin implementation is non-RTOS based. The test setup is shown in Fig. 10.

The protocol is tested on two radios for data rate and jitter performance. The tests were carried out to analyze data rate and jitter performance on User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) protocols. It was measured using jPerf (Java based GUI for running iPerf) ver 2.0.2.

The performance results are shown in Figs. 11 and 12.

The implementation results verify that the protocol can meet the QoS requirement of the user. It is observed from the results that with the new protocol (Fig. 12a), the bandwidth improves by 50% than the previous protocol (Fig. 12b). Hence the effective data rate improves with the allocation of the data slots to meet QoS.
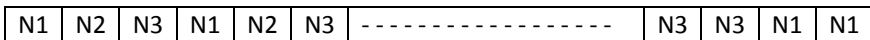
Fig. 9 Slot distribution for 3-node network with QoS slots

Fig. 10 Test setup
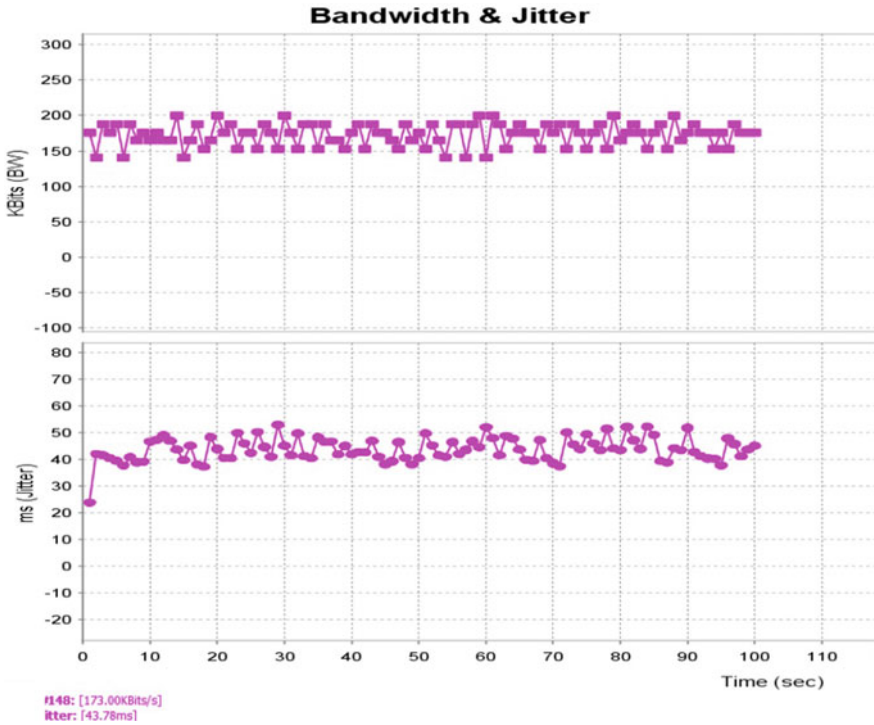
Fig. 11 Data rate and jitter for UDP

6 Conclusion

The proposed protocol is designed for a small network of 16 mobile hosts. The benefit of this method is that each node has some reserve capacity available for transmission, so limited size packet can be transmitted immediately without waiting for slot allocation. It reduces the latency in packet communication and improves the performance of the link.

It is clear from the implementation results that to meet QoS requirement regarding the data rate, the proposed protocol gives the desired output. In the tactical scenario, sometimes the bandwidth requirement is on high priority and the proposed protocol solves this problem.
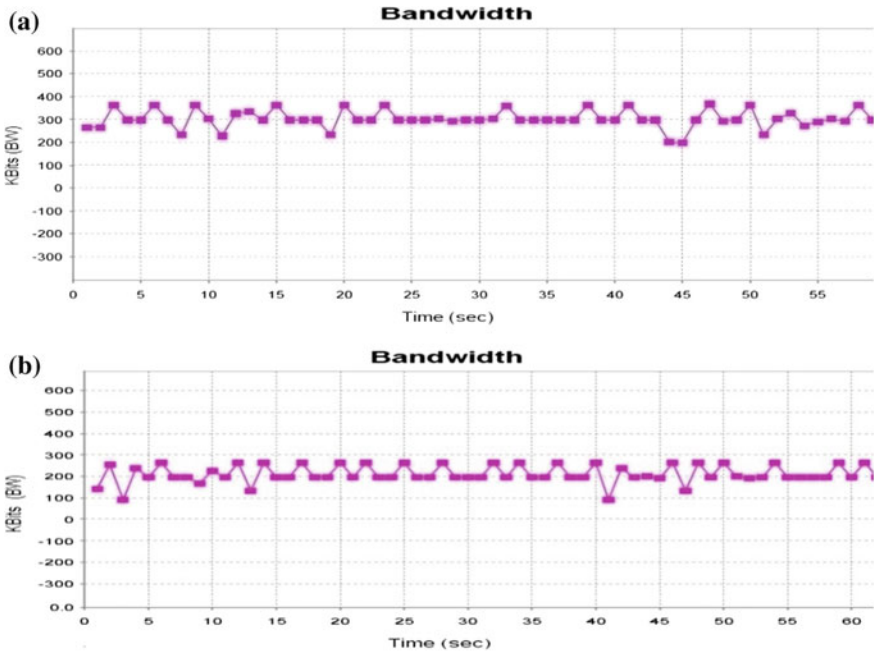
Fig. 12 **a** Data rate for TCP with allocation. **b** Data rate for TCP without allocation

References

1. IEEE Std 802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, 1999 edn
2. Huang, L., Lai, T-H.: On the scalability of IEEE 802.11 ad hoc networks. In: Proceedings of MobiHoc (pp. 173–182) (2002)
3. Lai, T.H., Zhou, D.: Efficient and scalable IEEE 802.11 ad-hoc-mode timing synchronization function. In: Proceedings of the 17th International Conference on Advanced Information Networking and Applications (pp. 318–323) (Mar. 2003)
4. Zhou, D., Lai, T-H.: Analysis and implementation of scalable clock synchronization protocols in IEEE 802.11 ad hoc networks. In: The First IEEE International Conference on Mobile and Sensor Systems (Oct. 2004)
5. Sheu, J-P., Chao, C-M., Sun, C-W.: A Clock Synchronization Algorithm for Multi-Hop Wireless Ad Hoc Networks (pp. 574–581). ICDCS (2004)
6. Young, C.D.: USAP multiple access: dynamic resource allocation for mobile multihop multi-channel wireless networking. In: Proceedings of IEEE MILCOM'99 (Nov. 1999)
7. Pande, H.K., Srivastava, K.K., Thapliyal, S., Mangal, L.C.: New resource allocation protocol for mobile ad-hoc network (MANET) in tactical scenario. In: Proceedings of the 5th IEEE International Conference on Computer and Communication Technology (Sep. 2014)
8. Nakamura, M., Thales-Raytheon System Company: Analysis of alternatives for tactical digital battlefield operation over low-band width legacy radios. In: MILCOM'2008 IEEE (Nov. 2008)

Comparative Study of Anomaly Detection in Wireless Sensor Networks Using Different Kernel Functions

Shashank Gavel, Ajay Singh Raghuvanshi and Sudarshan Tiwari

Abstract Wireless sensor network (WSN) is defined as an autonomous network composed of low power sensor nodes having limited computational, communication, and energy resources. Being short at resources they require efficient use of each resource to prolong network lifetime. Sensor networks are exposed to noise, compromised nodes, low battery levels, and damaged sensors, all these results in anomalous readings or anomaly. Presence of anomaly in system deteriorates the performance of WSN in terms of efficiency, accuracy, and reliability. Hence anomaly detection becomes a major challenge to decide the performance of network. Support vector machine (SVM) is a light weight, learning-based binary classifier that can classify the raw data into normal and anomalous. SVM suffers from computational complexity while handling large datasets, so sequential minimal optimization SVM (SMO-SVM) is used to reduce the complexity. In this paper, a comparative study is made on anomaly detection using SMO-SVM classifier utilizing different kernel functions.

S. Gavel (✉) · A. S. Raghuvanshi
Department of Electronics and Telecommunication, National Institute
of Technology, Raipur, India
e-mail: sgavel.phd2016.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, India
e-mail: stiwari@mnnit.ac.in

81

1 Introduction

Wireless sensor networks (WSNs) is a network made up of small active devices called sensor nodes which are used for monitoring or event detection purpose. Sensor nodes are autonomous low powered devices with sensors that provide different types of sensed data such as humidity, temperature, pressure, and vibrations. Nowadays WSNs have found intensive use in smart cities, smart grid, battlefields, medical sensing, etc. [13]. WSNs are resource constraint networks and the nodes are susceptible to noise, compromised nodes due to intrusion, low battery levels, and damaged sensors giving rise to anomaly in sensed data. These anomalies contribute adversely on the performance and lifetime of network. Hence detection of anomaly in WSN becomes major concern for the efficient use of resources [5].

Anomaly detection techniques that are used for WSN data can be categorized as machine learning, statistical, and signal processing-based approaches. Machine learning approach for detection and classification of anomalies in WSN is gaining a lot of interest by research community [2, 14].

Support vector machine is one of the efficient binary classification technique. The use of SVM for detection and classification has been done in [12]. Although SVM classifies data efficiently sometimes it possesses high computational complexity for the larger datasets. An optimized technique for SVM known as sequential minimal optimization SVM (SMO-SVM) is used in place of simple SVM for improved classification [9].

This paper presents a comparative study of SMO-SVMs classifier by utilizing the different kernel functions. The efficient use of kernel functions is governed by the nature of dataset. The dataset of two standard laboratory has been used to analyze the performance of different kernel functions for SMO-SVMs. The different types of anomaly that exist practically in WSN are manually inserted in each dataset. The performance of anomaly detection is compared using standard performance metrics like accuracy and F1-Measure.

The paper is organized in four sections: Sect. 2 presents literature review and theory of SMO-SVMs, results and discussion are presented in Sect. 3 followed by the conclusion in Sect. 4.

2 Literature Review and Theory of SMO-SVM

2.1 Literature Review

The concept of SVMs has been empirically applied in various fields such as machine learning, pattern recognition, and categorization of text. SVM as learning algorithm work efficiently for low dimension data but sometimes restricts the researcher to train the high dimensional data. So this becomes a problem while using this technique for high dimensional dataset problems. A new optimized algorithm for SVM was

proposed in [9] to overcome the difficulty of simple SVMs and to make the system perform better. This technique was SMO-SVMs which instead provides better results for complex quadratic programming problems. The main motivation behind the use of SMO-SVMs lies in its lesser computational complexity and hence lightweight. SMO-SVM can be used using different learning kernel functions, and to select appropriate kernel function is a major challenge for the classification of different types of anomaly present in WSN [1, 7].

The basic idea behind the SVMs is to fit the data into a hyperplane or hypersphere between different classes. The hyperplane is subspace in which data is fitted accordingly and the dimension is less than its ambient space. This hyperplane creates a large separation between the different classes of data. The method involves mapping of data into a higher dimensional space to make separation easier. After the data is mapped, kernel functions are applied for approximating the dot products between the mapped vectors into the feature space to find the hyperplane. This helps better in identifying the class of normal data with the anomalous data.

2.2 Sequential Minimal Optimization SVM

Here we consider a hypersphere in place of hyperplane, for SVM a set of training data is considered in a feature space, $X = (x_1, x_2, x_3, \ldots, x_n)$ where $x_i \in R^d$ $(1 \le i \le n)$ represents the d-dimensional data and n is the size of the training data. The data from the feature space is then trained. Considering this the optimization problem to be solved is given below:

$$\min R^2 + C \sum_{i=1}^{n} \varepsilon_i \tag{1}$$

$$s.t. \quad ||x_i - a||^2 \le R^2 + \varepsilon_i, i = 1, 2, \ldots, n.$$
$$\varepsilon_i \ge 0, i = 1, 2, \ldots, n.$$

where a and R are the center and radius of the hypersphere, respectively, in the feature space, ε_i is the slack variable which allows few training data outside the hypersphere and the penalty parameter C controls the trade-off between the number of target data outside and volume of the hypersphere [10, 12]. The mapping function $\phi(.)$ is used to map the data of input class to feature space in terms of $\phi(x_i)$.

This mapping variable $\phi(x_i)$ replaces the value of x_i. This allows the function to calculate the inner product of two vectors in feature space. The inner product is given by

$$K(x_i, x_j) = \phi(x_i) * \phi(x_j) \tag{2}$$

Although simple SVM is computed in the form of linear classifier where x is input and y is output, the objective function representing the linear classifier becomes

$$f(x) = w^T x + b \tag{3}$$

where w is a normal vector and b represents the threshold value. Since we are using binary classification to judge the anomalous and non-anomalous data, the value prediction will be in the terms of $y = 1$ if $f(x) \geq 0$ and $y = 0$ if $f(x) < 0$. By considering the inner product as given in (2), the function in (3) is given by

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b \tag{4}$$

where α_i represents the Lagrangian multiplier. Lagrange multiplier helps in finding the local maxima and minima of provided function [4, 6].

To optimize the above function Lagrangian multiplier plays an important role. This Lagrangian multiplier is to be optimized, and the constrained minimization problem is to be solved. The initial value of lagrangian multiplier which is to be updated is shown below:

$$\alpha_j = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} \tag{5}$$

In the above Eq. (5) E_k represents the error on kth value of training example and η is the second order derivative of objective function. The individual representation of the variable are shown below:

$$E_k = f(x_k) - y_k \tag{6}$$

$$\eta = 2K(x_i, x_j) - K(x_i, x_j) - K(x_j, x_k) \tag{7}$$

Thus depending upon the value (5) the new value of final Lagrangian multiplier becomes α_i, below value shows the new optimized value of the multiplier:

$$\alpha_i = \alpha_i + y_i y_j (\alpha_{j-1} - \alpha_j) \tag{8}$$

So the optimization problem is solved by replacing the new value of α_i from (8) to (4).

In this paper we have basically used the kernel functions value as mentioned in (2) for different kernels which are shown below [3]:

1. Gaussian Kernel

$$K(x_i, x_j) = e^{\gamma ||x_i - x_j||^2}$$

2. Linear Kernel

$$K(x_i, x_j) = x_i^T x_j$$

3. Sigmoid Kernel

$$K(x_i, x_j) = (\gamma (x_i^T x_j + 1))^d$$

4. Polynomial (poly3) Kernel

$$K(x_i, x_j) = \tan(\gamma x_i^T x_j + 1)$$

These four kernel functions are most commonly used as they perform better for working set selections.

3 Results and Discussion

For the analysis of suitability of kernel function with anomalous data, we have used SMO-SVM as a classifier to our paper. Data conditioning of data set from Labeled Dataset collection [11] and IBRL Dataset [8] is done, and the different types of anomalies are inserted to analyze the compatibility of different kernel function. Following are the datasets taken for analysis:

1. Multihop and Singlehop Datasets [11]:
   This dataset is taken from the network containing both multihop and singlehop scenario for anomaly detection. Both the multihop and singlehop network is setup and depending upon that the readings from the sensor node are taken.
2. IBRL Datasets [8]:
   This dataset was taken from the well-known Intel Berkeley Lab experiment where 54 sensors were deployed. We have managed to take and conditioned the data from $node8$ and $node9$ as they are close to each other and shows similar behavior.

   Following types of noise are inserted to the dataset:

1. Random Noise:
   This type of anomaly occurs where the sensor node supplies data and at the same time transient disturbance happens depending upon the random time distribution. This anomaly is used in multihop datasets from [11] example is shown in Fig. 1.
2. Regenerative Feedback Noise:
   This type of anomaly occurs in the network when the data from the sensor node
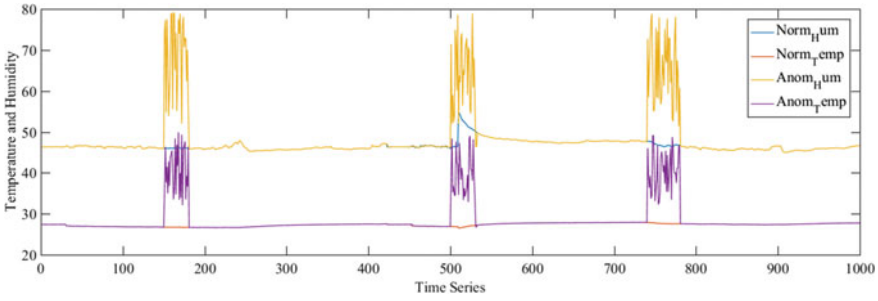
Fig. 1 Normal and anomalous dataset with respect to time series
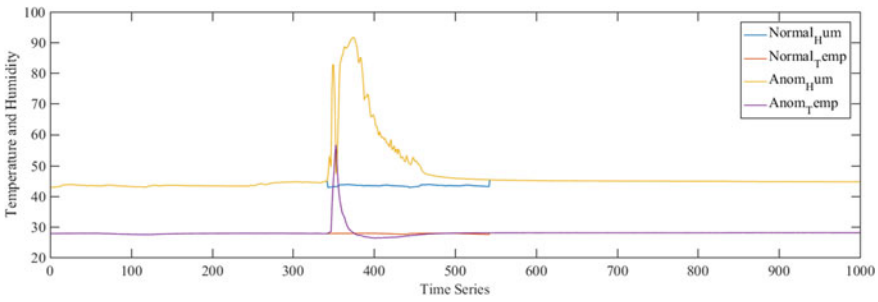
Fig. 2 Normal and anomalous dataset with respect to time series
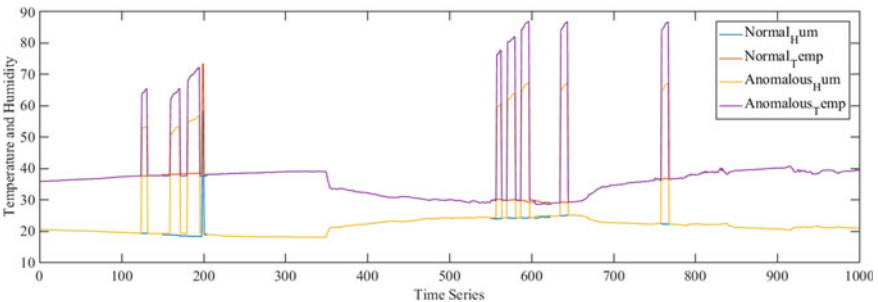
Fig. 3 Normal and anomalous dataset with respect to time series

keep on increasing as compared to the normal data. This anomaly is used in singlehop datasets from [11] example is shown in Fig. 2.

3. Shot Noise(Spikes):
   This anomaly occurs when the sensor data shows small spikes between the data. This anomaly is inserted to Node8 and Node9 of dataset from [8], example is shown in Fig. 3.

The comparison of the kernel functions are performed on the basis of performance metrics shown in (9) and (12). In which g is the g-means accuracy and $F1 - Measure$ shows the balance between the recall and precision value obtained from the dataset.

$$g = \sqrt{Acc_+ * Acc_-} \tag{9}$$

where

$$Acc_+ = \frac{No. \ \ of \ \ target \ \ samples \ \ correctly \ \ classified}{Total \ \ number \ \ of \ \ target \ \ samples} \tag{10}$$

$$Acc_- = \frac{No. \ \ of \ \ nontarget \ \ samples \ \ correctly \ \ classified}{Total \ \ number \ \ of \ \ nontarget \ \ samples} \tag{11}$$

and the F1-Measure is given by

$$F1 - Measure = \frac{2 * Precison * Recall}{Precison + Recall} \tag{12}$$

The graph is plotted in terms of accuracy and F1-Measure and this result shows the compatibility of different kernel functions with SMO-SVM classifier in terms of different kinds of anomalies. The comparison of results in terms of performance metrics can be shown in the figures. Figure 4 (a, b) shows the accuracy and F1-Measure for SMO-SVM method of multihop dataset, Fig. 5 (a, b) shows the accuracy
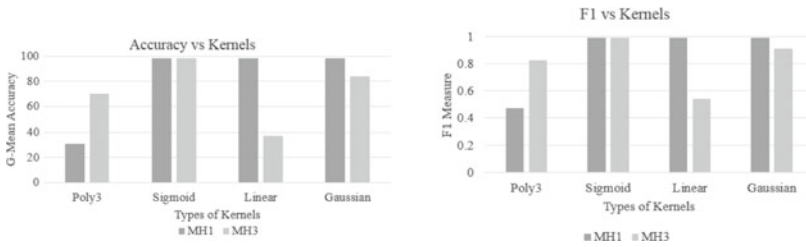
(a) Accuracy for Multihop dataset   (b) F1-Measure for Multihop dataset

Fig. 4 Accuracy and F1-Measure for multihop dataset
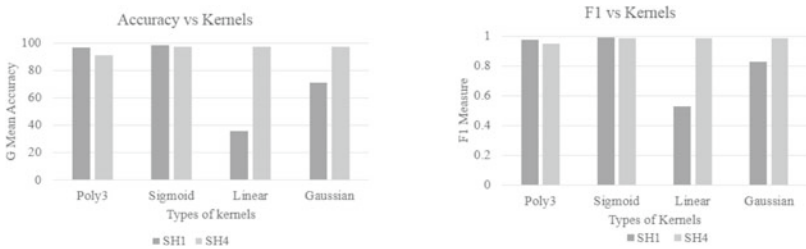
(a) Accuracy for SingleHop dataset   (b) F1-Measurefor SingleHop dataset

Fig. 5 Accuracy and F1-Measure for singlehop dataset

(a) Accuracy for IBRL dataset (b) F1-Measure for IBRL dataset
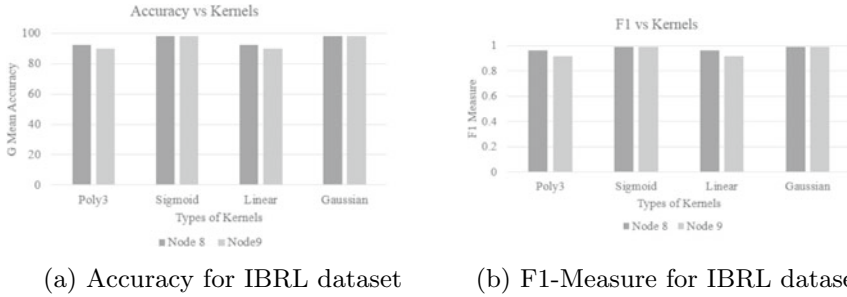
Fig. 6 Accuracy and F1-Measure for IBRL dataset

and F1-Measure for SMO-SVM method of singlehop dataset, and Fig. 6 (a, b) shows the accuracy and F1-Measure for SMO-SVM method of IBRL dataset. However the system performance for *Random Noise* the Sigmoid and Gaussian kernel show better compatibility with classifier, for *Regerative feedback Noise* the Sigmoid and poly3 kernel show better compatibility with classifier and for *Shot Noise* the all the four kernels show better compatibility with classifier. So it can be concluded that the accuracy of detection depends upon the selection of kernel function.

4   Conclusion

In this article we have used SMO-SVM as a detector and classifier for the WSN datasets. Three different types of noise are inserted in the datasets as explained in the above section, and to analyze this anomalous dataset various kernel functions are used with the SMO-SVM classifier. By analyzing the results obtained we may conclude that for network prone to random noise Sigmoid and Gaussian kernel function are better choice. For network prone to regenerative feedback noise sigmoid and poly3 kernel functions works better. For network prone to shot noise any of the kernel function among the four can be used. We have successfully compared the importance of kernel functions with the classifier and its compatibility for different anomalous conditions. In future we will be using this classifier to construct an Anomaly Detection System.

References

1. Branch, J.W., Giannella, C., Szymanski, B., Wolff, R., Kargupta, H.: In-network outlier detection in wireless sensor networks (2013)
2. Du, D., Qi, B., Fei, M., Wang, Z.: Quantized control of distributed event-triggered networked control systems with hybrid wired–wireless networks communication constraints (2017)

3. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using second order information for training support vector machines (2005)
4. Kang, W.S., Choi, J.Y.: Domain density description for multiclass pattern classification with reduced computational load (2008)
5. Leccese, F., Cagnetti, M., Trinca, D.: A smart city application: a fully controlled street lighting isle based on raspberry-pi card, a zigbee sensor network and wimax (2014)
6. Lee, S.W., Park, J., Lee, S.W.: Low resolution face recognition based on support vector data description (2006)
7. Moshtaghi, M., Leckie, C., Karunasekera, S., Rajasegarar, S.: An adaptive elliptical anomaly detection model for wireless sensor networks (2014)
8. Peter, B., Wei, H., C.G.S.M.M.P., Thibaux, R.: Ibrl dataset. http://db.csail.mit.edu/labdata/labdata.html
9. Platt, J.: Sequential minimal optimization: a fast algorithm for training support vector machines (1998)
10. Rajasegarar, S., Leckie, C., Palaniswami, M.: Hyperspherical cluster based distributed anomaly detection in wireless sensor networks (2014)
11. Suthaharan, S., Alzahrani, M., Rajasegarar, S., Leckie, C., Palaniswami, M.: Labelled data collection for anomaly detection in wireless sensor networks (2010)
12. Tax, D.M., Duin, R.P.: Support vector data description (2004)
13. Xie, M., Han, S., Tian, B., Parvin, S.: Anomaly detection in wireless sensor networks: a survey (2011)
14. Zamani, M., Movahedi, M.: Machine learning techniques for intrusion detection (2013)

Proactive Spectrum Handoff-Based MAC Protocol for Cognitive Radio Ad hoc Network

T. Akhil Chandran, Raghavendra Pal, Arun Prakash and Rajeev Tripathi

Abstract Cognitive radio technology significantly delivers solution to the problem of congested spectrum by exploiting radio spectrum unused by licensed users. The cognitive radio network consists of a number of channels and a group of unlicensed secondary users (SUs), which are ready to access the channel when the licensed primary users (PUs) are idle. To improve the performance of the cognitive radio network, spectrum handoff can be considered. This paper presents a proactive spectrum handoff-based medium access control protocol that is designed for cognitive radio ad hoc networks. The proposed protocol ensures to enable the SUs of the cognitive network for better accessibility and more flexibility with better performance.

Keywords Cognitive radio · Ad hoc network · MAC · Proactive handoff

1 Introduction

The usage of wireless application and devices are improving rapidly in our daily life. Hence there is an increasing demand in the spectrum resources. In the coming future, this might result in spectrum shortage and overcrowding. Cognitive radio technique allows a group of unlicensed secondary users (SUs) to resourcefully access channels originally allotted to licensed primary users (PUs) which provides significant improvement in the spectrum utilization.

T. Akhil Chandran · R. Pal (✉) · A. Prakash · R. Tripathi
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad 211004, India
e-mail: raghavendra.pal3@gmail.com

T. Akhil Chandran
e-mail: akhilchandrant@gmail.com

A. Prakash
e-mail: arun@mnnit.ac.in

R. Tripathi
e-mail: rt@mnnit.ac.in

To improve the performance of the cognitive radio network, spectrum handoff is considered. There are two types of spectrum handoff methods available in cognitive radio network, namely, reactive and proactive. In reactive spectrum handoff whenever handoff is requested by a SU, spectrum sensing is initiated to find an idle channel for SU to continue its unfinished transmission [1]. This method does not require the channel usage history details but it adds delay in order to find the idle channel for handoff. In proactive spectrum handoff SU predicts the arrival of a PU in the channel currently used by the SU and accordingly makes a decision for performing handoff. Then, SU switches to a new channel before the PU occupies the channel [2].

The essential work of controlling and synchronizing communication over wireless channels occur largely at the medium access control (MAC) layer and hence designing an efficient MAC protocol becomes a prime focus for successful implementation of any CR ad hoc network. The access toward channels for SUs is discontinuous in the sense that whenever the PUs are inactive, channels are available for SUs, and they can switch from any channel to another. The number of available channels depends on the time; and the number of available channels for each SU is different from each other. Since SUs are accessing channels used by PUs, the activity of SUs can create a certain amount of interference to the PU transmission. In this paper, a proactive spectrum handoff technique based on MAC protocol is proposed to avoid the interference to PUs and to resolve multichannel hidden terminal problem.

The remainder of the paper is organized as follows. Related work and previous studies of the MAC protocol and its classifications from the point view of the interference to PU and network reconfiguration overhead is presented in the Sect. 2. Section 3 presents the proposed system model and problem statement. Section 4 describes the operation of the proposed proactive spectrum handoff-based MAC protocol (PHMP). Extensive simulation analysis is carried out to evaluate the performance of PHMP, which is presented in Sect. 5. Finally, Sect. 6 concludes the paper and points out some future work.

2   Related Works

Based on the interference to PUs, multichannel hidden terminal problem and network reconfiguration overhead, it is possible to categorize the existing MAC protocols. The level of interference to PUs is an important parameter to consider while designing MAC protocol for CR network. While designing, it is good to address multichannel hidden terminal problem in all MAC protocols. This can be realized by exchanging certain control messages. The network reconfiguration is low when the MAC protocol uses dedicated common control channel, whereas it becomes moderate or high when the protocols use dynamic common control channel. In SYNMAC [3] interference to PUs is not considered, but multichannel hidden terminal problem is addressed and it experiences moderate network reconfiguration overhead. In DC-MAC [4] both interference to PUs and multichannel hidden terminal is addressed but it incurs high network reconfiguration overhead. Both interference to PUs and

multichannel hidden terminal is not addressed in O-MAC [5], but the QoS provision is addressed. In C-MAC [6] interference to PUs is not considered but multichannel hidden terminal problem is addressed and the network reconfiguration overhead is high. In HC-MAC [7] also, interference to PUs is not addressed but multichannel hidden terminal problem is considered, however, it had low network reconfiguration overhead. Both interference to PUs and multichannel hidden terminal is not addressed in CogMesh [8] and the network reconfiguration overhead is moderate, whereas in DOSS [9] interference to PUs is not considered and the multichannel hidden terminal problem is addressed but it is insufficient to handle the multichannel hidden terminal problem, it had moderate network reconfiguration overhead. In OSAMAC [10] both interference to PUs and multichannel hidden terminal is considered and it had low network reconfiguration overhead. From the analysis, it is clear that interference to PUs and multichannel hidden terminal problem should be addressed while designing a MAC protocol for CR ad hoc network. Hence to avoid the interference to PUs and multichannel hidden terminal problem, a proactive spectrum handoff technique is considered. Sensing outcomes are proactively monitored in CCMAC [11]. It provides immediate channel access to safety messages. Impact of protocol on safety and non-safety messages is clearly observed but infotainment messages are deprived of channel in case of high-safety traffic. MOCA [12] uses vehicles' direction, speed, etc. to predict the channel to be allotted to vehicles. It reduces frequent disconnections due to high mobility using cognitive radio technology.

3 Problem Statement and System Model

In cognitive radio due to the sudden changes in PU activities, the SUs have to frequently switch from one channel to another to continue their data transmission. This reduces the throughput of the network and increases average delay of the cognitive network. Figure 1 depicts the CR ad hoc network scenario in which a number of unlicensed SUs are ready to access the channels used by licensed PUs for data transmission, when the PUs become idle. Based on the requirement any two SUs can randomly form a SU pair (transmitter–receiver pair of secondary users) to transmit information between them.

In [13], the authors proposed a protocol in which each SU pair is allotted with a unique channel, i.e., channel A is allotted only for pair 1, channel B is allotted to pair 2 and so on. This method might reduce collision of data between the pairs, but when PU is active, the selected channel for a given pair is given back to the PU and the current secondary activity is paused. The pausing took place because a unique channel was allotted to individual pairs and hence there was no option of switching to an unused channel allotted to a different SU pair for continuing its data transmission.

It is inferred that multichannel hidden terminal problem and interference affect the efficient communication between SU pairs. Hence the main objective of this work is to handoff the SU activity when the ongoing data transmission is paused due to the arrival of PU and address the multichannel hidden terminal problem. Hence in this

Fig. 1 Scenario depicting PU and SU in the network

paper, a new protocol for improved data transfer in terms of throughput and average delay is proposed.

An ad hoc network is considered in the scenario and split phase method [13] is used to send control packets and data packets using the available channel in the network. This work is mainly concentrated on efficient data channel selection when there is PU activity so that the SUs are able to continue with their data transmission using handoff strategy.

4 Operation of the Proposed Protocol

SU pairs are formed using the random pairing algorithm [13]. Hence a particular user can connect to any other user and form a transmitter–receiver pair. Using the MAC protocol [13] channels for signaling phase and initial data transmission phase are assigned and with the help of cognitive radio concept, SUs are accessing the licensed channel depending upon the absence of PUs in their neighborhood. Hence the SU pair can exchange the data as long as the selected channel is available to them.

When the PU is active, the ongoing SU transmission is paused and PU gets access to the channel for its data transmission. Hence the SU has to wait for a time period (depends on PU activity) to get the selected channel access to continue its transmission. In order to maintain the SU transmission without any discontinuity a proactive spectrum handoff strategy is considered in this paper as described in Figs. 2 and 3. In Fig. 2b, PU is not active hence SUs are transmitting the data on channel A but in Fig. 2c, when PU becomes active the transmission between SUs is halted and in Fig. 2d, SUs transmits the data using channel B since PU is using the channel A.
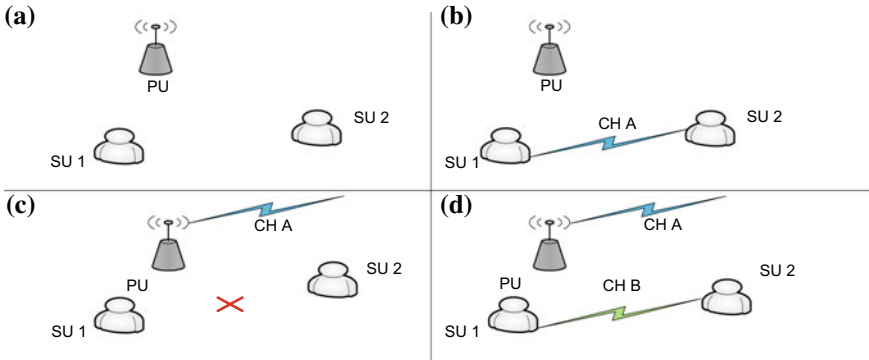
Fig. 2 **a** PU and SU are idle. **b** PU is idle, SU pair transmitting data in channel A. **c** PU is active, and SU stops transmitting data in channel A. **d** SU pair handoff data transmission to channel B
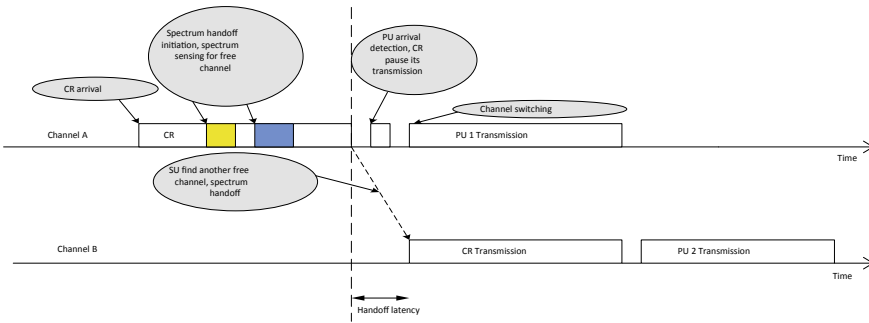
Fig. 3 Proactive spectrum handoff strategy

Algorithm for channel allocation during handoff is described in the following steps.

Algorithm Channel allocation during handoff	
Step 1	Initiate handoff strategy when the PU is arrived or active
Step 2	Initialize a count
Step 3	Compute the size of Common channels available for SU pair
Step 4	Select the next best channel from the list of common channels
Step 5	If the selected channel is free then continue the data transmission
Step 6	Else increment the count
Step 7	SU pair continue the data transmission on the selected channel

Finally, the flowchart of the protocol is proposed and presented in Fig. 4. When the PU becomes active, the particular SU pair which is in the range of PU has to

Table 1 CLC of the communication pairs

Pair order	SU pair	Common list of channels (CLC)			
1	[1, 3]	**B**	D	A	
2	[5, 8]	**C**	A		
3	[2, 6]	B	**D**	C	A
4	[4, 7]	D	C	**A**	

Bold letters show the selected channel for corresponding pair

Table 2 Channels assigned to each pair

Pair order	1	2	3	4
Selected channel	B	C	D	A

switch back the licensed spectrum to the licensed user. For example, consider Table 2 where the pair 1 is allotted with channel B to transmit data. When the PU which owns the license of channel B is turned ON, pair 1 has to initiate handoff and check for the next best channel from the corresponding list of channels common to the SU pair. In Table 1 the next best channel is channel D. If it is free, pair 1 can continue the data transmission in channel D. If channel D is not available, then check for the next best channel and so on. If none of the channel is available for handoff wait for the PU to turn OFF and then continue the data transmission.

4.1 Method to Check the Selected Channel Is Free or Not

In Fig. 5a, b, and c t_0, t_2 … represent PU arrival time whereas t_1, t_3 … represent PU departure time. "t" indicates the time at which handoff is initiated. The selected channel is available for SU communication, if "t" is between t_1 and t_2 and the energy between t_1 and t_2 is below the threshold value "Th." From Fig. 5a, it is clear that the channel is free and it can be used to continue the ongoing transmission of SU. t_2–t indicates the void time in which the SU can access the selected channel without any interference from PU. Whereas in Fig. 5b, the energy between t_1 and t_2 is above the T_h, signifying that the corresponding channel is used by other SU. In Fig. 5c, "t" is less than t_1, therefore the channel is accessed for PU activity.

5   Performance Evaluation

In this section, to evaluate the performance and adequateness of the proposed proactive handoff-based MAC protocol, throughput and average delay of the SUs are computed. Simulations are performed using ns-2 simulator version 2.31 with CRAHN

```
┌─────────────────────┐
│   ON GOING DATA     │
│   TRANSMISSION      │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   PRIMARY USER      │
│   ARRIVAL           │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   INITIATE          │
│   HANDOFF           │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│       K = 0         │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   L = SIZE OF       │
│   COMMON LIST       │
│   OF CHANNEL        │
└─────────────────────┘
          │
          ▼
┌─────────────────────┐
│   SELECT NEXT       │
│   BEST CHANNEL      │
│   FROM COMMON       │
│   LIST OF           │
│   CHANNEL           │
└─────────────────────┘
          │
          ▼
      IS SELECTED
      CHANNEL FREE?
          │
          ▼ YES
┌─────────────────────┐
│   DATA              │
│   TRANSMISSION      │
│   ON SELECTED       │
│   CHANNEL           │
└─────────────────────┘
          │
          ▼
        (  END  )
```

WAIT FOR PU TO BE TURN OFF

YES

NO

IF K < L

K = K+1

NO
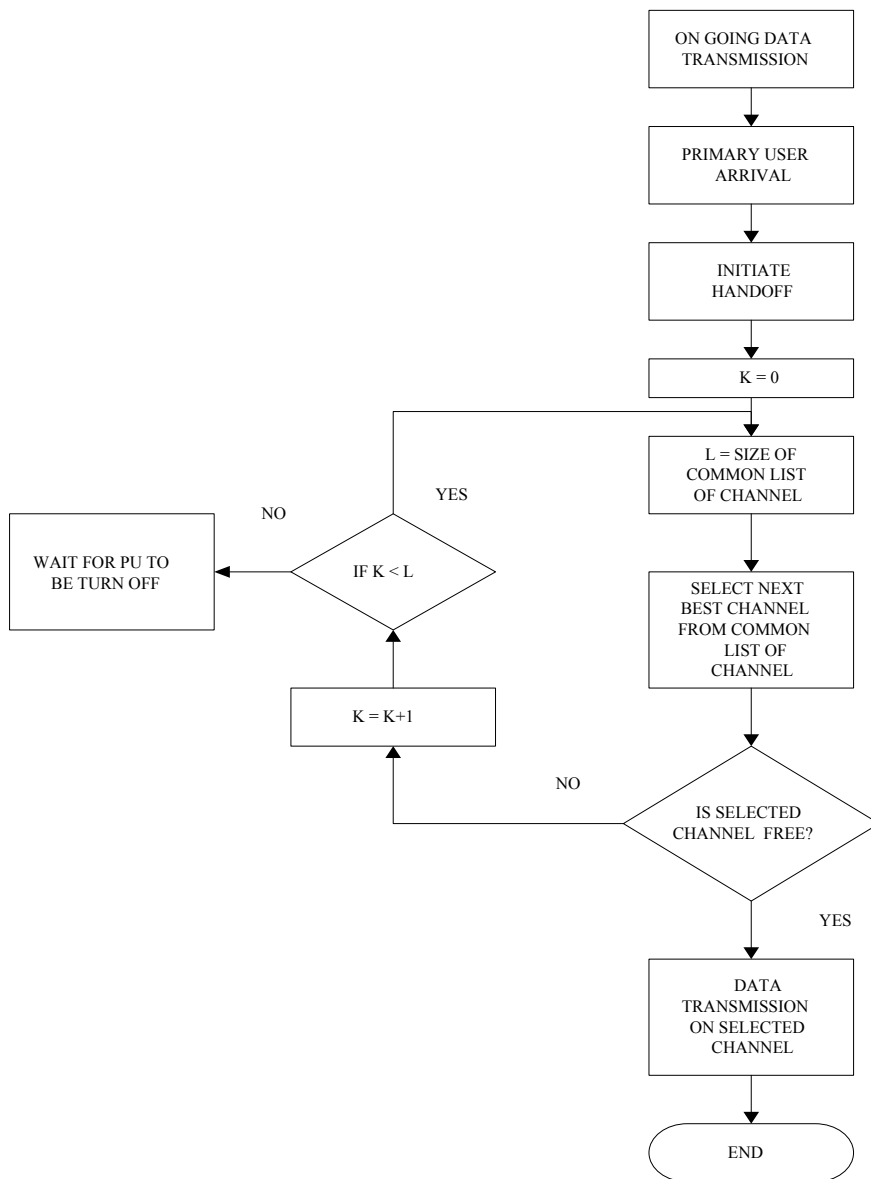
Fig. 4 Operation of the proposed handoff-based MAC protocol
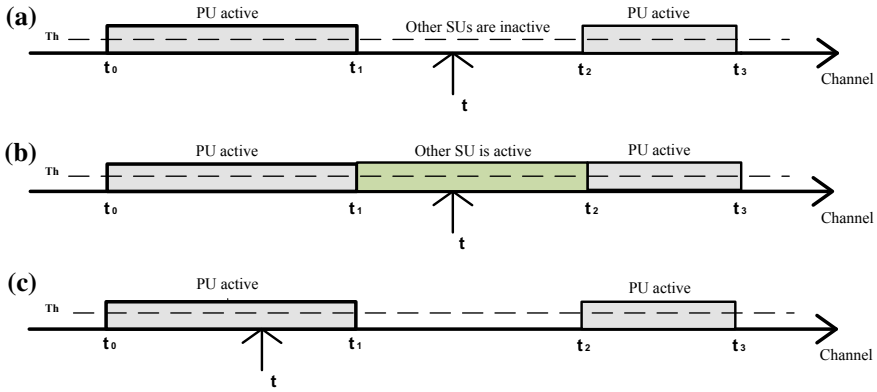
Fig. 5 **a** Selected channel is free at "t." **b** Selected channel is not free at "t" because it is already accessed by other SU. **c** Selected channel is not free at "t" because the PU is active

patch [14]. The simulation parameters of the proposed protocol are presented in Table 3.

A total of 8 SUs are deployed in a region of size 2000 m × 2000 m. The data transmission rate is set to 448 Kbps. Node speed is fixed to 15 m/s. The simulation time is 50 s. To facilitate the simulation in a reliable environment, we use two-ray ground fading model as the signal propagation model. 10 PUs are considered for the scenario. Different PU activity models have been proposed for CRN [15, 16] which include Markov process, queuing theory, time series, and ON/OFF periods. Here the static ON/OFF model is used where the PU spectrum occupancy statistics are known in advance. In reality, PU activities are more complex and cannot be fully modeled with ON/OFF period alone [15, 16].

Figure 6a shows the comparison between the throughput PHMP with the MAC protocol by Sukkar et al. [13]. The number of channels and SUs are 4 and 8, respec-

Parameters	Values
Simulation area	2000 m × 2000 m
No. of secondary users	8
Speed of secondary users (m/s)	15
Transmission range of SUs (m)	100
No. of channels	3, 4
No. of primary users	10
Coverage area of primary users (m)	100
Data transmission rate (Kbps)	448
Packet size (Bytes)	100
Simulation time (s)	50

Table 3 Simulation parameters
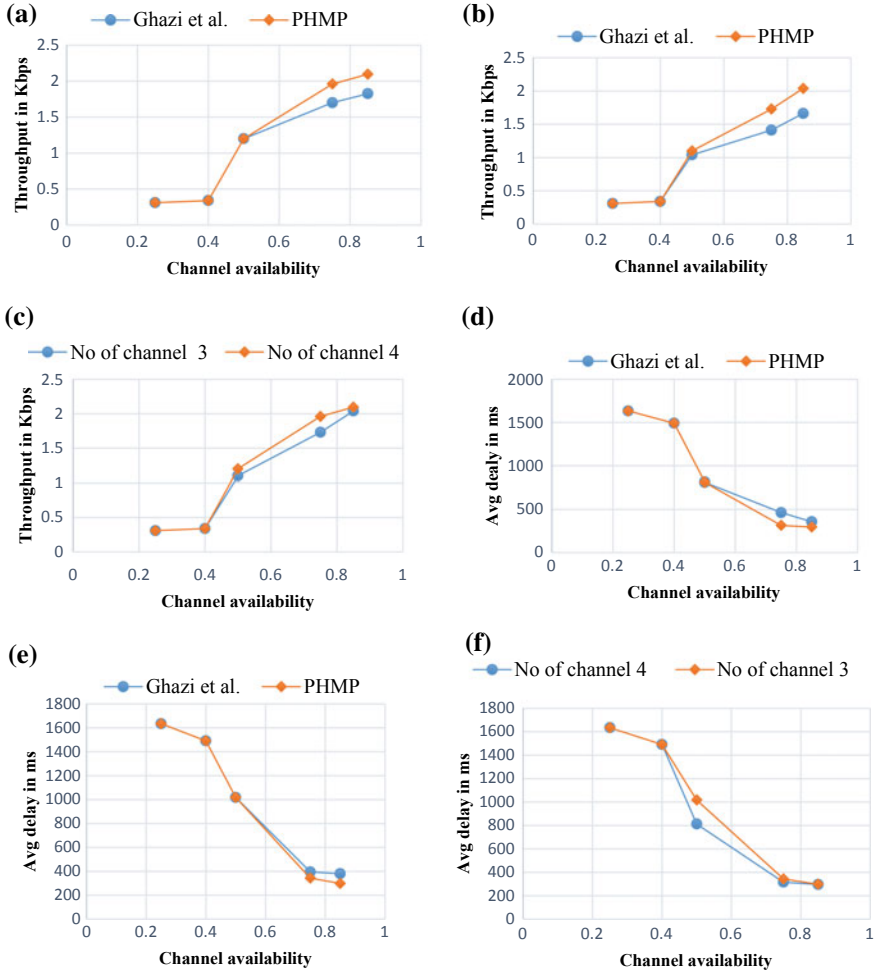
Fig. 6 **a** Throughput versus channel availability for the network with four channels Fig. 6. **b** Throughput versus channel probability for the network with three channels Fig. 6. **c** Throughput versus channel availability for PHMP Fig. 6. **d** Average delay versus channel availability for the network with four channels Fig. 6. **e** Average delay versus channel availability for the network with three channels Fig. 6. **f** Average delay versus probability for PHMP

tively. It's evident that the proposed protocol provides better throughput than the other protocol when the channel availability is greater than 0.5. Below 0.5 both the protocols provide same values of throughput because the channel availability is decreased and there is no channel to perform handoff strategy. Figure 6b shows the comparison between the throughput of PHMP with the MAC protocol by Sukkar et al. [13]. The number of channels and SUs are 3 and 8, respectively. Since the channel is reduced from four to three, there is a decrement in the throughput value.

The proposed protocol shows better results because of handoff strategy that reduces the loss of packets and thus increases the throughput.

Figure 6c gives the throughput of the cognitive network with eight SUs and number of channels available to the network as 3 and 4, respectively. When the channel availability is high, throughput increases whereas throughput is low when the channel availability is less. Figure 6d shows the comparison between the average delay of PHMP with the MAC protocol by Sukkar et al. [13]. The number of channels and SUs are 4 and 8, respectively. It is evident that the proposed protocol provides less average delay than the other protocol when the channel availability is greater than 0.5. Below 0.5 both the protocol is providing similar average delay because the channel availability is decreased and there is no sufficient channel to perform handoff strategy. Hence there is no difference between the proposed protocol and protocol by Sukkar et al. [13]. The proposed PHMP performs better since handoff makes data not to wait until the channel becomes free. Due to handoff, data is transmitted using a different channel.

Figure 6e shows the comparison between the average delay of PHMP with the MAC protocol by Sukkar et al. [13]. The number of channels and SUs are 3 and 8, respectively. Since the channel is reduced from 4 to 3, there is increase in the average delay. Figure 6f gives the average delay of the cognitive network with eight SUs and number of channels available to the network as 3 and 4, respectively. When the channel availability is high, average delay is low whereas average delay is high when the channel availability is less.

6   Conclusion

This paper proposes a proactive spectrum handoff-based MAC protocol (PHMP) with enhanced user connectivity in wireless cognitive radio ad hoc network. PHMP can initiate handoff for SU when the nearby PU becomes active. Besides this, PHMP makes better use of all available channels effectively. The results from simulation analysis indicate that the proposed PHMP provides better throughput and lesser average delay for the cognitive radio network when the channel availability is greater than 0.5. It is worth implying that this improvement is achieved when at least half the channels are free from PU activity and can be accessed by the unlicensed SUs. Future research direction may include QoS support for SUs, such that priorities can be assigned to SU pairs for channel selection based on the type of data they are transmitting.

References

1. Wang, C.W., Wang, L.C., Adachi, F.: Modeling and analysis for reactive-decision spectrum handoff in cognitive radio networks. In: IEEE Global Telecommunications Conference (GLOBECOM 2010), Tuttle South, USA, pp. 1–6 (2010)
2. Wang, C.W., Wang, L.C.: Modeling and analysis for proactive-decision spectrum handoff in cognitive radio networks. In: IEEE International Conference on Communications (ICC '09), Dresden, Germany, pp. 1–6 (2009)
3. Kondareddy, Y., Agrawal, P.: Synchronized MAC protocol for multihop cognitive radio networks. In: Proceedings IEEE International Conference on Communications (ICC '09), Beijing, China, pp. 3198–3202 (2008)
4. Zhao, Q., Tong, L., Swami, A., Chen, Y.: Decentralized cognitive mac for opportunistic spectrum access in ad hoc networks: a POMDP framework. IEEE J. Sel. Areas Commun. **25**(3), 589–600 (2007)
5. Su, H., Zhang, X.: Cross-layer based opportunistic MAC protocols for QoS provisioning's over cognitive radio wireless networks. IEEE J. Sel. Areas Commun. **26**(1), 118–129 (2008)
6. Cordeiro, C., Challapali, K.: C-MAC: a cognitive MAC protocol for multi-channel wireless networks. In: Proceedings IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN'07), Dublin, Ireland, pp. 147–157 (2007)
7. Jia, J., Zhang, Q., Shen, X.: HC-MAC: a hardware-constrained cognitive MAC for efficient spectrum management. IEEE J. Sel. Areas Commun. **26**(1), 106–117 (2008)
8. Chen, T., Zhang, H., Mario, G., Chlamtac, M.I.: CogMesh: a cluster-based cognitive radio network. In: Proceedings IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN'07), Dublin, Ireland, pp. 168–178 (2007)
9. Ma, L., Han, X., Shen, C.C.: Dynamic open spectrum sharing MAC protocol for wireless ad hoc networks. In: Proceedings IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN'07), Baltimore, MD, USA, pp. 203–213 (2005)
10. Le, L., Hossain, E.: OSA-MAC: a MAC protocol for opportunistic spectrum access in cognitive radio networks. In: Proceedings IEEE Wireless Communications and Networking Conference (WCNC'08), Las Vegas, NV, USA, pp. 1426–1430 (2008)
11. Gupta, N., Prakash, A., Tripathi, R.: Clustering based cognitive MAC protocol for channel allocation to prioritize safety message dissemination in vehicular ad-hoc network. Veh. Commun. **5**, 44–54 (2016)
12. Silva, C., Nogueira, M., Kim, D., Cerqueira, E., Santos, A.: Cognitive radio based connectivity management for resilient end-to-end communications in VANETs. Comput. Commun. **79**, 1–8 (2016)
13. Sukkar, G.A., Al-Damluji, M.N., Mansour, I.: A MAC protocol with an improved connectivity for cognitive radio networks. EURASIP J. Wirel. Commun. Netw. **116**, 1–9 (2016)
14. Di Felice, M., Chowdhury, K.R., Kim, W., Kassler, A., Bononi, L.: End-to-end protocols for cognitive radio ad hoc networks: an evaluation study. Perform. Eval. **68**(9), 859–875 (2011)
15. Saleem, Y., Rehmani, M.H.: Primary radio user activity models for cognitive radio networks: a survey. J. Netw. Comput. Appl. **43**, 1–16 (2014)
16. Chen, Y., Oh, H.S.: A survey of measurement-based spectrum occupancy modeling for cognitive radios. IEEE Commun. Surv. Tutor. **18**(1), 848–859 (2016)

An Energy-Efficient Framework Based on Random Waypoint Mobility Model in WSN-Assisted IoT

Anurag Shukla and Sarsij Tripathi

Abstract In this paper, we propose a hierarchical framework for mobile ad hoc IoT applications such as animal monitoring system, which supports the scalable network and uniform energy consumptions across all the nodes to prolong the network lifetime. The proposed framework uses the clustering algorithm to divide the network into equal zones. Each node belongs to some particular zone. Nodes are mobile (based on Random Waypoint Model) and able to move from one zone to another zone in the entire network. Some nodes select as zone head, zone coordinator, and relay node. Node forwards the data packets to the base station via a nearest relay node. Relay node forwards the packets to the zone head, and zone head transmits these packets to base stations via upper layer zone coordinator. The comparison has been made of the proposed framework with static model, LEACH and EESAA routing protocols. It identified that the proposed technique performs well in terms of energy consumption and network lifetime.

Keywords Internet of Things · Random Waypoint Model · Network lifetime

1 Introduction

Wireless Sensor Network (WSN) has an essential role in many IoT applications [1], and it is considered as a basic communication model in WSN-assisted Internet of Things (IoT) applications for transmitting data from the deployed network to Base Station (BS) [2]. WSN-assisted IoT is a very popular area in IoT that includes various applications such as smart parking system, smart healthcare, border security, and smart transport system. The main idea of the IoT applications is that it makes the physical objects conscious about real-world attributes such as hearing, feeling, monitoring, and performing actions with the coordination of other network devices [3]. Sensor nodes in such applications can be used for different roles in different networks. The sensor nodes are resource constrained in terms of energy, processing

A. Shukla (✉) · S. Tripathi
National Institute of Technology, Raipur, India
e-mail: anurag672@yahoo.com

power, and memory. The energy source in sensor nodes is non-rechargeable and a majority portion of sensor energy is depleted in communication. Therefore, to increase the network lifetime, energy conservation is a challenging task in IoT network. The number of participate nodes in IoT network is more as compared to WSN network. So, current WSN techniques cannot be applied directly on the scalable IoT network.

The network lifetime and scalability issue can be solved by hierarchical framework [4, 5], therefore, in this work, we followed the hierarchical clustering technique for node deployment. The IoT domain covers many applications, which make the daily human life more comfortable. The participate nodes in the IoT network either remain static or mobile, which depends on the requirement of the application. For example, environment monitoring and coal mine goof application support static network whereas vehicular ad hoc network and animal monitoring system support mobile network. In this paper, we propose hierarchical framework for node deployment and efficient routing protocols for ad hoc IoT applications.

To reduce energy depletion for overhead communication, efficient node selection for various roles (zone head, zone coordinator, and relay node) and their integration in minimum energy routing protocol are the prime goals of this paper. The proposed scheme is compared with traditional WSN techniques and observed that the proposed scheme is efficient in terms of network lifetime and energy consumptions. The research work carried out can be summarized as follows:

(1) In this work, we deploy sensor nodes in the hierarchical framework. This topology supports the scalability feature and can be extended up to any level. The communication among the sensor nodes in the zone and outside the zone follows some well-defined rules [6] to distribute the zone load so that network lifetime can be increased.
(2) To increase the network lifetime, we introduce an efficient node selection method for hierarchical framework to minimize the communication cost.
(3) To introduce mobility, we follow the Random Waypoint Mobility Model. The participate nodes can move in the entire terrain area.
(4) The simulation has been performed in MATLAB with standard parameters for randomly deployed nodes, and numerical results validate that the proposed technique is preferable than traditional WSN techniques for IoT application.

The structure of the paper is as follows: an overview of WSN routing protocols is given in Sect. 2. The classification of mobility model is given in Sect. 3. Section 4 covers the description of the system framework. Energy-efficient algorithm for the proposed framework is presented in Sect. 5. In Sect. 6, we will discuss the simulation results of the proposed framework and their comparative analysis with other standard routing protocols. Later, the conclusion is made.

2   Related Work

In WSN-assisted IoT network after nodes deployment, the primary objective of the sensor node is to send the data to the BS. The communication between sensor and BS can take place in two ways: direct transmission and multi-hop transmission. Direct transmission leads to fast energy depletion of sensor energy, and it causes the short network lifetime. To overcome these issues, many of the recent routing protocols use multi-hop communication for data transmission via some intermediate nodes like cluster head, relay node, and so on. Low Energy Adaptive Clustering Hierarchy (LEACH) is a multi-hop routing protocol [7], where the network is divided into multiple clusters and from every cluster; some of the nodes are selected as cluster head via probability function and residual energy level. The cluster head receives the sensor data from their local member node, and aggregate data send to the BS [8]. A multi-hop form of LEACH is M-LEACH [9], where CH sends the data to the BS via other intermediate CHs instead the direct transmission as LEACH. C-LEACH (Centralized-LEACH) [10] is one more version of LEACH, where the base station has the responsibility for the cluster formation. TL-LEACH (Two-Level LEACH) [11] is an enhanced version of M-LEACH, where primary and secondary CHs are formed for creating two hierarchical level structure. Sensor nodes send the data packets to primary CHs, and primary CH sends these data packets to secondary CH, which is above in the hierarchy level. The TL-LEACH increases the network lifetime. EESAA [12] adds residual energy as one more parameter for interchange between sleep and active modes to conserve energy. The Hybrid Energy Efficient Distributed (HEED) protocol [13] used a new method for CH selection, which is based on the hybridization of communication cost and residual energy.

Priyan et al. [14] proposed a framework for the Internet of Vehicle (IoV) based on Random Waypoint Mobility Model. This healthcare monitoring system includes mobile ambulance, mobile doctor, and patience. The selection of best ambulance (mobile node) among all the available ambulance is based on the Performance Rank (PR). PR is calculated for every ambulance by considering these parameters: Euclidean distance from the patient, medical capacity, and number of patients currently using this ambulance. The ambulance with minimum PR select for a needed patient.

The prime focus of above-discussed protocols is to achieve energy-efficient communication to increase the network lifetime for the static network. These routing protocols are not suited for IoT applications due to complex nature. They require more phases and time for cluster formation and less support for scalability. In our previous paper, we introduce an optimal RN selection method to increase the network lifetime for the static network [15]. In this paper, we give a framework for Ad hoc mobile IoT application by incorporating Random Waypoint Model with efficient relay node RN selection method.

3 Mobility Models Classification

Classifications of mobility models are based on the restrictions and requirements. Commonly used mobility models are listed below.

3.1 Random Based Mobility Model

This model is suitable for those applications; there are no restrictions or any dependencies. Random Waypoint Model comes under this category.

3.2 Temporal Dependencies

In temporal dependencies, the movement of nodes is based on the movement of the past nodes.

3.3 Geographic Restrictions

The node movement is restricted in the geographical restrictions mobility model. The node is allowed to move at their given domain only.

3.4 Hybrid Characteristics

As the name suggests this model is a hybrid of temporal, random, and geographical restrictions model.

4 System Model

4.1 Propose Framework

Figure 1 represents the multitier hierarchical framework, where the nodes are able to move from one zone to another zone across the entire network area. This framework is initially used in [16, 17] for static IoT network. We developed this framework for Ad hoc mobile network by applying Random Waypoint Model. The proposed framework can be used for those IoT applications, where most of the participate nodes in the
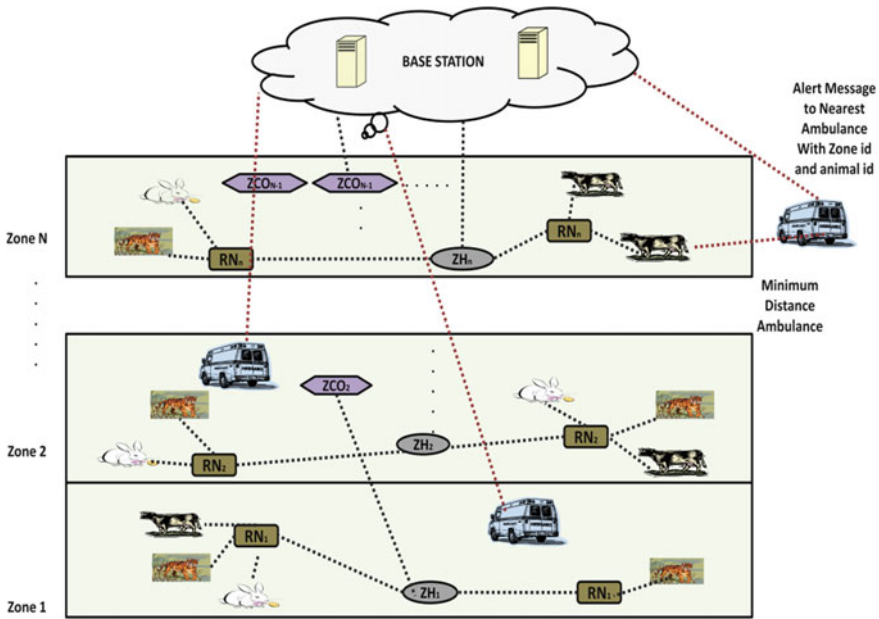
Fig. 1 A multitier hierarchical framework

network are moveable, for example, animal tracking and monitoring system [18], and Internet of Vehicles [14]. In Fig. 1, the top layer is configured as a base station layer and whole network area is divided into zones. The animals are attached with IoT wearable devices to collect their clinical data continuously. The clinical data is transmitted to the base station via the Relay Node (RN), Zone Head (ZH), and Zone Coordinator (ZCO). The clinical data is sent to nearest RN, where the RN forwards data to ZH and ZH sends the data to the BS via upper layer ZCO. Every zone has one ZCO except lower zone, which corresponds to lower zones. When the animal health is critical, the alert message is forwarded to the BS and BS broadcast this message to all the mobile ambulance with animal id and their current zone id so that the ambulance can reach with the minimum time. The wearable animal's IoT devices cannot be recharged or changed on a daily basis, so an efficient communication method is required to conserve energy in order to increase the network lifetime.

4.2 Assumptions

(1) All participating nodes in the network are moveable [14].
(2) The same set of nodes (SN, RN, ZH, and ZCO) have same attribute, e.g., transmission power, initial energy, energy depletion parameters, and so on.
(3) Nodes know their spot by some localization technique [19].

(4) Sensor nodes are running on the limited energy source, which cannot be recharged or replaceable.

(5) Node can send their data to the base station in a multi-hop way and base station is not restricted with respect to energy.

(6) Communication medium is error-free and reliable.

(7) The whole IoT network is connected, i.e., each participating node in a network has a path to a BS.

4.3 Energy Model

The main source of energy depletion in IoT network is data transmission and reception. Sensor nodes deplete the energy in sensing and processing is very less. In this work, we follow the first order radio model [8] for energy depletion in data communication, which is given below.

$$E_{tx} = L \times \left(E_{elec} + \in_X \times d^2 \right) \tag{1}$$

$$E_{rx} = L \times E_{elec} \tag{2}$$

$$E_{tx} = L \times \left(E_{elec} + \in_X \times d^4 \right) \tag{3}$$

ϵ_1, ϵ_2, ϵ_3, and ϵ_4 is used as a node's amplifier for SN, RN, CH, and CCO, respectively. The energy depletion in short distance and long distance communication is identified by Eqs. (1) and (3), respectively. Data reception at RNs, CHs, and CCOs nodes and their utilized energy in this task is computed by Eq. (2). Energy depletion per unit time for each node is computed by

$$E_p = \sum_{q \in RN} A_{pq}.F_{pq}.\left(E_{elec}^{NN} + \in_1 .d_{pq}^2 \right) \forall p \in SN \tag{4}$$

$$E_q = \sum_{p \in SN} A_{pq}.F_{pq}.E_{elec}^{RN} + \sum_{r \in ZH} A_{qr}.F_{qr}.\left(E_{elec}^{RN} + \in_2 .d_{qr}^2 \right) \forall q \in RN \tag{5}$$

$$E_r = \sum_{q \in RN} A_{qr}.F_{qr}.E_{elec}^{ZH} + \sum_{s \in ZCO} A_{rs}.F_{rs}.\left(E_{elec}^{ZH} + \in_3 .d_{rs}^2 \right) \forall r \in ZH \tag{6}$$

$$E_s = \sum_{r \in ZH \cup CCO} A_{rs}.F_{rs}.E_{elec}^{ZCO} + \sum_{t \in CCO \cup BS} A_{st}.F_{st}.\left(E_{elec}^{ZCO} + \in_4 .d_{st}^2 \right) s \in ZCO \tag{7}$$

$$E_t = \sum_{s \in ZCO} A_{st}.F_{st}.E_{elec}^{BS} \forall t \in BS \tag{8}$$

E_P, E_q, E_r, E_s, and E_t denote the energy depleted by SNs, RNs, ZHs, ZCOs, and BS, respectively in data transmission and reception. The symbols E_{elec}^{SN}, E_{elec}^{RN}, E_{elec}^{ZH}, E_{elec}^{ZCO}, and E_{elec}^{BS} indicate the energy consumption in the radio electronics of a SNs, RNs, ZHs, ZCOs, and BS. Equation (4) denotes the energy consumed in data transmission processes between SN and RN within the intra-zone. The energy consumed in sending the data to the ZH by RN and data receiving by the ZH is computed by Eq. (5). The energy consumed by ZH for data reception from RN and transmission to the upper layer ZCO is represented by Eq. (6). Equation (7) shows the energy consumed by ZCO for the data communication process to upper layer ZCO or the BS and for the data receiving processes either by lower layer ZH or ZCO. Equation (8) shows the energy consumed in the BS layer. All of the above equations exclude the energy consumption by signaling data because it is negligible [6] as compared to data transmission and receiving.

4.4 Network Life Time

In this paper, we consider three metrics for network lifetime comparison, which includes First Node Dead Statistics (FND_Statistics), Last Node Dead Statistics (LND_Statistics), and the number of dead nodes after each round. The number of packets received by BS through hierarchical topology (SN → RN → ZH → ZCO → BS) is denoted as one round [17]. FND_Statistics and LND_Statistics is the duration between the rounds, when the communication begins and the rounds were the first and last node dead [20, 21].

5 Efficient Communication Algorithm

The proposed framework can be used for animal monitoring in forest or zoo. In this paper, we make the analysis on the bases that animals do not leave the forest area and walk around the forest boundary. Random Waypoint Mobility Model is used to simulate the proposed model. Predefined ranges are used to mobile the nodes with random directions and speeds.

Algorithm 1. Energy Efficient Communication (EEC) Algorithm

1: **Procedure** Node Selection and Routing Phase (Number of Nodes N, Zone Head ZH,
    Zone Coordinator ZCO, Number of Zone M)
2: Network deployments in the static area   ▷ N sensor nodes are randomly distributed
3: **Divide** the network into **M zones** by subarea division algorithm with fixed boundaries
4: **declare** zone_id   ▷ Id for every zone 1 to M
5: Distribute the ids to every zones, where *clstr_id ≤ M.*
6: **declare** SN   ▷ All SN ∈ N
7: **declare** local_zone   ▷ Current zone.
8: **declare** ZH = (empty set)   ▷ To store selected zone head
9: **declare** ZCO = (empty set)   ▷ To store selected zone coordinator
10: **declare** T = travelled time of mobile node
11: **declare** θ1 = travelling direction of mobile node
12: **declare** V = travelling velocity of mobile node
                           % **Mobility Phase** %
13:    $X_n = X_{n-1} + V_n \times T \times \cos\theta_n$ ▷ All $X_n \in SN_X$, RN_X, ZH_X, ZCO_X
14:    $Y_n = Y_{n-1} + V_n \times T \times \sin\theta_n$ ▷ All $Y_n \in SN_Y$, RN_Y, ZH_Y, ZCO
                       % **RN Selection Phase** %
15:   **for** every zone, zone_id ≤ M **do**
16:     │ **RN selection** method
17:   **end for**
                  % **ZH and ZCO Selection Phase** %
18:   **for** every zone, zone_id ≤ M **do**
19:     │ **Select zh** and **zco** randomly in local zone, where **zh.energy ≥ energyThre
        │ shold** and **zco.energy ≥ energyThreshold**
20:     │ Number of **zco** in every zone = zone_id-1
21:     │ zh ∈ ZH and zco ∈ ZCO
22:   **end for**
                    % **Data Routing Phase** %
23: **for** every zone, zone_id ≤ M **do**
24:     │ **Send** the packet from u *(u ∈ SN ∈ local_zone)* to *v (v ∈ RN ∈ local_zone)*
25:     │ **Send** packets in order from v → w → x → y (w ∈ ZH, x ∈ ZCO, y ∈ BS)
26: **end for**
27: **for** every zone, zone_id ≤ M **do**
28:     │ **if Rn.energy** ≤ energyThreshold **then**
29:     │   **go to** RN Selection Phase (Line 16-20)
30:     │ **end if**
31:     │ **if** zh.energy and zco.energy ≤ energyThreshold **then**
32:     │   **go to** ZH and ZCO Selection Phase (Line 20-25)
33:     │ **else**
34:     │ **go to** Data Routing Phase
35:     │ **end if**
36: **end for**
37: **end Procedure**

    As we discuss in the upper section, energy-efficient communication method is required in the animal monitoring system to increase the network lifetime. The solution to the energy consumption problem is achieved by Algorithm 1, where Algorithm 1 elects the RN efficiently and routes the data from SN to BS. The proposed technique work in the following six steps. The first step (line 2) nodes are deployed randomly within the networks. Step 2 (line 3) divides the network into equal size by subarea division technique. Suppose if the area is 300 m^2 then the zone size is 300 m × 30 m

and number of zones is 10. In Step 3 (line 13) Random Waypoint Model is used for node mobility. In Step 4 (Line 15–17), RNs selection is done for all zones. In Step 5 (Line 18–22) ZH and ZCOs select randomly. In Step 6 (Line 23–26), the data is transmitted from SN to BS via RNs, ZHs, and ZCOs. To maintain robustness in large IoT network, energy of RN, ZH, and ZCO are compared with the threshold so that routing protocol can work even if some nodes become dead. The role of ZH, ZCOs, and RNs can be exchanged with high energy node.

6 Result Discussion

The simulation was performed in MATLAB with the Table 1 simulation parameter. Figure 2 shows the network lifetime comparison for a static and mobile framework with standard routing protocols. It can be noticed from the figure that the mobile node performs well than compared to static, LEACH, and EESAA routing protocol. To make it more understandable, we make a comparison with FND_Statistics and LND_Statistics. From Fig. 4 it can be observed that all the nodes are died in EESAA after 4164 rounds, in static after 4495 rounds, and in dynamic after 5049 rounds. The last node died in LEACH after 310 rounds, which is minimum among the compared protocols. It is because LEACH performs well when the number of nodes are less in the network. As soon as number of nodes increases in network, LEACH network lifetime goes down [17].

Figure 3 represents the total energy depletion of the compared protocols under from first round to last round. From the figure, we can see that proposed framework

Table 1 Simulation parameters

Parameters	Values
Number of sensor nodes (N)	1000
Network area (M × M)	200×200 m^2
E_0 (Initial energy of deployed node)	0.5 J
E^{elec} for all the nodes	50 nJ/bit
E_l represents the energy depletion for long distance communication	0.0013 pJ/bit/m^4
E_s represents the energy depletion for short distance communication	10 pJ/bit/m^2
Energy used in beamforming E_{bf}	5 nJ/bit
L (number of bits in a packet)	4000 bits
RN communication range	40 m
E_{da} (data aggregation energy per bit)	5 nJ
Number of zones	10
Speed interval	[0.2–2.2] m/s
Walk interval	[0–1] m/s
Pause interval	[2–6] s

Fig. 2 Network lifetime comparison for the topology: Area = 200 m², Nodes = 1000
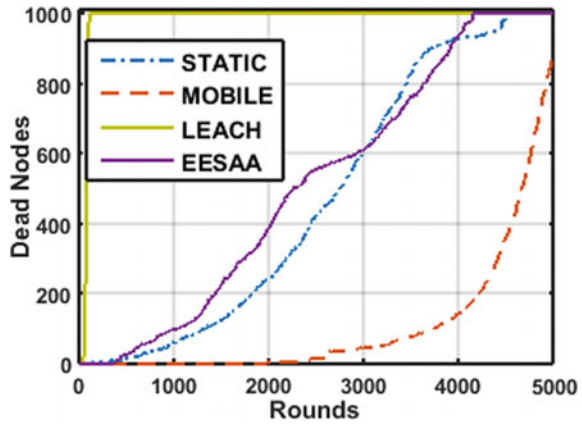
Fig. 3 Total energy depletion for the topology: Area = 200 m², Nodes = 1000
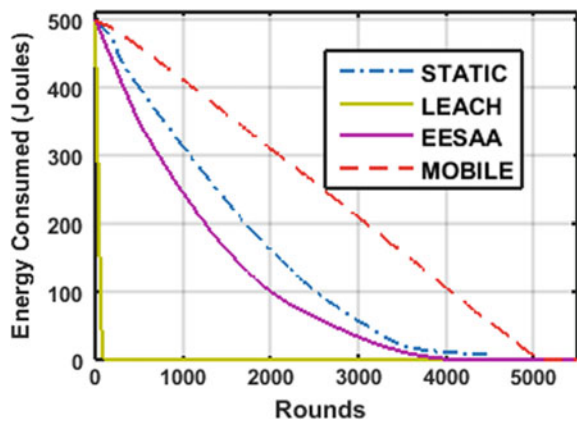
has less and static energy depletion curve as compared to other routing protocols. It is because all the nodes are mobile in the proposed framework and in every round fresh node is selected for high reasonability job as RN and ZH. So, the energy depletion is uniformed for all nodes in the proposed framework instead of some individual nodes, which makes the network's lifetime of proposed framework better (Fig. 4).

7 Conclusion

Energy depletion and network lifetime are always an issue for sensor nodes. The sensor nodes are an important part of IoT applications and most of the participating devices in IoT applications are sensors. In this paper, we developed a framework for Mobile Ad Hoc IoT applications in such a way that network lifetime can increase. The proposed framework is based on Random Waypoint Model and uses the hierarchical
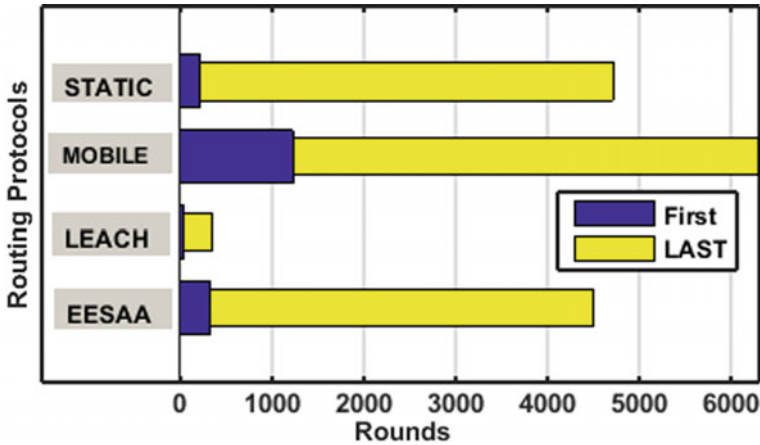
Fig. 4 FND_Statistics & LND_ Statistics for topology: Area $= 200$ m^2, Nodes $= 1000$

topology for network deployment to support the core IoT feature scalability. The proposed framework is compared with LEACH, EESAA, and static topology and simulation results validate the effectiveness of the proposed framework in order to increase the network lifetime.

References

1. Ejaz, W., Naeem, M., Basharat, M., Anpalagan, A., Kandeepan, S.: Efficient wireless power transfer in software-defined wireless sensor networks. IEEE Sens. J. **16**(20), 7409–7420 (2016)
2. Aijaz, A., Aghvami, A.H.: Cognitive machine-to-machine communications for Internet-of-Things: a protocol stack perspective. IEEE Internet Things J. **2**(2), 103–112 (2015)
3. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Internet of Things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. **17**(4), 2347–2376 (2015)
4. Younis, M., Youssef, M., Arisha, K.: Energy-aware management for cluster-based sensor networks. Comput. Netw. **43**(5), 649–668 (2003)
5. Liu, J.S., Lin, C.H.R.: Energy-efficiency clustering protocol in wireless sensor networks. Ad Hoc Netw. **3**(3), 371–388 (2005)
6. Shen, J., Wang, A., Wang, C., Hung, P.C., Lai, C.F.: An efficient centroid-based routing protocol for energy management in WSN-assisted IoT. IEEE Access **5**, 18469–18479 (2017)
7. Xia, H., Zhang, R.H., Yu, J., Pan, Z.K.: Energy-efficient routing algorithm based on unequal clustering and connected graph in wireless sensor networks. Int. J. Wirel. Inf. Netw. **23**(2), 141–150 (2016)
8. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: 2000. Proceedings of the 33rd Annual Hawaii International Conference on System sciences, 2000, 10 pp. IEEE (2000)
9. Mhatre, V., Rosenberg, C.: Homogeneous vs heterogeneous clustered sensor networks: a comparative study. In: ICC, pp. 3646–3651 (2004)

10. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**(4), 660–670 (2002)
11. Loscri, V., Morabito, G., Marano, S.: A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH). In: Vehicular Technology Conference, 2005. VTC-2005-Fall. 2005 IEEE 62nd, vol. 3, pp. 1809–1813. IEEE (2005)
12. Shah, T., Javaid, N., Qureshi, T.N.: Energy efficient sleep awake aware (EESAA) intelligent sensor network routing protocol. In: Multitopic Conference (INMIC), 2012 15th International, pp. 317–322. IEEE (2012)
13. Younis, O., Fahmy, S.: HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mob. Comput. **3**(4), 366–379 (2004)
14. Priyan, M.K., Devi, G.U.: Energy efficient node selection algorithm based on node performance index and random waypoint mobility model in Internet of Vehicles. Clust. Comput. 1–15 (2017)
15. Shukla, A., Tripathi, S.: An optimal relay node selection technique to support green Internet of Things. J. Intell. Fuzzy Syst. (Preprint), 1–14
16. Huang, J., Meng, Y., Gong, X., Liu, Y., Duan, Q.: A novel deployment scheme for green Internet of Things. IEEE Internet Things J. **1**(2), 196–205 (2014)
17. Rani, S., Talwar, R., Malhotra, J., Ahmed, S.H., Sarkar, M., Song, H.: A novel scheme for an energy efficient Internet of Things based on wireless sensor networks. Sensors **15**(11), 28603–28626 (2015)
18. Xu, J., Solmaz, G., Rahmatizadeh, R., Turgut, D., Boloni, L.: Internet of Things applications: animal monitoring with unmanned aerial vehicle. arXiv preprint arXiv:1610.05287 (2016)
19. Niculescu, D., Nath, B.: Ad hoc positioning system (APS) using AOA. In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 3, pp. 1734–1743. IEEE (2003)
20. Han, Z., Wu, J., Zhang, J., Liu, L., Tian, K.: A general self-organized tree-based energy-balance routing protocol for wireless sensor network. IEEE Trans. Nucl. Sci. **61**(2), 732–740 (2014)
21. Chang, J.H., Tassiulas, L.: Energy conserving routing in wireless ad-hoc networks. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 22–31. IEEE (2000)

A Scheduling Algorithm Including Deadline of Messages in Vehicular Ad hoc Network

Abhishek Agarwal, Raghavendra Pal and Arun Prakash

Abstract Vehicular Ad hoc Network is the area which is attracting a large number of researchers. The problem faced by this network is high mobility, large topological area, and frequent changes in connection and location. Messages are broadly divided into safety and non-safety messages. For proper communication and safety of life, proper scheduling of messages is required in VANET so that important messages are given the highest priority. In this paper, a scheduling strategy is proposed which schedules the messages according to the parameters like deadline, size, and static factor of message as well as the speed of the vehicle. Based on these parameters the priority of messages is determined and the messages are then rescheduled accordingly. The proposed algorithm is then simulated to demonstrate improvement in comparison to Taherkhani et al.

Keywords Vehicular Ad hoc network · Vehicle-to-vehicle · Vehicle-to-infrastructure · Scheduling · Prioritization

1 Introduction

Vehicular Ad hoc Network (VANET) is a wireless technology that allows communication of vehicles with other vehicles (V2V) and also with the infrastructure along roadsides (V2I). Being first mentioned and introduced in 2001 they have become one of the most potential fields for research. With the burgeoning population in major

A. Agarwal (✉)
Department of Electronics and Communication Engineering, Bundelkhand Institute of Engineering and Technology, Jhansi 284128, India
e-mail: abhishekagarwal482@gmail.com

R. Pal · A. Prakash
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India
e-mail: raghavendra.pal3@gmail.com

A. Prakash
e-mail: arun@mnnit.ac.in

metropolitan cities across the world and increasing congestion on the roads the use of VANET is becoming necessary. Based on Mobile Ad hoc Network (MANETs) principles in VANET the vehicles act as mobile nodes which are constantly changing their locations and communicating with the vehicles in their range. The links with the vehicles and RSUs (Road Side Units, which is basically roadside infrastructure) is constantly being made and break due to the fast-changing location of vehicles. In VANET 802.11p [1] and WAVE [2] are the standards used for efficient communication between the vehicles.

The protocols and standards for V2V communication and V2I communication is being defined by Dedicated Short Range Communication [2] Scheme. The VANET is characterized by fast-changing location, large topological areas, and frequent connections and disconnections. All these pose a challenge in VANET networks. Further increasing congestion and large amount of requests leads to choking of channels and degradation in the performance of VANET networks. Consequently, various parameters like packet delivery ratio, throughput decrease while the delay in transmission of messages increases. There have been various strategies and algorithms proposed for reducing the congestion and efficient transfer of messages by prioritizing the messages. These scheduling schemes help in efficient transmission of information and improving the various parameters of VANET.

DSRC uses a 75 MHz bandwidth at 5.9 GHz for transferring the messages. It is composed of seven channels out of which six channels are Service Channels (SCH) and one channel is Control Channel (CCH) [2]. Between each channel a guard band is provided. The high priority safety messages are transmitted by the control channel while the low priority non-safety messages are processed by the service channels. The duration of each interval is 46 ms while that of guard interval is 4 ms. The guard interval is the transition zone in which no communication takes place and at end of its Communication of the next Channel Interval (CCHI or SCHI) in sequence proceeds.

This paper proposes a scheduling algorithm which helps in increasing packet delivery ratio, throughput, and reducing jitter as compared to the earlier proposed algorithms. The packets are scheduled on the basis of their speed, size, deadline, and static factor. The static factor comprising whether a message is safety or non-safety message. The paper is organized as follows: In Sect. 2, the existing RSU scheduling schemes are described. In Sect. 3, a basic system model of the proposed scheme is explained. In Sect. 4 a new collaborative scheduling scheme to solve the problems in the existing RSU scheduling schemes is proposed. In Sect. 5, a performance evaluation of the proposed scheme is presented to prove the effectiveness of the scheme by comparing it with the existing scheme. Finally, Sect. 5 concludes the paper.

2 Related Works

First In First Out (FIFO) scheduling algorithm is one in which the message request is served according to its arrival, i.e., the first message request is served first then the second message request is served and so on. In First Deadline First (FDF) algorithm the messages are served according to their deadline. In Smallest Data Size First (SDF) [3] the message whose size is the least is being served first.

Bai et al. [4] proposed Context Awareness Beacon Scheduling (CABS). In this, the congestion within high-density vehicular networks was controlled by scheduling the beacon messages which were being transmitted at a very high broadcasting rate. Taherkhani and Pierre [5], proposed Uni-Objective Tabu search (UOTabu) congestion-controlled strategy. In this, the channel usage was monitored to detect the congestion and then Tabu Search algorithm was applied. In this algorithm, it was showed that it could minimize the delay and packet loss compared to other strategies.

Zhang et al. [6] proposed D*S RSU scheduling scheme in which the packets were scheduled according to their deadline and size. The packet with the smallest DS_ value was scheduled first. Gui and Chan [7] proposed a motion prediction based scheduling scheme in which an RSU can transfer the vehicular request to another RSU in the direction of motion of vehicle if itself is facing a heavy load. In ROADCAST (Popularity Aware Content Sharing) [8] the most relevant information is being transferred among vehicles. It consisted of two components popularity aware content retrieval and popularity aware data replacement that took care of transmission of most popular data and at the same time managing delay of less popular data.

In [9] to prioritize and schedule safety messages two congestion control strategies are proposed. There are two units, namely, priority assignment unit and message scheduling unit. The priority assignment unit assigned priority to the message on the basis of static and dynamic factors while the message scheduling unit reschedules these messages in service and control channel queues based on their priorities determined by priority assignment unit. These strategies are then compared with earlier proposed strategies on the basis of various parameters like number of packets lost, packet loss ratio, average delay, and average throughput.

3 Problem Statement and System Model

Since there are multiple priorities of messages in VANET depending upon various factors. Hence, there is a need for scheduling packets in the queue of a vehicle in such a way so that maximum QoS is achieved. One of the main parameters in scheduling is deadline that is the time after which a packet loses its significance. Hence, an algorithm is needed that uses deadline of packet while deciding its priority. In the proposed system model, each vehicle has an On-Board unit (OBU), contains a transceiver for transmission and reception of message. It works on IEEE802.11p/WAVE 1609 standard. Each OBU has a GPS that provides position, velocity, and other GPS data. The
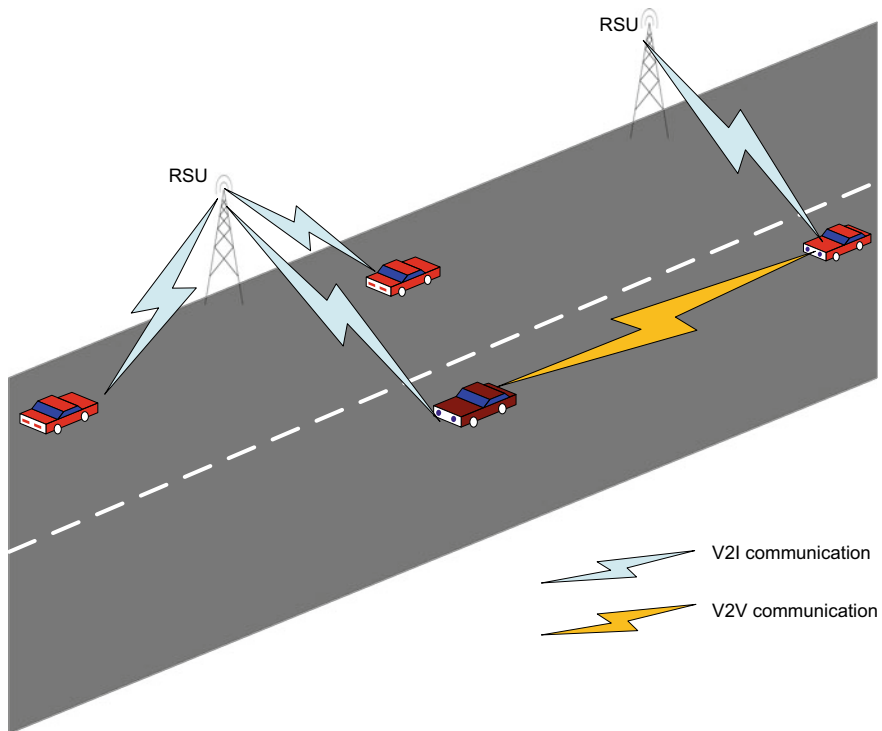
Fig. 1 System model for VANET

RSUs also contain transceivers for transmission and reception of messages. Figure 1 shows the system model for VANET with V2I and V2V communication.

4 Proposed Scheme

In the proposed scheme an algorithm is presented to reschedule the messages in the RSUs. There is one queue in RSU that receives the service requests from all the vehicles in its communication range and then reschedules the messages to be serviced according to the proposed algorithm. The results of the proposed algorithm are then evaluated and compared with a most relevant algorithm [9] on the basis of parameters such as packet delivery ratio, delay, throughput, and jitter.

The message scheduling is performed according to the proposed algorithm. In the proposed algorithm when the RSU receives messages from the vehicles it enques them in its queue and then sorts the message according to the algorithm as given in Fig. 2. This algorithm takes mainly three parameters into account, speed of the vehicle, size, and deadline of the message. Deadline of the message is calculated as
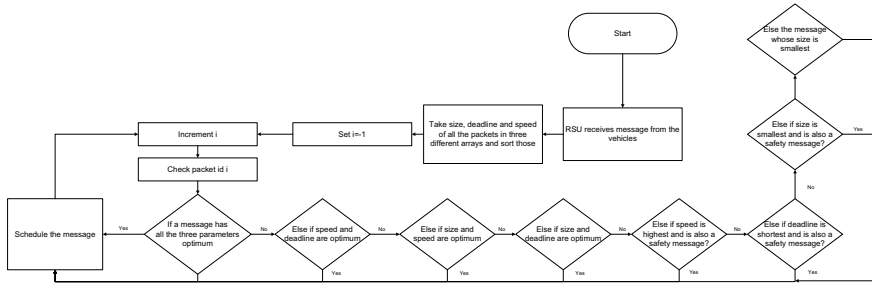
Fig. 2 Flowchart of the proposed algorithm

follows:

$$deadline = \text{current time} + 10 \, \text{ms}$$

The deadline of each message is assigned to it as in the above equation at the time of its generation. As the simulation time increases the deadline of message decreases and hence its priority increases according to Fig. 2.

The static factor helps in determining whether the message is safety or non-safety message as in the equation given below

$$static\,factor = \begin{pmatrix} 0; for\,safety\,messages \\ 1; for\,non\text{-}safety\,messages \end{pmatrix}$$

There are three temporary variables taken which stores in each iteration of for loop the smallest size message, the fastest speed vehicle, and the smallest deadline message along with the position of three parameters in the respective arrays. It then advances to compare the position of three parameters. The further mechanism is explained through the help of algorithm illustrated in Table 1.

5 Results and Discussions

Results are obtained by simulation on Network Simulator (NS) version 2.34. Simulation parameters are presented in Table 2. Following parameters have been used to test the proposed algorithm and they are defined as follows:

Packet Delivery Ratio (PDR): The ratio of no. of packets received by no. of packets sent. It should be as maximum as possible.

Average delay: The average time taken for information to reach the receiver from the sender. It should be as minimum as possible.

Average throughput: Information successfully transmitted per unit time. It should be as maximum as possible.

Table 1 Algorithm for scheduling mechanism

Step 1	Set i = −1
Step 2	Increment i
Step 3	Check packet id i
Step 4	If ith packet has all the three parameters optimum then
	Schedule the packet and Goto Step 2
Step 5	Else if deadline and speed of ith packet are optimum then
	Schedule the packet and Goto Step 2
Step 6	Else if size and speed of ith packet are optimum then
	Schedule the packet and Goto Step 2
Step 7	Else if size and deadline of ith packet are optimum then
	Schedule the packet and Goto Step 2
Step 8	Else if speed is optimum and is also a safety message then
	Schedule the packet and Goto Step 2
Step 9	Else if deadline is optimum and is also a safety message then
	Schedule the packet and Goto Step 2
Step 10	Else if size is optimum and is also a safety message then
	Schedule the packet and Goto Step 2
Step 11	Else the packet whose size is optimum Schedule it and Goto Step 2

Table 2 Simulation parameters

Parameters	Value
Simulation area	2000 m × 2000 m
Number of vehicles	20–100
Vehicle speed	10–20 m/sec
Transmission rate	3 Mbps
Message size	511–611 bytes
MAC interface	Mac/802_11Ext
Simulation time	20 s
Transmission range	300 m, 1000 m

Average jitter: Variation in delay. It should be as minimum as possible.

Figure 3 shows comparison of simulation results of packet delivery ratio by varying the number of vehicles. The packet delivery ratio increases with the number of vehicles. The simulation results show better performance of the proposed algorithm in comparison to the Taherkhani et al. This is because in the proposed algorithm highest priority is being given to the deadline of message sent by vehicle and speed of the vehicle which in turn increases packet delivery ratio. Since the message is being broadcasted, therefore, for each packet sent there are multiple number of receivers, hence PDR is greater than 100%.

Fig. 3 Packet delivery ratio versus No. of vehicles
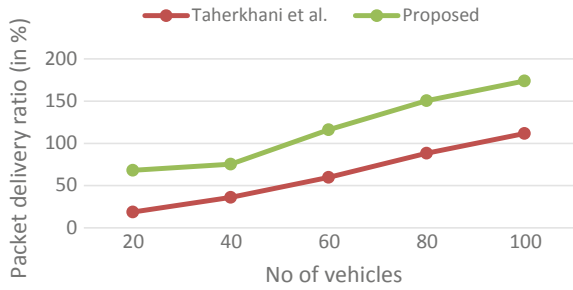
Figure 4 shows comparison of simulation results of delay with increasing number of vehicles. The proposed algorithm shows an initial decrease in delay but then corresponding increase in delay as compared to Taherkhani et al. This is because in the proposed algorithm more emphasis has been laid on the deadline of message within which it should be transmitted rather than size of message to ensure that information lost is minimum and maximum packets are transmitted. Delay is compromised in order to maximize the no. of packets transmitted and hence the information.

Figure 5 shows simulation results comparison of throughput with increasing number of vehicles. The throughput increases with increasing no. of vehicles. The proposed algorithm shows better performance in comparison to Taherkhani et al. The reason is that in the proposed algorithm more emphasis has been laid upon the

Fig. 4 Delay versus No. of vehicles
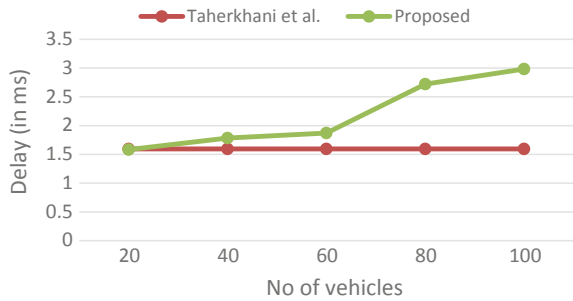
Fig. 5 Throughput versus No. of vehicles
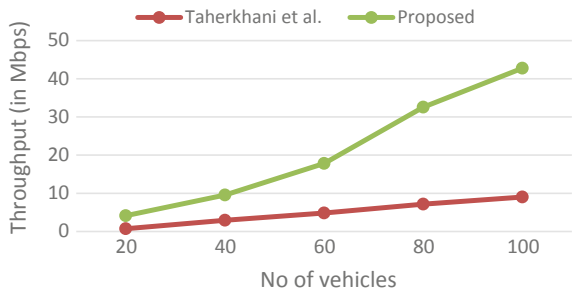
Fig. 6 Jitter versus No. of
vehicles
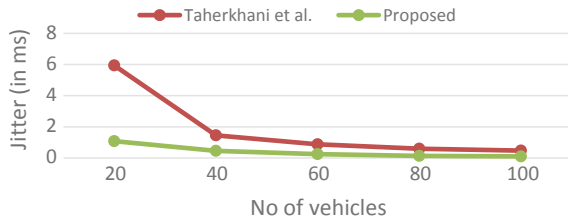
deadline of message and speed of vehicle which ensures more no. of packets being
transmitted and reduced number of packets dropped which in turn increases the
throughput.

Figure 6 shows comparison of simulation results of jitter with increasing no.
of vehicles. The jitter decreases with increasing number of vehicles. The proposed
algorithm shows better results in comparison to Taherkhani et al.

6 Conclusion

In this paper, a scheduling algorithm is proposed to schedule the messages according
to the deadline, size, and static factor of message and speed of the vehicle which
is sending the message. The messages are first dequeued from service and control
channels and then after rescheduling them according to the proposed algorithm,
enqueued back to service and control channels. Simulation is carried out to compare
the results of the proposed algorithm with Taherkhani et al. Better scheduling results
are obtained owing to prioritizing the messages with short deadline and faster speed
of vehicles first. The parameters like packet delivery ratio, throughput are increased
while jitter is reduced showing a better way of scheduling of message requests. Hence
more reliable and efficient VANET system is implemented.

References

1. Ayyapan, B., Kumar, P.M.: Vehicular Ad hoc Networks (VANETS): Architecture, Methodologies
   and Design Issues. https://doi.org/10.1109/iconstem.2016.7560946
2. Kenney, J.B.: Dedicated short-range communications (DSRC) standards in the United States. In:
   Proceedings of the IEEE, vol. 99, pp. 1162–1182. IEEE (2011). https://doi.org/10.1109/jproc.
   2011.2132790
3. Kumar, V., Chand, N.: Data scheduling in VANETs: a review. Int. J. Comput. Sci. Commun. **1**,
   399–403 (2010)
4. Bai, S., Oh, J., Jung, J.: Context awareness beacon scheduling scheme for congestion control in
   vehicle to vehicle safety communication. Ad Hoc Netw. **11**, 2049–2058 (2013). https://doi.org/
   10.1016/j.adhoc.2012.02.014
5. Taherkhani, N., Pierre, S.: Congestion control in vehicular ad hoc networks using meta-heuristic
   techniques. In: Proceedings of the Second ACM International Symposium on Design and Analy-

sis of Intelligent Vehicular Networks and Applications, pp. 47–54. ACM Digital Library, Paphos, Cyprus (2012). https://doi.org/10.1145/2386958.2386966

6. Zhang, Y., Zhao, J., Cao, G.: On scheduling vehicle-roadside data access. In: International Workshop on Vehicular Ad hoc Networks, pp. 9–18 (2007). https://doi.org/10.1145/1287748.1287751

7. Gui, Y., Chan, E.: A motion prediction based cooperative scheduling scheme for vehicle-roadside data access. In: International Conference on Networking and Information Technology, pp. 195–204 (2011)

8. Zhang, Y., Zhao, J., Cao, G.: Roadcast: A Popularity Aware Content Sharing Scheme in VANETs, Mobile Computing and Communications Review, vol. 13, pp. 1–14 (2009). https://doi.org/10.1145/1740437.1740439

9. Taherkhani, N., pierre, S.: Prioritizing and scheduling messages for congestion control in vehicular Ad hoc networks. Comput. Netw. 108, 15–28 (2016). https://doi.org/10.1016/j.comnet.2016.06.027

Hardware Implementation of Simeck Cipher as a Lightweight Hash Function

Prachin Bhoyar, Sanjay Dhok and Raghavendra Deshmukh

Abstract   Internet of Things brings a galore of lightweight connected devices which triggered a need for befitting security mechanisms. A lightweight hash function is one of the major step taken to accomplish that requirement. In this paper, capabilities of Simeck cipher for implementing hash function are explored. A novel architecture is proposed that serve as a cipher as well as hash function. It nullifies the use of two separate functions in power constrained devices where both the operations are required. The preliminary requirements for power and area are ascertained by implementing it on FPGA and developing an ASIC prototype. The hardware utilization and a number of clocks required are compared with SHA and another hash function and the proposed hash function performs better in these aspects. The basic security analysis of the hash function is also performed.

Keywords   Simeck · Simon · IoT · Security · Lightweight hash function

1   Introduction

The internet of things (IoT) is continuously expanding its boundaries and reaching to the levels which are yet unexplored. In [4], various communication technologies for IoT and their associated security challenges are presented. Growing use of low-cost radio frequency identification (RFID) tags in IoT applications put forward many security challenges to resolve [12]. Many hash functions have been proposed and implemented in the literature. The cost of implementation of two popularly known

P. Bhoyar (✉) · S. Dhok · R. Deshmukh
Visvesvaraya National Institute of Technology, Nagpur, India
e-mail: prachin127@gmail.com

S. Dhok
e-mail: sbdhok@ece.vnit.ac.in

R. Deshmukh
e-mail: rbdeshmukh@ece.vnit.ac.in

hash families SHA and MD5 is over 8k gates equivalent (GE) [6] which takes a significant area considering the limitations of RFID tags.

Lightweight hash function becomes a new paradigm in recent studies. Few lightweight hash functions like QUARK, GLUON, SPONGENT, and PHOTON have already been introduced [1, 5, 7, 10]. Researchers are still working on the cryptanalysis of these functions and improvements in their software and hardware architectures. Hardware implementation of these lightweight hash functions needs around 1.5k GE.

The lightweight block cipher families, Simon, and Specks are proposed by the National Security Agency of USA and their performances in terms of power, area, throughput is better [2]. Moreover, Simeck family [15] combines the advantages of both Simeck and Specks and shows improved performance in terms of hardware implementation. Simeck cipher can be implemented in bit-serial or parallel architecture. The GE required for the implementation of Simeck is very much less compared to the other block ciphers and that makes it suitable for power constrained devices.

In this paper, it is proposed that Simeck cipher can be utilized to design a lightweight hash function. It offers required security to the power constrained devices. In the paper, circuit implementation of Simeck32/64 is presented, however, the circuit implementation of all the remaining ciphers in the family can be realized in a similar fashion.

2 Simeck Cipher

Simeck is symmetric, block cipher family. It is represented as Simeck $2n/mn$, where $2n$ is plain text size and mn is key size. Three distinct sizes are made available for word size (n) viz. 16, 24, and 32 while key words (m) may vary from two to four. In the hardware architecture, it consists of two major blocks, round function and key scheduling algorithm. The value of m and n decides the number of rounds to be executed for the completion of operation. Number of rounds and key size according to m and n is shown in Table 1.

Table 1 Simeck parameters

Word size (n)	Key words (m)	Key size (mn)	Number of rounds
16	4	64	32
24	3, 4	72, 96	36, 36
32	3, 4	96, 128	42, 44
48	2, 3	96, 144	52, 54
64	2, 3, 4	128, 192, 256	68, 69, 72

Here, plaintext is split into two equal sized words. Thereafter, round operation is performed on these words. Let h_o represent the higher word and l_o represent the lower words of plaintext. The round function is defined as:

$$R(h_i, l_i) = (l_i \oplus f(h_i) \oplus k_i, h_i) \tag{1}$$

where k_i is the round key and function f is defined as

$$f(h_i) = (h_i \wedge (h_i \ll 5) \oplus (h_i \ll 1)) \tag{2}$$

where '$\wedge$' and '$\ll$' are AND operator and left shift operator respectively.

Figure 1 illustrates round function of Simeck cipher. Mainly, "XOR", "AND", and "SHIFT" operations are performed on the words. Key expansion algorithm provides a specific Round Key k_i to each round. The main key (K) is parted into four words (m_2, m_1, m_0, k_0). For updating key values, the round function gets executed.

$$\begin{aligned} k_{i+1} &= m_i \\ m_{i+3} = k_i \oplus f(m_i) &\oplus C \oplus (Z_j)_i \end{aligned} \tag{3}$$

where Z_j is simeck constant and $(Z_j)_i$ represent the ith bit of Z_j while $C = 2^n - 4$, is a binary constant.

Fig. 1 Round function of Simeck Cipher
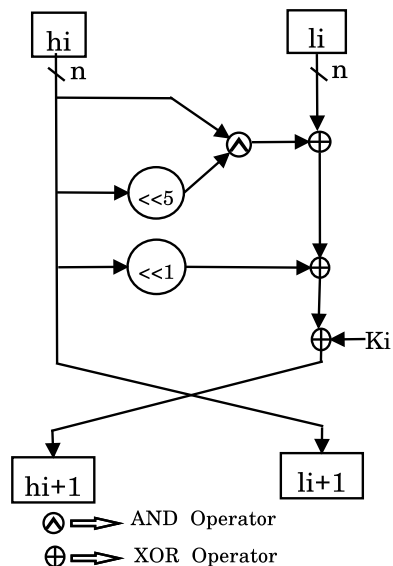
Fig. 2 Matyas-Meyer-Oseas
scheme
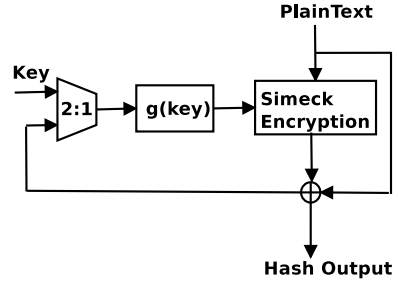
3 Simeck Hash Generator

The Matyas-Meyer-Oseas compression function is used in simeck hash generator. The scheme takes plain text (x_i) and key (y_i) as inputs is shown in Fig. 2. A function "g" is performed on the key. In this function, every eighth bit in the key is deleted and at last eight zeroes are appended at MSB. Then second and last bit from last is replaced by "10". As Simeck32/64 is studied for this paper, function "g" for key size 64 is as follows:

Let $key = y_{63}y_{62}y_{61}y_{60}\ldots y_0$ then

$$g(key) = 01000000y_{63}y_{62}y_{61}y_{60}y_{59}y_{58}y_{57}y_{55}\ldots y_0 \qquad (4)$$

The output of "g" function is then given to Simeck cipher. The output of the simeck cipher again XORed with the plain text to obtained the next hash value. The final output y_o is defined by the iterated formula

$$y_i = E_{g(y_{i1})}(x_i) \oplus x_i \qquad (5)$$

for $1 \leq i \leq t$ and t represents a constant predefined initial value 4. After 4 cycles, the final hash output is obtained.

3.1 Simeck Architecture

For encryption, Simeck32/64 with parallel architecture proposed in our previous work [3] is used. In this system, as there is no use of decryption, it is removed from the architecture. The remaining datapath of Simeck32/64 is same and shown in Fig. 3.

Various lightweight RFID authentication protocols can be used for IoT devices [8, 13, 14]. A unique hash function has been generated for authentication of tag and reader, and tag information is then correlated with the server database.

Fig. 3 Data path for
Simeck32/64 encryption
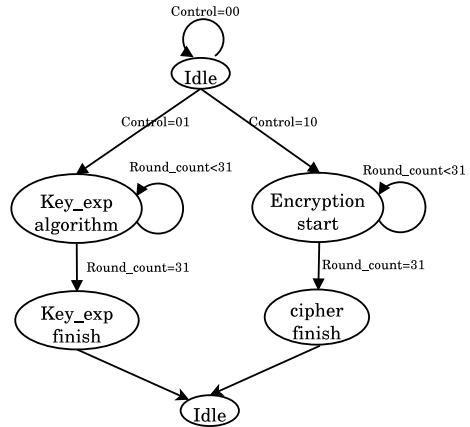
4   Hardware Implementation

The proposed hash function is simulated on Xilinx Isim and then synthesized on Kintex-7 KC705 evaluation board. The clock frequency of 1 MHz is set for the operation. The system consumes 420 μW dynamic power at that frequency. Hardware utilization summary for the proposed system is given in Table 2.

Application-specific integrated circuit (ASIC) prototype of the proposed system is also designed using SCL 180 nm CMOS library with Cadence tool. As the system uses parallel architecture, 2148 GE is required while the dynamic power is 204.246 μW.

Table 3 shows the comparison of the proposed hash with other hash architectures. Being a lightweight, it requires lesser GE. The number can be further reduced by employing bit serial architecture at the cost of additional clock cycles.

Table 2   Hardware utilization for the proposed hash function

No of clock cycles	Slice registers	Slice LUTS	Dynamic Power in μW
68	237	165	420

Table 3   Comparision of the proposed system with other hash functions

Architecture	GE	Clock required
ν-Tav-160 [9]	5030	448
SHA-1 [11]	6122	344
SPONGENT-160 [5]	2406	90
PHOTON-160 [5]	2849	180
D-QUARK [5]	3695	88
The proposed hash	2148	68

5   Security Analysis

5.1   *Preimage Resistance*

Preimage resistance gives the information difficulty the hash function impose while recovering the original message from the hash output. As simeck cipher involve 32 rounds and it is already that out of those maximum 20 rounds can be attacked with time and data complexity of 2^{34} and $2^{31.5}$ [15]. The function "g" introduced and four outer rounds in Matyas-Meyer-Oseas Scheme increase the complexity multiple times. Therefore, it is very difficult to recover the original message with brute force.

5.2   *Second Preimage Resistance*

The second preimage resistance represents the collision of two plain text messages which generate the same hash output. Simeck cipher follows the avalanche criteria and single bit change in input produce more than 50% changes in the output. The criteria are tested in our previous work [3]. The modifications in the proposed architecture further boost the claim.

5.3   *Collision Attack*

In a collision attack, two inputs produce the same hash value. Covariance and histogram analysis has been performed on simeck and it shows that bit patterning in simeck is stirred, therefore chances of two input producing same output are nullified. Furthermore, the four outer rounds in Matyas-Meyer-Oseas Scheme improves the resistivity against such attacks.

6   Conclusion

A hash function is derived from Simeck cipher which is smaller, faster and can be used as an alternative to SHA-X. It has low GE count, lesser complexity and offers a required security measure, therefore it is a viable option for IoT applications such as RFID. Furthermore, it can be used along with the encryption block, that needs only few changes in the architecture. A control mux can be used to use select Simeck as a hash function or encryption or to deselect the entire block. The behavior of the proposed hash function along with application layer protocols such as MQTT or CoAP can also be explored further. A study of the entire simeck family and various architectural strategies for implementing a more efficient hash function can be performed.

References

1. Aumasson, J.P., Henzen, L., Meier, W., Naya-Plasencia, M.: Quark: a lightweight hash. J. Cryptol. **26**(2), 313–339 (2013)
2. Beaulieu, R., Treatman-Clark, S., Shors, D., Weeks, B., Smith, J., Wingers, L.: The simon and speck lightweight block ciphers. In: Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE, pp. 1–6. IEEE (2015)
3. Bhoyar, P., Dhok, S., Deshmukh, R.: Hardware implementation of secure and lightweight Simeck32/64 cipher for IEEE 802.15. 4 transceiver. AEU-Int. J. Electron. Commun. **90**, 147–154 (2018)
4. Bhoyar, P., Sahare, P., Dhok, S., Deshmukh, R.: Communication technologies and security challenges for internet of things: a comprehensive review. AEU-Int. J. Electron. Commun. (2018)
5. Bogdanov, A., Knezevic, M., Leander, G., Toz, D., Varici, K., Verbauwhede, I.: Spongent: the design space of lightweight cryptographic hashing. IEEE Trans. Comput. **62**(10), 2041–2053 (2013)
6. Feldhofer, M., Rechberger, C.: A case against currently used hash functions in RFID protocols. In: OTM Confederated International Conferences on the Move to Meaningful Internet Systems, pp. 372–381. Springer (2006)
7. Guo, J., Peyrin, T., Poschmann, A.: The photon family of lightweight hash functions. In: Annual Cryptology Conference, pp. 222–239. Springer (2011)
8. He, D., Zeadally, S.: An analysis of rfid authentication schemes for internet of things in healthcare environment using elliptic curve cryptography. IEEE Int. Things J. **2**(1), 72–83 (2015)
9. Martin, H., Lopez, P.P., San Millan, E., Tapiador, J.E.: A lightweight implementation of the tav-128 hash function. IEICE Electron. Express **14**(11), 20161255–20161255 (2017)
10. Mukundan, P.M., Manayankath, S., Srinivasan, C., Sethumadhavan, M.: Hash-one: a lightweight cryptographic hash function. IET Inf. Secur. **10**(5), 225–231 (2016)
11. ONeill, M., et al.: Low-cost sha-1 hash function architecture for RFID tags. RFIDSec **8**, 41–51 (2008)
12. Pateriya, R., Sharma, S.: The evolution of RFID security and privacy: a research survey. In: 2011 International Conference on Communication Systems and Network Technologies (CSNT), pp. 115–119. IEEE (2011)
13. Peris-Lopez, P., Hernandez-Castro, J.C., Estevez-Tapiador, J.M., Ribagorda, A.: Lightweight cryptography for lowcost RFID tags. Secur. RFID Sens. Netw. 121–150 (2016)
14. Tewari, A., Gupta, B.: Cryptanalysis of a novel ultra-lightweight mutual authentication protocol for iot devices using rfid tags. J. Supercomput. **73**(3), 1085–1102 (2017)
15. Yang, G., Zhu, B., Suder, V., Aagaard, M.D., Gong, G.: The simeck family of lightweight block ciphers. In: International Workshop on Cryptographic Hardware and Embedded Systems, pp. 307–329. Springer (2015)

Comparative Study of PSO-Based Hybrid Clustering Algorithms for Wireless Sensor Networks

Ghanshyam Singh, Shashank Gavel and Ajay Singh Raghuvanshi

Abstract Clustering is a task which creates groups depending upon the presence of similarity between the data objects. Many clustering algorithms exist, which are capable of creating well-defined clusters. One of the popular algorithms is K-means, which is generally used for data clustering where performance is dependable on initial state of centroid but have limitation of trapping in local optima. Besides K-means, K-harmonic means, and Fuzzy C-means are also popular algorithms used for data clustering but again they have the same limitation of trapping in local optima. So this creates problem while handling anomaly existing dataset in wireless sensor network. In this paper, an analysis of best suitable hybrid clustering algorithm is brought for a congregation of normal and anomalous dataset by using a stochastic tool Particle Swarm Optimization (PSO) by utilizing different sensor datasets. The results are encouraging in terms of best suitable fitness function and low computational time.

Keywords Wireless sensor network · K-Means · K-Harmonic means · Fuzzy C-Means · Particle swarm optimization

1 Introduction

Presently, Wireless Sensor Networks (WSN) are widely used for many applications such as healthcare, military, industrial monitoring, and Internet of Things where nodes are deployed in large numbers to gather information by which large number of datasets generated. In a complete system different type of datasets are produced by different node architectures. Anomaly is one of the major issues which some

G. Singh (✉) · S. Gavel · A. S. Raghuvanshi
Department of Electronics and Telecommunication, National Institute of Technology, Raipur 492001, India
e-mail: gsingh.phd2016.etc@nitrr.ac.in

S. Gavel
e-mail: sgavel.phd2016.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

time occurs in WSN dataset in the form of noise. For proper handling of anomalous data present in WSN, clustering plays an important role. Clustering algorithms help the system in grouping the normal and anomalous data separately [1]. To extract efficient information and knowledge from these large number of datasets data mining technique is used. In data mining, a task is required to categorize same type of datasets in a particular group, so that suitable processing techniques are applied on similar type of datasets to overcome the limitations of large processing time and complexity [2].

Clustering is a popular task in data mining to group the same type of data objects in a manner that the similarity between data objects of same cluster is maximum but quite dissimilar from data objects of different clusters [3]. In clustering, there is no requirement of any supervision to recognize and find the patterns and trends in large sum of data that is why it is also called unsupervised learning. Clustering algorithms are used in different applications such as image processing [4], data compaction [5], medicine [6], marketing [7], and anomaly detection[8].

There are a large number of algorithms available to perform clustering, which is further distinguished as hierarchical and partitioning algorithms. In hierarchical algorithms, structure of data is created by splitting and merging data based on parallel criterion whereas in partitioning algorithms clustering criterion is optimized by proper shuffling of data among clusters [9]. It has been observed that partitioning algorithms have upper hand on hierarchical algorithms in terms of efficiency and popularity [10]. Due to a wide range of available applications, data types and objectives, a single algorithm cannot serve all the requirements.

The widely used partitioning algorithm is K-means algorithm due to its ease of implementation, simplicity, and fast processing with linear time complexity. However, it has limitation due to high dependence on cluster centroid initial selection which further directed to get stuck in local minima during the process of minimizing the objective function [11].

An improved version of K-means, K-harmonic means algorithm was proposed which have the same partition mechanism of data objects into K different clusters depending on application purposes. K-harmonic mean algorithm overcomes the limitation of sensitivity toward initial cluster centers but cannot resolve the limitation of trapping into local minima [12].

After the fuzzy theory introduced, researchers applied fuzzy theory in clustering through which data objects partially assigned to multiple clusters. Fuzzy C-means is one of the most popular and effective fuzzy clustering algorithms but it applied center point selection in random manner which further results falling into local optima through iterative process [13].

Particle Swarm Optimization (PSO) is a popular optimization tool based on population. Various function optimization problems can be easily solved by applying PSO. A Fuzzy version of Particle Swarm Optimization (FPSO) is also described in [14].

In this paper, we have optimized the different clustering techniques, i.e., K-means, K-harmonic means, and fuzzy C-means algorithm using PSO. For analyzing the compatibility of PSO with different clustering algorithms, different sensor datasets

from various standard laboratories are used. Manual anomalies are inserted to these datasets for better clustering in the form of normal and anomalous data and fitness cost is calculated with computational time. The rest of the paper is organized in a manner as in Sect. 2 we investigate the related background of clustering algorithms and Sect. 3 introduces results and discussion on proposed work. Finally, Sect. 4 concludes this work.

2   Background on Clustering Algorithms

This paper focuses basically on three of the most popular and efficient clustering algorithms named as K-means algorithm, K-harmonic means algorithm, and fuzzy C-means algorithm. PSO as a tool is used to calculate the Euclidian distance based on the output of previously mentioned clustering algorithms which provide a comparative analysis on best suitable hybrid clustering algorithm to cluster the normal and anomalous dataset. So, in this section a brief description of these clustering algorithms is mentioned for understanding the mechanism of clustering algorithms.

2.1   K-Means (KM) [5]

In K-means clustering, each of the objects represents a value in an individual cluster. This cluster is formed depending upon the data objects close to one another. The cluster centroid is obtained by calculating the mean for the data objects for a particular cluster. The distance between data objects and cluster centroid is found using the Euclidean distance metrics. Let $X = (x_1, ..., x_n)$ be a set of data objects, $C = (c_1, ..., c_k)$ be the cluster centers. The clustering algorithm aims to find a partition of k clusters that includes the data within the similar group which are close to each other. The objective function for K-mean clustering is given below:

$$f(X, C) = \sum_{l=1}^{k} \sum_{X_i \in c_i} d(X_i, Z_l)^2 \tag{1}$$

where $d(X_i, Z_l)$ is the variation between the data object and the cluster center C (Z_l). This variation is obtained using the mean value of different objects within the cluster. In general Euclidean distance is used to find variation between data objects and the cluster center, which is expressed as follows:

$$d(X_i, X_j) = \sqrt{\sum_{p=1}^{d} \left(x_i^p - x_j^p \right)^2} \tag{2}$$

The K-means clustering algorithm is a popular and efficient clustering method used in many applications, but while handling large datasets, this algorithm gets stuck to a local optimum value and is unable to find the global optimum. This problem tends to local conjunction and convergence of algorithm. This problem can be solved using an efficient optimization algorithm.

2.2 K-Harmonic Means (KHM) [15, 16]

K-harmonic means addresses the intrinsic problem of K-means by using harmonic mean for finding the centroid that replaces the obtained minimum distance between the data point to the centers. For each data point close to anyone center this harmonic mean provides a good score, which is one of its properties. Using the value of different data objects X and cluster centroid C as explained in K-means clustering, the objective function representing K-harmonic means clustering algorithm is given by

$$KHM(X, C) = \sum_{i=1}^{n} \frac{k}{\sum_{j=1}^{k} \frac{1}{||x_i - c_j||^p}} \tag{3}$$

where p is an input parameter ($p \geq 2$).

2.3 Fuzzy C-Means (FCM) [13]

In fuzzy C-means algorithm, partition of n data objects is done in R^d dimensional space into k fuzzy clusters. This fuzzy clustering of data objects is obtained using fuzzy matrix denoted by μ_{ij} where i signifies number of rows and j signifies number of columns. The objective function of FCM algorithm is given by

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m d_{ij} \tag{4}$$

where d_{ij} is the Euclidean distance and m is the weighting exponent which controls the fuzziness of data object with respect to resulting cluster. Combining the data object X and cluster center C, Euclidean distance is given by

$$d_{ij} = ||X_i - C_j|| \tag{5}$$

$$C_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m X_i}{\sum_{i-1}^{n} \mu_{ij}^m} \tag{6}$$

where C_j is the centroid for jth cluster.

FCM is very sensitive to the initial values of cluster centers which likely falls it into local optima. To cluster large data using FCM, this problem is to be resolved.

2.4 Particle Swarm Optimization (PSO)

PSO is one of the popular metaheuristic population-based optimization technique which is inspired by flocking of birds originally introduced and designed by Kennedy et al. [17]. The flow of this optimization algorithm depends upon the initial position and velocity of the particle. These attributes are used for number of iterations for which the *pbest* and *gbest* values are calculated which represent the personal best and global best of the system. The values of velocity and position for an individual particle is obtained using Eq. (7)

$$V(t+1) = w.V(t) + c_1 r_1 (pbest(t) - X(t))$$
$$+ c_2 r_2 (gbest(t) - X(t)); \quad k = 1, 2, \ldots P \quad (7)$$

$$X(t+1) = X(t) + V(t+1)$$

where X and V represent the position and velocity of a particle, w represents the inertia weight whereas c_1 and c_2 are positive acceleration constant values [18]. P represents the number of particle and r_1 and r_2 are the random values $(r_1, r2) \sim [0, 1]$.

3   Results and Discussion

For the analysis of best suitable hybrid clustering algorithm to cluster the normal and anomalous dataset, we have used PSO as mentioned in Eq. (7). Different hybrid clustering algorithms are obtained by optimizing the individual techniques using PSO. For PSOKMeans Eq. (1) is optimized using Eq. (7), for PSOKHarm Eq. (3) is optimized using Eq. (7), and for PSOFCM Eq. (4) is optimized using Eq. (7). Data conditioning of different datasets is carried out by inserting different types of anomalies to each and every node data. These datasets are taken from various standard laboratories such as **sensor dataset 1** is taken from [19], **sensor dataset 2** is taken from [20], and **sensor dataset 3** is taken from [21]. The detailed data conditioning is explained below:

1. **Sensor Dataset 1** [19]:
   This dataset is obtained by deploying nodes in different manners, i.e., singlehop and multihop network. For each of the network, a dataset is obtained which contains both normal and anomalous data.

2. **Sensor Dataset 2** [20]:

   The dataset was generated by well-known Intel Berkeley lab experiment, 54 sensor nodes were used for monitoring the environment. We have utilized two-node data, i.e., node 7 and node 9 and have inserted manual anomalous data in the form of random noise.

3. **Sensor Dataset 3** [21]:

   This dataset was taken from the chemical sensors deployed for continuous monitoring of the environment. Depending upon the scenario we have managed to take data from each environment such as background activity, banana, and wine. In these datasets manual anomaly in the form of shot noise is inserted.

By utilizing the abovementioned datasets, analysis of fitness functions for different hybrid methods is carried out. These hybrid algorithms are PSOKMeans, PSOKHarm, and PSOFCM. The results are compared in Tables 1, 2 and 3, in which Table 1 shows the comparison of Sensor Dataset 1 and the results show better performance for PSOFCM as compared to other hybrid algorithms. Similarly Table 2 shows the comparison of Sensor Dataset 2 and the results show better performance for PSOKMeans as compared to other hybrid algorithms. Table 3 shows the comparison of Sensor Dataset 3 and the results show better performance for PSOKHarm as compared to other hybrid algorithms. So by analyzing and comparing the different hybrid clustering algorithms it can be concluded that performance of these algorithms varies while handling various kinds of anomalous datasets.

Table 1 Comparison of fitness cost for Sensor Dataset 1 having regenerative feedback noise

Dataset	PSOKMeans	PSOKHarm	PSOFCM
Singlehop 1	**5.0936**	5.1163	5.1281
Singlehop 4	3.0470	3.1496	**1.6589**
Multihop 1	2.7901	2.7887	**1.5153**
Multihop 3	**5.9375**	6.0849	6.0180
Computational time (in s)	6.014	**5.7711**	5.805

Table 2 Comparison of fitness cost for Sensor Dataset 2 having random noise

Dataset	PSOKMeans	PSOKHarm	PSOFCM
Node 7	**2.3168**	2.4810	2.3168
Node 9	**8.3497**	8.3474	8.3712
Computational time (in s)	**6.82**	7.159	6.95

Table 3 Comparison of fitness cost for Sensor Dataset 3 having shot noise

Dataset	PSOKMeans	PSOKHarm	PSOFCM
Background activity	0.4014	**0.3923**	0.3951
Banana dataset	**0.2566**	0.2568	0.2566
Wine dataset	0.0808	0.0808	**0.0807**
Computational time (in s)	6.11	**5.6723**	5.8622

4 Conclusion

Different individual clustering algorithms face problem of local sticking at initial value. Although for handling large noisy datasets in WSN, these clustering algorithms limits their performance which degrades the overall performance of the system. On the other hand, PSO, an optimization algorithm helps this clustering algorithm to resolve their problems and in finding the best possible global optimum values. In this paper, we have analyzed three hybrid algorithms PSOKMeans, PSOKHarm, and PSOFCM in terms of fitness function and computational time. These algorithms are tested using different anomalous WSN datasets. The results show variation in fitness cost and computational time for different types of anomalies present in these datasets. This paper also compares the performance in presence of different types of noisy datasets for WSN and also suggests the best suitable hybrid clustering method. In future, this work can be extended for designing entropy-based correlative system for identification of anomaly.

References

1. Patcha, A., Park, J.M.: An overview of anomaly detection techniques: existing solutions and latest technological trends. Comput. Netw. **51**(12), 3448–3470 (2007)
2. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier (2011)
3. Jain, A.K.: Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. **31**(8), 651–666 (2010)
4. Xia, Y., Wang, T., Zhao, R., Zhang, Y.: Image segmentation by clustering of spatial patterns. Pattern Recogn. Lett. **28**(12), 1548–1555 (2007)
5. Yang, S., Wu, R., Wang, M., Jiao, L.: Evolutionary clustering based vector quantization and SPIHT coding for image compression. Pattern Recogn. Lett. **31**(13), 1773–1780 (2010)
6. Liao, L., Lin, T., Li, B.: MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach. Pattern Recogn. Lett. **29**(10), 1580–1588 (2008)
7. Sağlam, B., Salman, F.S., Sayın, S., Türkay, M.: A mixed-integer programming approach to the clustering problem with an application in customer segmentation. Eur. J. Oper. Res. **173**(3), 866–879 (2006)
8. Moshtaghi, M., Havens, T.C., Bezdek, J.C., Park, L., Leckie, C., Rajasegarar, S., Palaniswami, M.: Clustering ellipses for anomaly detection. Pattern Recogn. **44**(1), 55–69 (2011)
9. Berkhin, P.: A survey of clustering data mining techniques. In: Grouping Multidimensional Data, pp. 25–71. Springer, Berlin, Heidelberg (2006)

10. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. (CSUR) **31**(3), 264–323 (1999)
11. Kao, Y.T., Zahara, E., Kao, I.W.: A hybridized approach to data clustering. Expert Syst. Appl. **34**(3), 1754–1762 (2008)
12. Cui, X., Potok, T. E., Palathingal, P.: Document clustering using particle swarm optimization. In: Proceedings IEEE of Swarm Intelligence Symposium, 2005, pp. 185–191 (2005)
13. Bezdek, J.C.: Fuzzy mathematics in pattern classification. Ph. D. Dissertation, Applied Mathematics, Cornell University (1973)
14. Pang, W., Wang, K., Zhou, C., Dong, L.: Fuzzy discrete particle swarm optimization for solving traveling salesman problem. In: Proceeding of Fourth International Conference on Computer and Information Technology, pp. 796–800. IEEE CS Press (1973)
15. Hamerly, G., Elkan, C.: Alternatives to the k-means algorithm that find better clusterings. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management, pp. 600–607. ACM (2002)
16. Ünler, A., Güngör, Z.: Applying K-harmonic means clustering to the part-machine classification problem. Expert Syst. Appl. **36**(2), 1179–1194 (2009)
17. Kennedy, J., Eberhart, R.C., Shi, Y.: Swarm Intelligence. Morgan Kaufmann Publishers. Inc., San Francisco, CA (2001)
18. Izakian, H., Abraham, A.: Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. Expert Syst. Appl. **38**(3), 1835–1838 (2011)
19. Suthaharan, S., Alzahrani, M., Rajasegarar, S., Leckie, C., Palaniswami, M.: Labelled data collection for anomaly detection in wireless sensor networks. In: 2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 269–274. IEEE (2010)
20. C.G.S.M.M.P. Bodik, P., Hong, W., Thibaux, R.: Ibrl dataset. http://db.csail.mit.edu/labdata/labdata.html
21. Huerta, R., Mosqueiro, T., Fonollosa, J., Rulkov, N.F., Rodriguez-Lujan, I.: Online decorrelation of humidity and temperature in chemical sensors for continuous monitoring. Chemometr. Intell. Lab. Syst. **157**, 169–176 (2016)

Modified Cluster Head Election Scheme Based on LEACH Protocol for MI-Driven UGWSNs

A. Laxmi Prasanna and Vinay Kumar

Abstract To enlarge the lifetime of any microsensor network the factors like energy consumption rate, efficient routing protocols, and media access plays an important role. In this article, we incorporated the concept of low-energy adaptive clustering hierarchy (LEACH) protocol in homogeneous magnetic induction (MI) communication-based underground wireless sensor network (UGWSN), i.e., in dry soil medium. Furthermore, we extended our study to its heterogeneous counterpart. Herein, a distinct energy model is used for energy calculations. Apart from this, we developed an energy-efficient threshold-based cluster head (CH) election technique which considers the initial energy of every sensor node in each round. The simulation results show that there is a consistent performance in prolonging the wireless sensor network lifetime which is expressed in terms of energy dissipation of network and the number of alive nodes.

Keywords Underground wireless sensor networks (UGWSNs) · MI communication · LEACH protocol and energy heterogeneity

1 Introduction

MI-based communication is currently traversed approach in nonconventional media. For nonconventional media like underground and underwater, electro magnetic (EM) wave system is not suitable due to few problems like high path loss, large antenna size, and dynamic channel nature. Hence, MI waveguide-based approach for communication is proposed to rectify these effects in nonconventional media (soil, rocks, and water) [1]. This MI waveguide approach for communication is implemented to convey the effect of high attenuation in MI waves across the soil (dry or wet). MI

A. L. Prasanna · V. Kumar (✉)
Visvesvaraya National Institute of Technology, Nagpur 440010, India
e-mail: vinayrel01@gmail.com

A. L. Prasanna
e-mail: akulalaxmiprasanna@gmail.com

approach uses the near field of magnetic field in low frequency to initiate wireless communication. In MI-based communication, primary and secondary coils of a transformer act as transmitter and receiver, respectively. Various parameters such as frequency of operation, number of turns, length of coils, and medium (properties of soil) play a key role in MI communication. In MI communication, magnetic field is generated by a time-varying current-carrying coil [2, 3]. In terrestrial environment, MI waves attenuate rapidly and can attain a good range in nonconventional media. In WSNs, powered batteries are required for sustaining the sensor nodes to some more extent. Since the transmission circuitry in the network consumes more amount of energy, an energy-efficient protocol needed to be designed to reduce the energy dissipation of the whole sensor network and to increase the network lifetime [4–6]. Hence, an energy-efficient protocol named as low-energy adaptive clustering hierarchy (LEACH) [7] is implemented in homogeneous environment.

As yet, there is no related work on classical routing protocols such as direct transmission, minimum transmission energy (MTE), and LEACH in the era of MI communication. Hence this paper explains about LEACH protocol (homogeneous and heterogeneous) in MI era. In the following sections, related work on LEACH protocol, contributions, and implementation of this protocol in homogeneous and heterogeneous model followed by new proposed LEACH heterogeneous model in MI perspective are shown.

1.1 Motivation

We have effectively implemented the LEACH protocol based on energy heterogeneity in MI communication in dry soil medium. In this module, the related work in MI-based communication is discussed.

Sharma et al. [8] developed a review on magnetic induction-based nonconventional media communications. This paper mainly explains various applications in MI-based approach and also discussed qualitative analysis on widely used communication techniques in nonconventional media.

Sun and Akyildiz [9] proposed a model that characterizes the two important factors of MI system, namely, bandwidth and path loss. In this paper, MI-based waveguide technique is developed to reduce the high path loss of EM wave system and also a comparative analysis is made on traditional EM wave system, MI system and developed MI waveguide system in nonconventional media.

Kumar [10] explains about the design of MI-based energy-efficient WSNs for nonconventional media using multilayer transmitter enabled novel energy model. In this paper, four feasible compensation circuits and multilayer coil structure are developed for the layout of MI transceiver. A novel energy model which illustrates the dissipation energy in the transceiver circuit on MI-based in nonconventional media is proposed.

Heinzelman et al. [7] implemented an application-specific protocol architecture for wireless microsensor networks in which an energy-efficient protocol is designed known as low-energy adaptive clustering hierarchy (LEACH). This paper aims at

analyzing the protocol architecture that explains efficient routing and media access to acquire the better network lifetime. Also this protocol explains a new distributed cluster formation technique and algorithms for adapting clusters efficiently.

Zhou et al. [11] proposed a stable election protocol based on energy heterogeneity in wireless sensor networks (WSNs) and also a new energy-efficient routing technique, i.e., energy dissipation forecast method (EDFM) is proposed for the efficient election of cluster heads. In this article, a mathematical model on energy consumption in heterogeneous WSNs is observed in detail.

1.2 Contributions

In this paper, the notion of energy-efficient LEACH protocol on MI-based communication is introduced to prolong the microsensor network lifetime in terms of energy consumption, media access, and efficient routing in underground wireless sensor network (dry soil medium).

The principal contributions of this paper are as follows:

1. Mathematical analysis of energy formulation of LEACH homogeneous and heterogeneous environment in MI perspective.
2. Evaluation of the election of cluster head probability per round in heterogeneous WSN model.
3. Comparative analysis of LEACH protocol among the three models (homogeneous, heterogeneous, and proposed heterogeneous) on MI-based approach in terms of average energy of each node and number of alive nodes.

2 LEACH Protocol

Several energy-efficient routing protocols and clustering techniques are proposed to increase the sensor network lifetime, to improve low-energy media access control and to reduce the dissipation energy of whole system. Therefore, one of the basic primitive protocol known as LEACH in homogeneous model explains the abovementioned requirements. LEACH protocol has some assumptions such as sensor nodes having enough transmit power, computational ability, and equal initial energy which maybe possible due to modern advancement in technology [12]. This protocol action is segregated into rounds, each round consists of setup phase and steady-state phase. Setup phase comprises cluster head (CH) election and cluster formation mechanisms while, steady-state phase explains about the data transmission. In the election of CH, the probability of each node i to become a CH per round r in homogeneous model is given as

$$P_i = \frac{P}{1 - Pmod(r, round(\frac{1}{P}))} \tag{1}$$

where P is the average CH election probability

Table 1 Parameters used in
the paper and it's definition

Parameter	Defination
N	Total MI nodes in the network
K	Number of clusters
M	Side of sensing region (m)
l	Number of bits
$E_{Tx-elec}$	Tx electronics energy
$E_{Rx-elec}$	Rx electronics energy
d_{toCH}^6	Distance from non-CH to CH
d_{toBS}^6	Distance from CH to BS
I	Current in transmitter coil (A)
a	Radius of coil (m)
n	Number of turns of transceiver coils
d	Distance between the Tx and Rx (m)
G	Attenuation factor
δ	Skin depth
P	Avg CH election probability
m	Fraction of advanced nodes (0.1)

After the election of CHs, each CH node gives an acknowledgement through an advertisement message (ADV) to all non-CHs using nonpersistent carrier sense multiple access (CSMA) MAC protocol. Also each non-CH sends a message request to the selected cluster head using this protocol and now each CH establishes a TDMA schedule for the data transmission within the cluster. This TDMA schedule reduces the energy consumption in non-CHs and collisions among the messages. Next in steady-state phase, few assumptions are made like nodes are perfectly synchronized in time and start the setup phase at equal times. Using TDMA schedule, data from non-CHs is transmitted to CHs in allocated time intervals. To reduce the effect of inter-cluster interference, every cluster convey the information employing direct sequence spread spectrum (DSSS) and uses a unique spreading code [13]. Now the data is sent from CHs to sink or base station (BS) using CSMA and a fixed spreading code. The following table represents the various parameters that are considered in this paper and their definitions (Table 1).

2.1 Proposed Heterogeneous WSN on MI-Based Approach

In this segment, we discuss the energy formulation on magnetic induction (MI)-based heterogeneous WSN model in underground medium (dry soil). In general, unlike homogeneous WSNs, heterogeneous WSNs with MI sensor nodes are having different sensing power, computational ability, and various energy levels.

We assume that WSN deploy the sensor nodes uniformly over the square sensing region of side M. This network model comprises two types of sensor nodes, namely, normal nodes and "m" fraction of advanced nodes and assuming the position of sink or base station at the center of square sensing region. Let the normal nodes has initial energy as E_0 and advanced nodes has $(1 + \lambda)E_0$, where advanced nodes have λ times more the initial energy. P be the optimal election probability of a node to become a cluster head and N be the total number of sensors in the network. The election probabilities of two types of nodes are as follows:

$$P_n = \frac{P}{1 + \lambda m} \tag{2}$$

$$P_a = \frac{P(1 + \lambda)}{1 + \lambda m} \tag{3}$$

According to the traditional LEACH protocol (homogeneous model), initially all the sensor nodes have equal energy and enough transmission and computational power. Let K be the number of clusters ($K < N$). In this protocol, nodes categorize into clusters and there will be $\frac{N}{K}$ number of nodes in a cluster with one cluster head (CH), while $(\frac{N}{K}) - 1$ number of non-cluster head nodes. Non-cluster head nodes assemble the data and send to cluster head then, CH nodes further broadcast the information to sink or base station (BS) for processing through MI waveguide.

In general, WSNs in MI communication are restricted to the amount of energy. These battery operated sensor nodes have no source for recharging and can not survive for long time. This shows an impact on the sensor nodes to minimize the dissipation energy. Hence an energy model is required in governing some aspects like efficient routing, clustering, and MAC protocols. As stated by radio energy model, the dissipation energies of cluster heads and non-cluster heads per round to transmit l bits of data in MI perspective [10] are represented as

$$E_{CH} = l(\frac{N}{K} - 1)E_{Rx-elect} + lE_{Tx-elect} + l\alpha E_{MI}(d_{toBS}^6) \tag{4}$$

$$E_{NonCH} = lE_{elect} + l\alpha E_{MI}(d_{toCH}^6) \tag{5}$$

Here, $E_{Tx-elect}$ and $E_{Rx-elect}$ are the electronics energies at transmitter and receiver, respectively, which depends on digital modulation and coding approach, E_{MI} is the dissipation energy in a current-carrying coil and α is a factor that depends merely on skin depth (δ) of medium which is inversely proportional to the frequency of operation and mutual inductance (MI) between transmitter and receiver coils. The value of magnetic permeability remains almost constant in nonconventional media (soil,water and rocks).

$$E_{MI} = \frac{\mu I^2 a^4 n^2}{8d^6} \tag{6}$$

$$MI = \frac{\mu \pi n^2 a^4}{2d^3} G \tag{7}$$

$$G = e^{-\frac{d}{\delta}} \tag{8}$$

where μ is magnetic permeability, d^6_{toCH} is the distance from sensor node to cluster head, and d^6_{toBS} is the distance from cluster head to the sink. We assumed that each cluster is in circular shape and the position of base station (BS) is at center of square sensing region. Hence, the expected values of these distances are shown below,

$$E[d^6_{toCH}] = 0.008 \frac{M^6}{K^3} \tag{9}$$

$$E[d^6_{toBS}] = 0.011 M^6 \tag{10}$$

Our proposed technique is similar to the heterogeneous LEACH protocol except in case of cluster head selection of two types nodes. Equation 1 is the generalized expression of any MI sensor node that represents the probability of each MI node to become a cluster head and is referred from [7]. The following two equations represent the probabilities of each normal and advanced node to become a cluster head at rth round multiplied by a term which is the initial energy of respective nodes (normal and advanced). Where the parameters λ and m are constants.

$$P_{nrm}(i) = \frac{P_n}{1 - P_n mod\,(r,\,round\,(\frac{1}{Pn}))} (\frac{E(i)}{1 + \lambda m}) \tag{11}$$

$$P_{adv}(i) = \frac{P_a}{1 - P_a mod\,(r,\,round\,(\frac{1}{Pa}))} (\frac{E(i)(1 + \lambda)}{1 + \lambda m}) \tag{12}$$

where $E(i)$ is the initial energy of ith sensor node (normal or advanced node) in round r. Excluding the CH election in proposed heterogeneous model, cluster formation algorithm, and steady-state phase mechanism resembles with the traditional LEACH protocol.

3 Results and Analysis

In this segment, the evaluations of standard LEACH protocol (homogeneous and heterogeneous models) and proposed model are performed through MATLAB. The following results (Figs. 1 and 2) show that at 1 KHz frequency of operation, proposed, and heterogeneous models exhibit better performances than homogeneous model in

Fig. 1 Number of alive nodes versus number of rounds in dry soil medium at 1 KHz of frequency
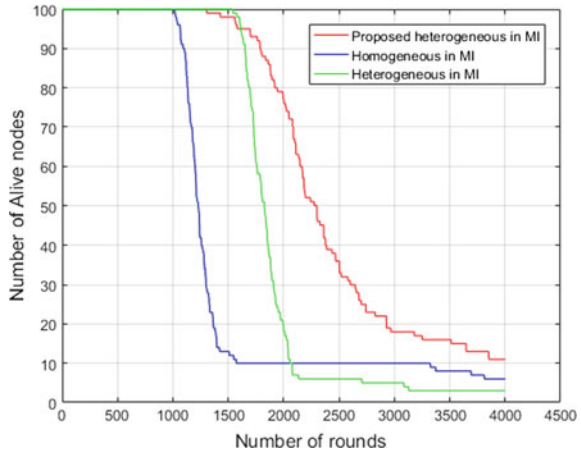
Fig. 2 Average energy of each sensor node versus number of rounds in dry soil medium at 1 KHz of frequency
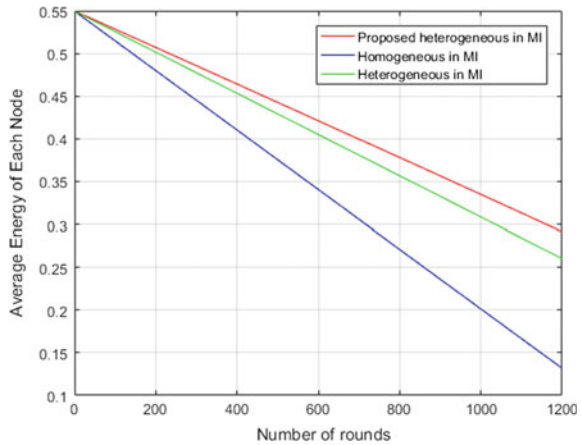
terms of number of alive nodes and average energy of each node. In performing the simulations through MATLAB, Eqs. 2, 3, 11, and 12 are used to evaluate the number of dead nodes and alive nodes per round in the sensor network. Similarly, Eqs. 2, 3, 4, 5, 11, and 12 are used to calculate the average energy of each MI sensor node per round which also depends on the factors like E_{MI}, M, G, d_{toCH}^6, and d_{toBS}^6.

From Fig. 3 it is inferred that as the frequency of operation increases, the energy needed for the transmission decreases and hence the first sensor node dies early in the network. Therefore, this leads to the decrease in the number of alive nodes as the frequency of operation increases. Also, the dissipation energy at transmitter side depends on the coil and circuit parameters such as number of turns of transceiver coils, radius of coil, and distance between the transceivers.

Fig. 3 Number of alive nodes versus number of rounds in dry soil medium at 1 and 3 KHz of frequency
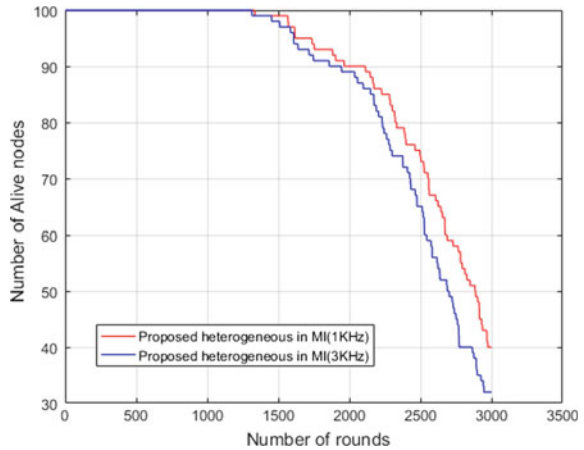
4 Conclusion

In this paper, we have incorporated the concept of traditional LEACH protocol in homogeneous and heterogeneous environment in magnetic induction (MI) communication-based underground wireless sensor network (UGWSN). We developed an energy-efficient threshold-based cluster head (CH) election technique which considers the initial energy of every sensor node in each round. The simulation shows that, as compared to homogeneous energy model, the heterogeneous counterpart provides a consistent sensor network lifetime (in terms of number of alive nodes) and energy efficiency (in terms of residual energy of each sensor per round).

References

1. Sandeep, D.N., Kumar, V.: Review on clustering, coverage and connectivity in underwater wireless sensor networks: a communication techniques perspective. IEEE Access J. **5**, 11176–111999 (2017)
2. Pathak, V., Kumar, V., Barik, R.: Magnetic induction communication-based transceiver coil and waveguide structure modeling for non-conventional WSNs. In: Accepted for Publication in 9th ICCCNT 2018, at IISc Bangalore. IEEE (2018)
3. Tambe, S., Kumar, V., Bhusari, R.: Magnetic induction-based cluster optimization in non-conventional WSNs: a cross layer approach. AEU-Int. J. Electron. Commun. **93**, 53–62 (2018)
4. Kumar, V., Sandeep, D., Yadav, S., Barik, R.K., Tripathi, R., Tiwari, S.: Multi-hop communication based optimal clustering in hexagon and voronoi cell structured WSNs. AEU-Int. J. Electron. Commun. **93**, 305–316 (2018)
5. Yadav, S., Kumar, V.: Optimal clustering in underwater wireless sensor networks: acoustic, EM and FSO communication compliant technique. IEEE Access J. **5**, 12761–12776. https://doi.org/10.1109/ACCESS.2017.2723506

6. Chatterjee, R., Kumar, V.: Energy efficient routing protocol via chain formation in gaussian distributed wireless sensor networks. Int. J. Electron. Lett. Taylor Francis USA. https://doi.org/10.1080/21681724.2017.1279223
7. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An-application protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**(4) (2002)
8. Sharma, A.K., Yadav, S., Dandu, S.N., Kumar, V., Senguptha, J., Dhok, S.B., Kumar, S.: Magnetic Induction-based non-conventional media communications: a review. IEEE Sens. J. **17**(4) (2017)
9. Sun, Z., Akyildiz, I.F.: Magnetic induction communications for wireless underground sensor networks. IEEE Trans. Antennas Propag. **58**(7) (2010)
10. Kumar, V.: Design of magnetic induction based energy efficient WSNs for non-conventional media using multilayer transmitter enabled novel energy model. Accepted IEEE Syst. J. (2018). https://doi.org/10.1109/JSYST.2018.2852487
11. Zhou, H., Wu, Y., Xie, G.: A stable election protocol based on energy dissipation forecast method for clustered heterogeneous wireless sensor networks. In: International Conference on Wireless Communications, Networking and Mobile Computing (2009). https://doi.org/10.1109/WICOM.2009.5304152
12. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of 33rd Hawaii International Conference on System Sciences (2000)
13. Sun, Z., Akyildiz, I.F.: Channel modeling and analysis for wireless networks in underground mines and road tunnels. IEEE Trans. Commun. **58**(6) (2010)

Stable Energy-Efficient Routing Algorithm for Dynamic Heterogeneous Wireless Sensor Networks

Akshay Verma [ID]**, Sunil Kumar** [ID]**, Prateek Raj Gautam** [ID]
and Arvind Kumar [ID]

Abstract The advancement in recent research in the field of dynamic wireless sensor networks opens the major challenges in terms of Qos (Quality of Service), connectivity, reliability, and many more. The work that is being presented in this literature overcomes such issues by implementing a Stable Energy-Efficient Routing Algorithm (SERA) for Dynamic Heterogeneous Wireless Sensor Networks (DHWSNs). The SERA protocol dynamically selects the cluster heads (CHs) and also provides the re-rotation of such selected CHs at the proper time for sync energy consumption in the network. In SERA, the normalized concept of distance and energy is used for selecting the CHs and the connectivity of such CHs with its neighbors is provided on the basis of threshold distance d_o, thereby, increasing the reliability, QoS, and connectivity in the network. The simulation results justified that the SERA protocol withstands for DHWSNs in terms of QoS, connectivity, reliability, and stability.

Keywords Wireless sensor network · QoS · Reliability · Connectivity

1 Introduction

In recent research, dynamic wireless sensor networks (DWSNs) are the major focus for routing developers. A number of users are utilizing the DWSNs to extract their interested purposes. The major challenges in the DWSNs are mobility and connectivity with the users. Since WSN is made up of sensor nodes which are equipped with intelligent but limited-power microcontrollers that can sense, process, receive and/or transmit, and store the data, reliable energy-efficient routing algorithms should be designed. The goal of the routing algorithm is to forward the desired data efficiently to destination/base station (BS). A number of routing algorithms have been designed

A. Verma (✉) · S. Kumar · P. R. Gautam · A. Kumar
Motilal Nehru National Institute of Technology Allahabad, Prayagraj, India
e-mail: akshayv.verma49@gmail.com

A. Kumar
e-mail: arvindk@mnnit.ac.in

on the basis of proactive and reactive topology basis to serve such purposes [8]. Since the reactive topology provides on-demand response to the users/mobile sinks (MSs), it is more popular than that of proactive approaches [8]. But it has the issues related to QoS, reliability, and connectivity among the nodes. Since the sensor's power is limited, sensor nodes are intelligently connected together to form a reliable and energy-efficient routing path so that the overall network performance is enhanced. The broad application areas of WSNs are in military operations, pollution control, vehicle motion smart cities, etc. [13].

A number of cluster-based routing protocols are designed which work on the randomly distributed immobile sensor nodes [2, 6, 7, 9, 14]. But there are very few clustering protocols which work on the routing and connectivity of mobile sensor nodes. These are AODV [10], PUMA [12], and SLMRP [1]. The PUMA protocol routes the data by flooding the data in the network and forming the mesh with the help of a connectivity list and assigning a core to the network. The limitation of this is that the protocol periodically floods the data in the mesh regardless of the interest of the nodes which increases the overhead in the network. To overcome such issue, Abdulwahid et al. developed SLMRP which uses the multicast routing activation timer (MRAT) and path timeout timer (PTT) for load distribution. The discussed protocols [1, 10, 12] have applied the random way point mobility model and applied the partial mobility in the network to the (MSs). Further, in recent research, a Clustering-based Optimized HEED protocols for WSNs using bacterial foraging optimization and fuzzy logic system [4], and designing of energy-efficient stable clustering protocols based on BFOA for WSNs [3] are implemented. These protocols also focus on the partial mobility of nodes but the work is remarkable as far as stable clustering is concerned.

The proposed protocol SERA also applied the random way point mobility model to all the sensor nodes which remove the partial mobility in the network and then we discussed the novel clustering method which routes the data to BS and provides the proper re-rotation of CHs at the proper time which improves the QoS, connectivity, and reliability issue of the network. The protocol is more dynamic than that of the existing protocols and can withstand in a dynamically changing environment of WSNs.

The organization of the paper is as follows: Sect. 2 provides the details of SERA protocol. Section 3 presents the simulation results. Finally, the conclusion is concluded in Sect. 4.

2 SERA Protocol

Since in SERA protocol, all the sensor nodes are mobile and have limited battery power, there should be a dynamic clustering approach which selects the proper CHs and provides the proper route toward BS when the position of mobile nodes changes with time. The proposed protocol utilizes the concept of normalized node distance $d_{norm}^{ij}(t)$ and energy $E_{norm}^{i}(t)$ in proportion manner that can be evaluated by

$$d_{norm}^{ij}(t) = \frac{d_i(t)}{d_{max}^j} \tag{1}$$

$$E_{norm}^i(t) = \frac{E_i(t)}{E_{max}(t)} \tag{2}$$

where $d_{norm}^{ij}(t)$ denotes the normalized distance for a node i to center/BS j at time t, $d_i(t)$ is the node i distance from BS j, $d_{max}^j = 50\sqrt{2}$ m is the maximum possible distance from the BS j which depends on network filed, $E_{norm}^i(t)$ represents the normalized energy at time t, $E_{max}(t)$ is the maximum residual energy present in the network at time t, and $E_i(t)$ denotes the residual energy of sensor node i.

The Eqs. (1), (2) represents the cost function in proportion manner which can be evaluated as

$$Cost - f_i(t) = \frac{E_{norm}^i(t)}{d_{norm}^{ij}(t)} \tag{3}$$

The proposed cost function as shown in Eq. (3) selects the proper CHs when the location and energy of sensor nodes change. The cost function selects the CHs until optimum number of CHs k_{opt} are selected which can be identified from Eq. (4) as given by [5, 11]

$$k_{opt} = \begin{cases} \sqrt{\frac{n}{2\pi}} \cdot \frac{M}{d_{avg-toBS}} = \sqrt{\frac{n}{2*\pi}} \cdot \frac{2}{0.765}, & for \ d \leq d_o \\ \sqrt{\frac{n}{2\pi}} \cdot \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}} \cdot \frac{M}{d_{avg-toBS}^2}, & for \ d \geq d_o \end{cases} \tag{4}$$

where n denotes the total number of nodes deployed in $M \times M$ region, ϵ_{fs} or ϵ_{mp} is the power amplifier energy consumption, and $d_{avg-toBS}$ represents the average distance from CHs to BS which can be evaluated as [11]

$$d_{avg-toBS} = 0.765\frac{M}{2}$$

The proposed cost function tries to minimize the energy consumption in the network by normalizing the energy and distance factors in proportional manner so that there would be sync energy consumption in the network. Thereby, it maintains the CHs distribution effect in the network and make the lifetime of network longer due to $d_{norm}^{ij}(t)$ factor.

For network performance analysis, we have adopted the radio energy model as provided by Heinzelman et al. [5]. As per [5], the energy expended by the radio over a distance d for transmitting and/or receiving an l-bit message can be evaluated as

$$E_{Tx}(t) = \begin{cases} l.E_{elec} + l.\epsilon_{fs}.d^2 & d < d_o \\ l.E_{elec} + l.\epsilon_{mp}.d^4 & d \geq d_o \end{cases} \tag{5}$$

where $E_{Tx}(t)$ is the transceiver energy dissipation at time t, d denotes the distance between the sender and receiver, E_{elec} denotes the transmitter/receiver circuit energy dissipation per bit, $\epsilon_{fs}.d^2$ or $\epsilon_{mp}.d^4$ is the amplifier energy that depends on acceptable bit error rate, the distance to the receiver and the $d_o = \sqrt{\frac{\epsilon_{fs}}{\epsilon_{mp}}}$ denotes the threshold distance.

Since the network is heterogeneous, the n sensor nodes occupying with different energies in between $[E_o\ 2E_o]$ are dispersed uniformly within an $M \times M$ region initially. Thus, the total initial energy E_{total} of the HWSN can be obtained by

$$E_{total} = \sum_{i=1}^{n} E_o(1 + \beta_i) \qquad (6)$$

where E_o is the initial energy in the nodes, β is the incremental energy factor.
In addition, connectivity with the mobile nodes is also a major issue which is introduce in the proposed protocol to figure out the network connectivity performance. Since the mobile nodes keep changing its direction and location with time, beyond the distance d_o such nodes would lose the link with which they are connected and lose the information. The transmission of data is only for the distance $d < d_o$. Moreover, the proposed protocol also maintain the re-rotation of CHs with time t with the help of *sub-epoch* as shown in Eq. (7). The nodes once selected as CHs will not take part in the CH selection process untill *sub-epoch* Eq. (7)is satisfied. Equation (7) also helps in sync energy consumption of sensor nodes during the data processing. It increases the network lifetime, minimizes the energy consumption, and increases the network throughput. Consequently, it improves the QoS with reliable connectivity in the dynamic network.

$$sub\text{-}epoch_i(t) = \begin{cases} mod\left(r, round\left(\frac{1}{Cost-f_i(t)}\right)\right) = 0, \ i \in G_c \\ Otherwise, \ i \notin G_c \end{cases} \qquad (7)$$

where G_c is the group of nodes that take part in the CHs selection process at time t.

In the SERA protocol, with prior knowledge of deployed sensor nodes energies and their locations, BS estimates the $Cost - f_i(t)$ and relatively selects CHs as per higher values of the $Cost - f_i(t)$ until k_{opt} nodes are selected. This process is followed in every round until the end of the network lifetime.

3   Simulation Results

This section describes the performance of the SERA protocol using DHWSN. The parameters used in the proposed protocol are tabulated in Table 1 and then the network performance analysis is done by using MATLAB. We consider a network in which

Table 1 Simulation parameters

Symbol	Description	Value
E_{elec}	Transmitter/receiver electronics	5 nJ/bit/m^2
E_o	Sensor node energy	0.1 J
n	Nodes	100
β	Incremental energy factor	1
l	Packet size	500 bytes
ϵ_{fs}	Transmit amplifier ($d < d_o$)	10 pJ/bit/m^2
ϵ_{mp}	Transmit amplifier ($d \geq d_o$)	0.0013 pJ/bit/m^4
E_{DA}	Data aggregation energy	5 nJ/bit/message
Time		20 s
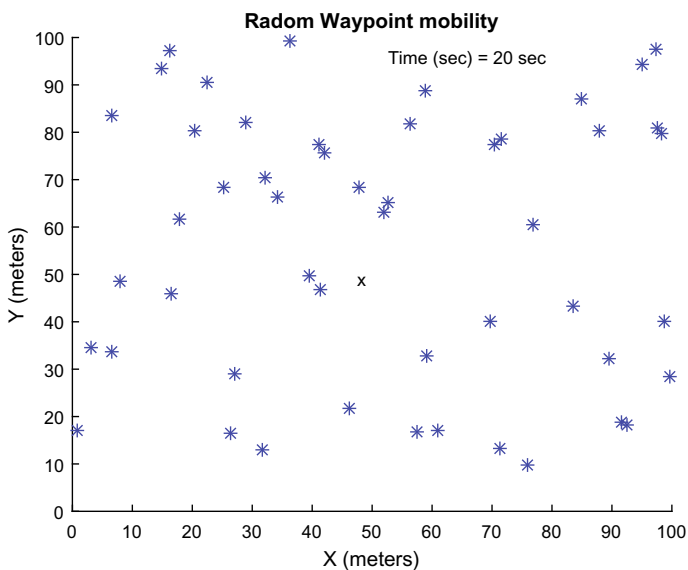
Fig. 1 Snapshot of mobile nodes at 20 s in 100 m × 100 m network field

$n = 100$ mobile sensor nodes are distributed randomly in 100 m × 100 m field at time 0 s with BS located at the center of the sensing region as shown in Fig. 1.

From Eq. (6), the simulation results are presented for $\beta = [0.1\ 1]$ which provides the energy heterogeneity to the nodes. In the SERA protocol, the CHs selection is independent of p_{opt} and $p_i(r)$ as used by [5]. The presented experiment model is simulated for dynamically changing mobile nodes. However, [1, 10, 12] protocols are simulated for partial mobile nodes, i.e., most of the nodes are stationary and fewer are mobile. Therefore, the comparative analysis of the proposed SERA protocol with [1, 10, 12] protocols is not possible. Alive nodes versus rounds and network energy surviving versus rounds for DHWSNs are plotted in Fig. 2 and Fig. 3, respectively.
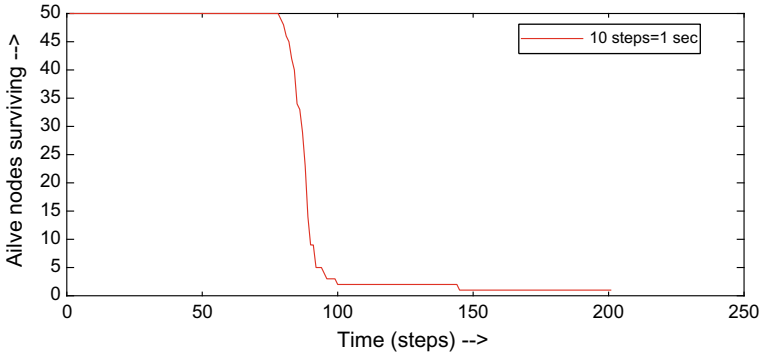
Fig. 2 Alive nodes versus rounds
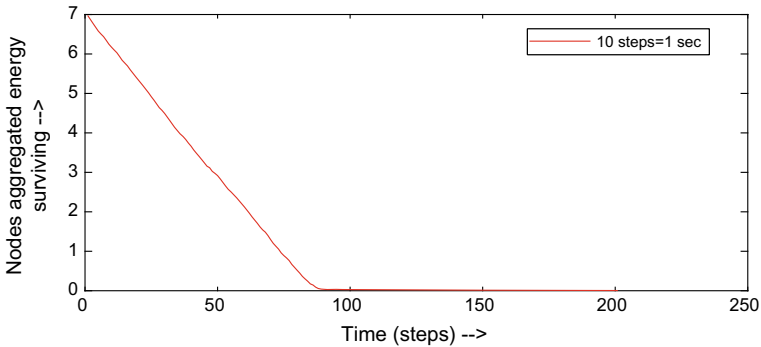
Fig. 3 Nodes energy surviving versus rounds

Figure 2 indicates that the die-up of alive nodes starts sharply with time which maintains the network coverage as long as possible and there is a sharp cutoff in alive nodes after 90% die-up of nodes because of sync energy consumption as plotted in Fig. 3. Figures 4, 5, and 6 represent the successful packet received at BS, CHs, and cumulative network energy surviving with variation in incremental energy factor β. Figure 7 represents the individual node energy status at different time which implies that the energy consumption of nodes is almost in sync. Figures 8 and 9 represent the packet received at BS and CHs respectively which is almost in an equal pattern, i.e., the proposed protocol also maintains the equal load distribution at CHs and BS of the network. Figure 10 describes the stack flow of nodes surviving from a stability period to 100% die-up of nodes. It indicates that with increase in β, the lifetime of network increases because of increase in network energy. Also, each stack (from bottom to top) at particular β represents the stability period, 10, 20 , 30, 40, 50, 60, 70, 80, 90, and 100% die-up of nodes. It is clear from the discussed figures that the overall performance of the network in terms of QoS, connectivity, reliability, and stability period can withstand within the constrained region for a completely mobile network.
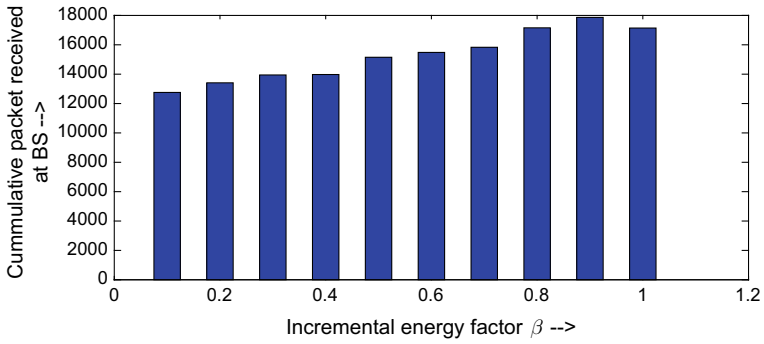
Fig. 4 Cumulative packet received at BS with variation in β
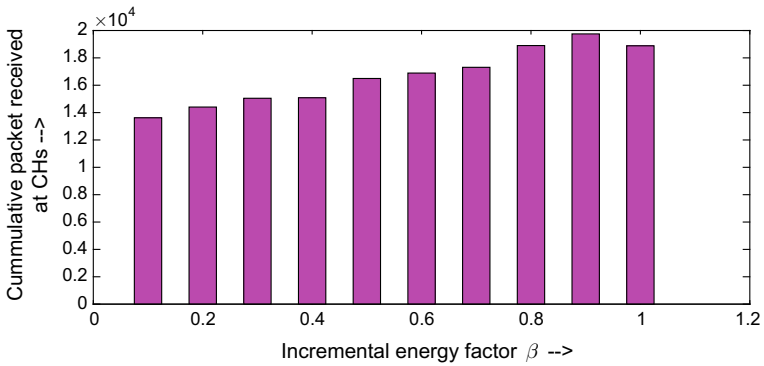
Fig. 5 Cumulative packet received at CHs with variation in β
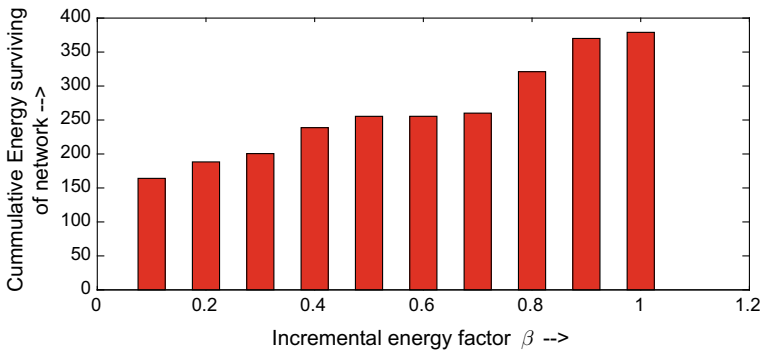
Fig. 6 Cumulative network energy surviving with variation in β
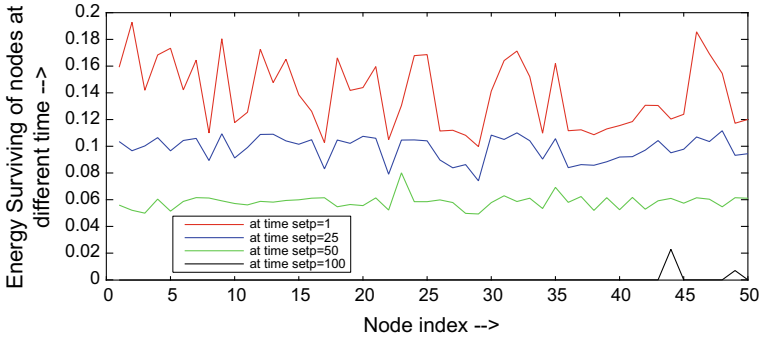
Fig. 7 Energy surviving of nodes at different time with node index
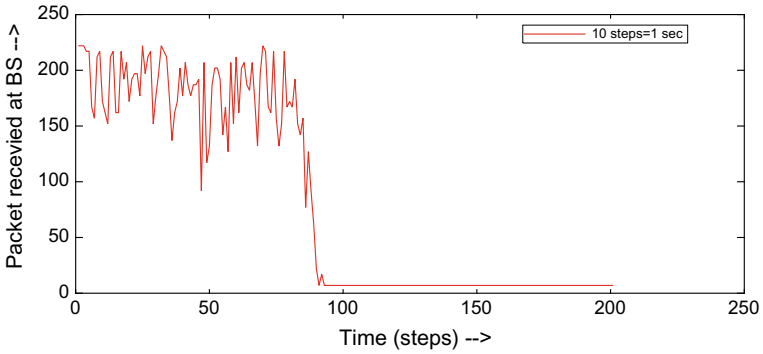
Fig. 8 Packet received at BS with time
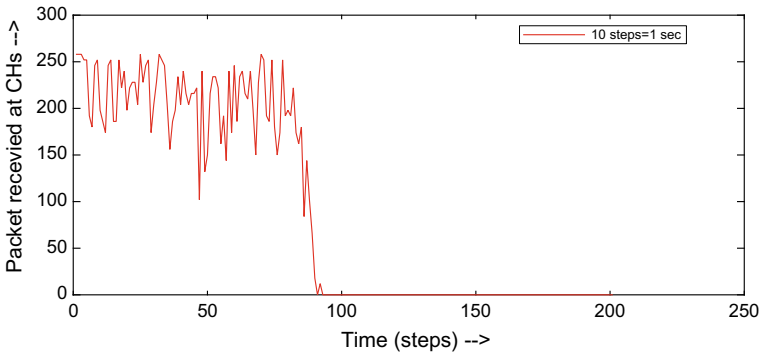
Fig. 9 Packet received at CHs with time
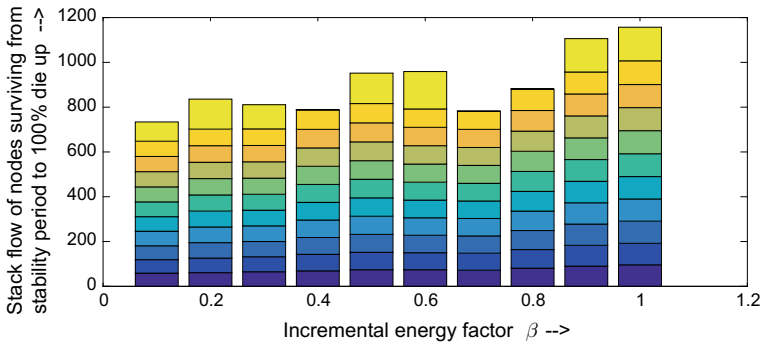
Fig. 10 Nodes surviving with variation in β

4 Conclusion

In this paper, the SERA (Stable and Energy-Efficient) protocol is proposed a proportional approach for analyzing the network performance of the dynamic Heterogeneous WSNs (MHWSNs). It provides a generalized cost function for DHWSNs. The SEAR protocol focuses on the normalized distance and energy factor in the network for finding the optimum number of cluster heads (CHs) and also provides proper re-rotation of such CHs at the proper time with the help of *sub-epoch*. It maintains the CHs distribution in the network which provides sync energy consumption in the network and makes the network lifetime longer. Simulation results justify that the SERA protocol is reliable and efficient for DHWSNs. The SERA protocol maintains the QoS in the network by increasing the throughput, network lifetime and decreasing the energy consumption of sensor nodes. Thus, the overall performance of the network is well-balanced in the proposed SERA protocol. In future, the proposed protocol also focuses on the large-scale dynamically changing WSNs with completely mobile sensor nodes and then analyzes the overall network performance on different environmental factors.

References

1. Abdulwahid, H., Dai, B., Huang, B., Chen, Z.: Scheduled-links multicast routing protocol in manets. J. Netw. Comput. Appl. **63**, 56–67 (2016)
2. Afsar, M.M., Tayarani-N, M.H.: Clustering in sensor networks: a literature survey. J. Netw. Comput. Appl. **46**, 198–226 (2014)
3. Gupta, P., Sharma, K, A.: Designing of energy efficient stable clustering protocols based on BFOA for WSNS. J. Ambient Intell. Humanized Comput. 1–20 (2018)
4. Gupta, P., Sharma, A.K.: Clustering-based optimized heed protocols for WSNS using bacterial foraging optimization and fuzzy logic system. Soft Comput. 1–20 (2017)
5. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wirel. Commun. **1**(4), 660–670 (2002)

6. Krishna, P., Vaidya, N.H., Chatterjee, M., Pradhan, D.K.: A cluster-based approach for routing in dynamic networks. ACM SIGCOMM Comput. Commun. Rev. **27**(2), 49–64 (1997)

7. McDonald, A.B., Znati, T.F.: Design and performance of a distributed dynamic clustering algorithm for ad-hoc networks. In: Proceedings of 34th Annual Simulation Symposium, 2001, pp. 27–35. IEEE (2001)

8. Pantazis, N.A., Nikolidakis, S.A., Vergados, D.D.: Energy-efficient routing protocols in wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **15**(2), 551–591 (2013)

9. Qing, L., Zhu, Q., Wang, M.: Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. Comput. Commun. **29**(12), 2230–2237 (2006)

10. Royer, E.M., Perkins, C.E.: Ad-hoc on-demand distance vector routing. In: Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, vol. 2, pp. 90–100 (1999)

11. Smaragdakis, G., Matta, I., Bestavros, A.: Sep: a stable election protocol for clustered heterogeneous wireless sensor networks. Boston University Computer Science Department, Technical Report (2004)

12. Vaishampayan, R., Garcia-Luna-Aceves, J.J.: Efficient and robust multicast routing in mobile ad hoc networks. In: 2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems, pp. 304–313. IEEE (2004)

13. Xu, L., Collier, R., O'Hare, G.M.: A survey of clustering techniques in WSNS and consideration of the challenges of applying such to 5G IoT scenarios. IEEE Internet Things J. **4**(5), 1229–1249 (2017)

14. Zhen, H., Li, Y., Zhang, G.J.: Efficient and dynamic clustering scheme for heterogeneous multi-level wireless sensor networks. Acta Automatica Sinica **39**(4), 454–460 (2013)

Comparative Study of Different Routing Protocols for IEEE 802.15.4-Enabled Mobile Sink Wireless Sensor Network

Pallavi Joshi, Ghanshyam Singh and A. S. Raghuvanshi

Abstract Routing protocols play a very significant role in Wireless Sensor Networks (WSN). Routing is responsible for establishing an optimal path between the source and sink node. WSN are self-organized networks of spatially distributed dedicated sensors which collect the data and send it to the base station. IEEE 802.15.4 is an IEEE standard developed for Low Rate-Wireless Personal Area Network (LR-WPAN). Most real-time WSN applications use IEEE 802.15.4 standard for Media Access Control (MAC) and Physical layers in order to enhance the functionality of the nodes. This paper aims to provide a comparative performance evaluation of four chosen routing protocols most suited for WSN applications, which are, Ad-hoc On-demand Distance Vector routing (AODV), Dynamic MANET On-demand routing (DYMO), Bellman–Ford and Optimized Link State Routing (OLSR) protocols. The performance of each protocol is compared using standard Quality of Service metrics (QoS) like average throughput, average end-to-end delay, average energy consumption, etc. The developed scenario is studied for both stationary and mobile sink condition. Simulation results presented show that DYMO performs best in terms of average throughput, average energy consumption, and total packet received at sink.

Keywords Wireless sensor networks · Ad-hoc networks · AODV · DYMO · Bellman–Ford · OLSR · IEEE 802.15.4 · Routing protocols

1 Introduction

Sensor networks are basically the networks consisting of different or same types of nodes. WSNs are very similar to wireless Ad-hoc networks. These nodes are sensor nodes which sense the parameters like pollutants in the air, temperature,

P. Joshi (✉) · G. Singh · A. S. Raghuvanshi
Department of Electronics and Telecommunication, NIT Raipur, Raipur, India
e-mail: pjoshi.phd2017.etc@nitrr.ac.in

G. Singh
e-mail: gsingh.phd2016.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

vibration. etc.; these parameters are transmitted to the base station via a path called Route. The algorithms used to find the paths from the source to sink are called Routing Algorithms. Data-Centric Routing is generally implemented in wireless sensor networks [1]. WSN does not need any infrastructure. According to the need of the application, it can configure itself into any topology [2].

Routing is the term which refers to the technique of finding the best path between the source and destination in very less time and also by utilizing very less energy. Delay and energy consumption are the main constraints in wireless sensor networks. Routing is simply relaying of data, but to do it efficiently and in an optimized manner, we have different routing protocols. These protocols are basically algorithms which handle the task of finding the path very efficiently [3].

Mobility is a profound feature in wireless networks. Mobility may avoid unnecessary delays and errors in wireless sensor networks [4]. Mobility in wireless networks can solve many problems:

- Plenty of applications can be implemented with mobile nodes like military surveillance, cattle and habitat monitoring, etc. [5].
- If the mobile nodes are present in a network, then there is hardly any chance of link breakage; also, it gives relief from traffic congestion [4].
- Since there will be no link break, it will solve the problem of network connectivity with a few nodes moving and the other nodes in their fixed location [6]. The moving nodes act as full-function devices and have the ability to collect data from the stationary nodes.

Comparison study of four protocols is done by simulation in Exata/Cyber software. This will help other researchers to select the appropriate protocol for a particular QoS parameter in which the protocol gives best results.

The paper organizes five sections: Sect. 1 is the Introduction of wireless sensor networks and Ad-hoc wireless network. Section 2 explains ZigBee/IEEE 802.15.4 technology is MAC and PHY protocol which is used in the scenario. Section 3 discusses the routing protocols implemented in the WSN scenario which enable IEEE 802.15.4. Section 4 contributes to the proposed work. Section 5 deals with the simulation and parameters analysis. Section 6 manifests results and discussions and in Sect. 7, the study is concluded.

2   IEEE 802.15.4

It is LR-WPAN which works at low cost, transfers data efficiently for upper layers of the protocol stack, and efficiently uses energy [7]. It is defined in both the Physical as well as the MAC layer of the network. At the Physical layer, it uses the modulation scheme as Offset-QPSK and Radio type as 802.15.4 Radio. The Physical layer has three frequency bands of 868 MHz (1 channel), 915 MHz (10 channels), and 2400 MHz (16 channels) [8].

For the MAC layer, the IEEE 802.15.4 standard employs two types of devices that are FFD (full-function device) which act as a network coordinator. FFD can synchronize, establish communication, and join in the network by sending beacons. The RFDs (reduced function devices) can sense and detect the parameters and are called end devices. RFD can interact with one FFD. The PAN coordinator sends a beacon which has a super frame structure as shown in Fig. 1. The frame beacon has two periods—Active and Inactive period (sleep period). The Active period has 16 slots out of which 9 slots are allotted for contention access period and 7 slots are for contention-free period. The whole beacon period (consisting of Active and Inactive periods) is called Beacon Interval and this structure regulates the sending of beacon by a PAN coordinator.

3   Routing Protocols

Routing protocols are used with respect to the nature and topology of the sensor networks, so they can be categorized as

- Static and Dynamic
- Reactive (On-demand), proactive (Table-driven), and hybrid
- Link state and Distance vector

The reactive routing protocols like AODV, DYMO are on-demand routing protocols. For these, the routes are found only when the source node desires to transmit the message to the sink node. Reactive protocols make routes on demand. Proactive protocols like OLSR are Table-Driven protocols make use of routing tables. The hybrid protocols like ZRP characterizes both reactive and proactive protocols [3, 9].

AODV—AODV (Ad-hoc On-demand Distance Vector) is designed specifically for Ad-hoc networks for establishing routes to sink nodes when the source demands it, hence the name "On-Demand". It finds routes only when a source node requests thereby avoiding extra traffic [10]. It is suitable for both multicast and uni-cast routing. As long as the source node needs the routes, they are maintained otherwise discarded. Initially, the source node which wants to transmit initiates route discovery and transmits route request packet to the neighbors thereby avoiding unnecessary broadcasts [11]. When it reaches the sink, the sink sends a reply message toward a
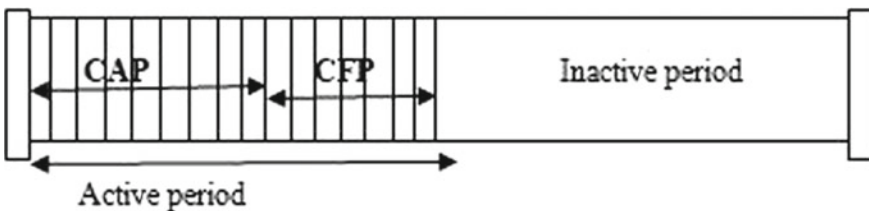
Fig. 1   IEEE 802.15.4 super frame for MAC

reverse path to the source. Error is handled by sending RERR packets. Thus, AODV avoids looping and unnecessary long routes.

DYMO—DYMO (Dynamic MANET On-demand) is a modified version of AODV. Both go well with MAC IEEE 802.15.4 [12, 13]. DYMO is a simple but fast reactive protocol, basically for multi-hop wireless networks. DYMO constitutes two operations—Route discovery and Route maintenance. It repeats the Route discovery process for avoiding congestion in the network [14]. Unless the suitable route path is obtained, it continues to send beacon messages and waits until the route is discovered. This protocol saves memory to a larger extent because only the routing information of the concerned source and destination nodes are maintained by every node [15]. In case of a link breakage, RERR is issued and the route is discarded; again the same process starts and a new route is established.

BELLMAN–FORD—This protocol overestimates the length of the path from the first vertex to other vertices. Iteratively, the same process continues and replaces the new shorter paths by previous overestimated paths thereby getting the optimized paths as a result.

It is similar to Dijkstra's shortest path algorithm. It searches the shortest path between the given source and sink in the graph by means of the shortest branches in the graph. Bellman–Ford is suitable for both weighted and non-weighted graphs. It handles negative weights in the graph and they are discarded in the search of the shortest distance.

OLSR—OLSR (Optimized Link State Routing) is a link-state protocol specifically designed for MANETS with high density of nodes. This protocol has the ability to control random traffic. It goes well with mobile networks and can exchange topology with different nodes in the network. It is a proactive protocol which has some data forwarding and traffic control nodes called MPRs (multipoint relays). These MPRs are efficient in flooding mechanism and can reduce the unnecessary transmissions in the network.

4 Proposed Work

The random mobility model which is popular for measuring the performance of the routing protocol of Ad-hoc networks was proposed by Johnson and Maltz. In our scenario, we are introducing five ground vehicles which possess random waypoint mobility and collect the sensed data from other stationary nodes. The 100 stationary nodes are randomly deployed and some nodes (source nodes) are connected via a CBR link to the base station. We are basically focusing on how reactive and proactive protocols behave for an IEEE 802.15.4-enabled wireless sensor network scenario with large number of nodes. Four protocols are implemented: AODV, DYMO, Bellman–Ford, and OLSR and the QoS metrics are analyzed.

5 Simulation and Parameter Analysis

Exata/Cyber network simulator is used for generating a random scenario of 100 sensor nodes and five unmanned ground vehicles which have random waypoint mobility [16]. These vehicles are acting as temporary base stations for collecting data from other nodes. A destination station is placed in the scenario which is connected through CBR link with 10 nodes acting as source nodes. Figure 2 shows the scenario in Exata/Cyber. Now the scenario is saved and the simulation is started with all four protocols implemented each time. Some 10–15 simulations were done with different SEED values.

SEED is the parameter which regulates the initialization of the location of the nodes. Since the nodes are randomly distributed, for every simulation the location of the node will get slightly changed (Table 1).
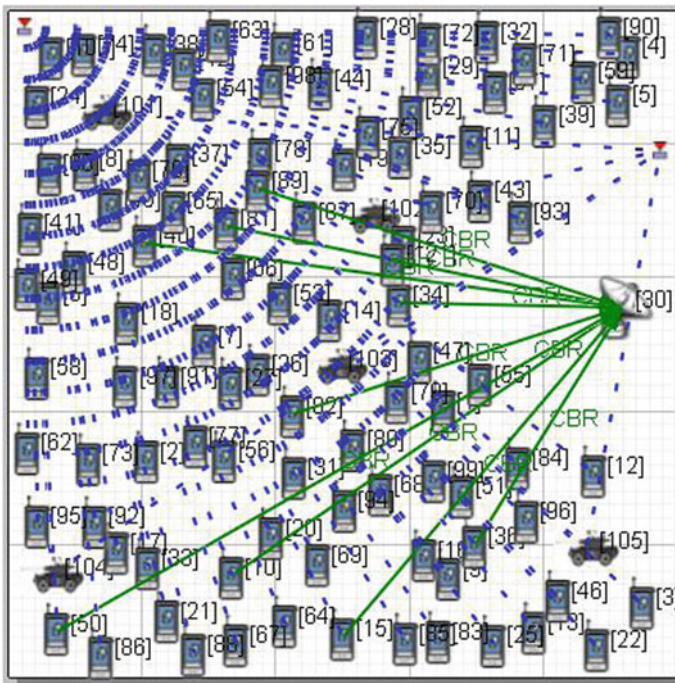
Fig. 2 A model scenario with 100 stationary and 5 mobile nodes

Table 1 Simulation parameters

Parameters	Values
Scenario dimensions (m)	500 m × 500 m
No. of nodes	105
No. of mobile nodes	5
MAC protocol	MAC 802.15.4
Routing protocol	AODV, DYMO, Bellman–Ford, OLSR
Mobility model	Random waypoint
Simulation time (s)	800
Bandwidth (Mbps)	2
Traffic type	CBR
Data packet size (bytes)	512

6 Results and Discussions

The scenario implements the following technologies in Exata/Cyber

- The wireless standard IEEE 802.15.4 for MAC and Physical layers.
- The random waypoint mobility.
- AODV, DYMO, Bellman–Ford, and OLSR routing protocols for the same scenario.

The same scenario with different parameters is run in Exata/Cyber and the results of different performance metrics which are obtained in the analyzer of Exata/Cyber have been observed for two different protocols. The paper compares QoS metrics for all four protocols [17]. The results are as follows.

Throughput is the division of the total data collected by sink node from the source by the time taken by the sink to get the final packet. Its unit is bit/sec. The higher the mobility of the nodes, the more will be the throughput and is high in case of DYMO as shown in the graph. Figure 3 shows the graph for average throughput.

Due to some reasons like queuing and various routing paths, a packet may experience delay in reaching to the sink node. The average end-to-end delay is the total time interval taken by a data packet that started from the source to be received entirely by the sink. Its unit is seconds. For mobile nodes, Bellman–Ford gives negligible end-to-end delay. For reactive protocols like AODV and DYMO, the delay is approximately the same. Figure 4 demonstrates the graph for the average end-to-end delay.

The jitter is the term used for the variations of time in between the incoming of packets. Due to drift in time, congestion, and change in route paths in the network, this factor comes into the picture. The lower the value of jitter, the better it is for the network. For mobility model, proactive protocols behave excellently for this QoS metric whereas DYMO shows comparatively less jitter in a dense network scenario. Figure 5 shows the graph for Average Jitter.

Fig. 3 Average throughput at the sink node
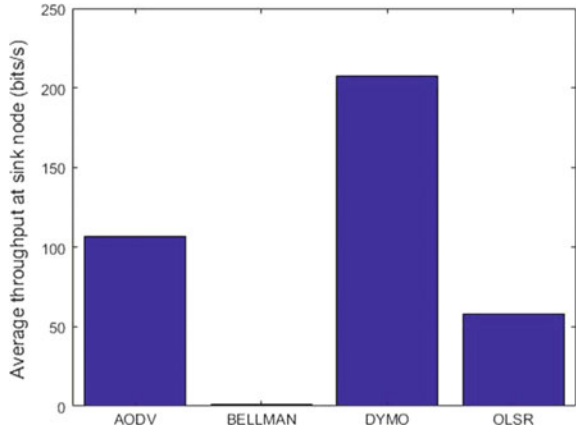
Fig. 4 Average end-to-end delay
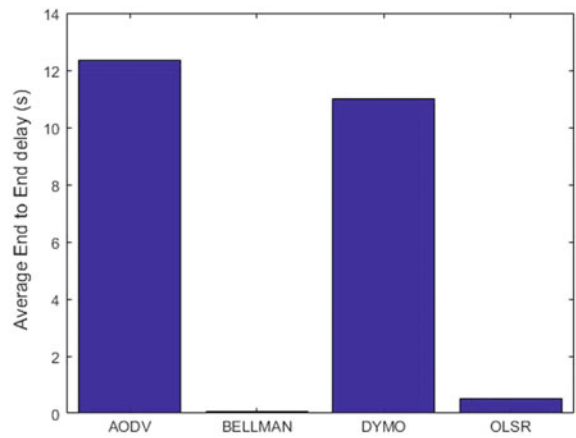
Fig. 5 Average jitter
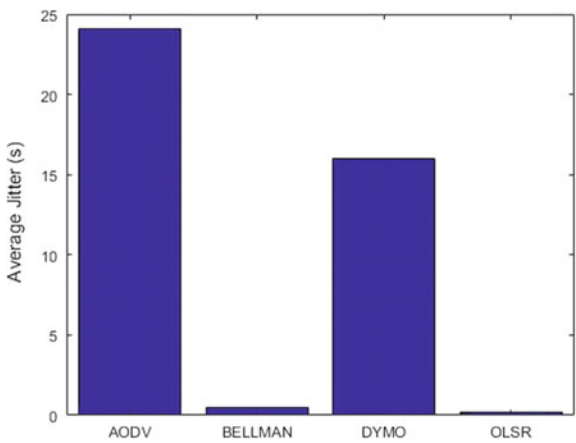
The scenario is equipped with both stationary and mobile nodes. Mobile nodes are prone to more consumption of energy. The main constraint of any Ad-hoc network is the energy consumption by the nodes, which should be less. We can easily observe that DYMO protocol saves much energy than the rest of the protocols. It can be noticed that the consumption of energy increases with the density of nodes in the WSN. Figure 6 is the graph for average energy consumption.

The packets sent from all sources may get dropped in the way due to delay, traffic, etc. The highest number of packets received at the sink node is for the scenario implementing DYMO. Figure 7 shows the graph for total packets received at the sink node (Table 2).

DYMO is the fast routing protocol and is able to search unicast routes thereby serving best for on-demand and dynamic topology-based scenarios. It has to maintain little routing state by only taking care of the information of active sources and

Fig. 6 Average energy consumed
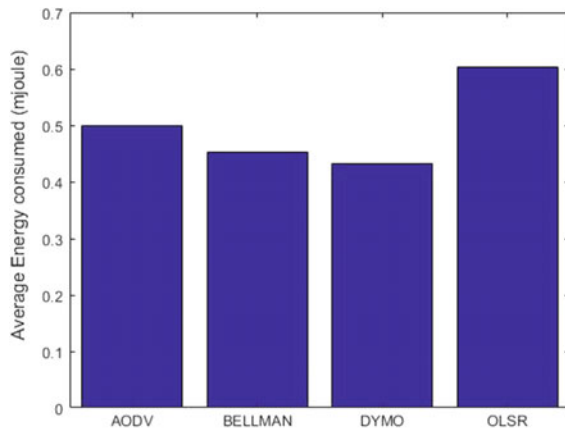
Fig. 7 Total packets received
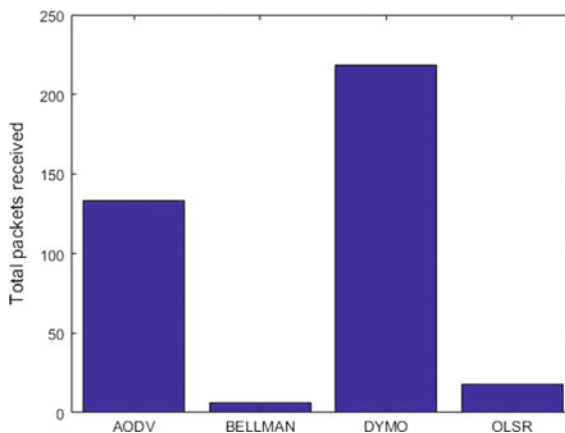
Table 2 For the same scenario with 105 nodes, the comparison for four different protocols for different performance parameters is done

S No.	Utilized protocols	Avg. throughput (bits/sec)	Avg. end-to-end delay (s)	Avg. jitter (s)	Avg. energy consumed (mJ)	Total packets received
1	AODV	112	12.14	24.87	0.498	130
2	Bellman–Ford	0.231	0.02737	0.135	0.452	0.125
3	DYMO	222	11.39	16.31	0.432	220
4	OLSR	57	0.11124	0.0381	0.604	21.34

destinations making it resource and memory constrained. This is the reason why DYMO outperforms the AODV, Bellman–Ford, and OLSR protocols in the case of average throughput, energy consumption and total packets received.

7 Conclusion

The paper demonstrates the example for analyzing IEEE 802.15.4 network performance for reactive and proactive protocols. Delay and energy consumption are the main constraints in wireless sensor networks. After analyzing the results, it is observed that for a scenario of mobile nodes, the average throughput, average energy consumption, and total packets received is excellent for DYMO but at the cost of average jitter and average end-to-end delay [18]. It can be concluded that Bellman–Ford and OLSR protocols gives very less jitter and delay but their throughput is very less which is not idle for the wireless network [19].

The energy consumed in the scenario when DYMO was implemented is much less as compared to that of other protocols [20]. So it can be concluded that DYMO is the leading reactive protocol when IEEE 802.15.4-enabled sensor networks are employed with high density of nodes.

References

1. Akyildiz, I.F., Su, W., Sankasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. 393–422 (2002)
2. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. **52**, 2229–2330 (2008). Elsevier Publications
3. Basavaraju, T.G., Sarkar, S.K.: Ad Hoc Mobile Wireless Networks: Principles, Protocols and Applications. CRC Press (2016)
4. Al-Jemeli, M., Hussin, F.A.: An energy efficient cross-layer network operation model for IEEE 802.15.4-based mobile wireless sensor networks. IEEE Sensor J. **15**(2) (Feb. 2015)
5. Wang, X., Lin, X., Wang, Q., Luan, W.: Mobility increases the connectivity of wireless networks. IEEE/ACM Trans. Netw. **21**(2), 440–454 (Apr. 2013)

6. Cattani, M., Guna, S., Picco, G.P.: Group monitoring in mobile wireless sensor networks. In: Proceedings of the International Conference on Distributed Computer Sensor Systems Workshops (DCOSS) (pp. 1–8) (Jun. 2011)
7. Zen, K., Habibi, D., Rassau, A., Ahmad, I.: Performance evaluation of IEEE 802.15.4 for mobile sensor networks. In: Proceedings of the 5th IFIP International Conference on Wireless and Optical Communication Networks (WOCN) (pp. 1–5) (2008)
8. Huang, Y.K., Pang, A.C., Hung, H.N.: A comprehensive analysis of low power operation of beacon-enabled IEEE802.15.4 wireless networks. IEEE Trans. Wireless Commun. **8**(11), 5601–5611 (Nov. 2009)
9. Broch, J.,Jhonson, D.,Maltz, D.: The dynamic source routing protocol for mobile ad-hoc networks for IPv4. IETF RFC 4728 (Feb. 2007)
10. Perkins, C., Royer, E., Das, S.: Ad hoc on demand distance vector (AODV) routing. IETF RFC No. 3561 (July 2003)
11. Perkins, C.E, Royer, E.M.: Ad-hoc on-demand distance vector routing. In: Proceeding of 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMSA 1999) (pp. 90–100) (Feb. 1999)
12. ZigBee Specification Version 2007, ZigBee Document [Online]. http://www.zigbee.org 0534747r17 (Jan. 2008)
13. ZigBee.: Wireless Technology for Low Power sensor Networks [Online]. http://eetimes.com/design/communicationsdesign/4017853/ZigBee. Wireless Technology for Low Power Sensor Networks. Accessed 2 May (2013)
14. Chakeres, I., Perkins, C.: Dynamic MANET on-demand (DYMO) routing. IETF Internet Draft, V.19 (Mar. 2010)
15. Raghuvanshi, A.S., Tiwari, S.: DYMO as routing protocol for IEEE-802.15. 4 enabled wireless sensor networks. In: 2010 Sixth International Conference on Wireless Communication and Sensor Networks (WCSN) (pp. 1–6) (Feb. 2011)
16. Exata/cyber User's Guide [Online]. http://www.scalablenetworks.com/
17. Ahmed, B., Ben Islam, M.K., Rahman, J.: Simulation, analysis and performance comparison among different routing protocols for wireless sensor network. Comput. Eng. Intell. Syst. **2**(4), 61–71 (2011). ISSN 2222-1719 (Paper) ISSN 2222-2863
18. Malany, A.B., Dhulipala, V.S., Chandrasekaran, R.M.: Through put and delay comparison of MANET routing protocols. Int. J. Open Problems Compt. Math **2**(3), 461–468 (2009)
19. Kum, D.W., Park, J.S., Cho, Y.Z., Cheon, B.Y.: Performance Evaluation of AODV and DYMO Routing protocols in MANET CCNC. IEEE (2010)
20. Spaho, E., Mino, G., Barolli, L., Xhafa, F.: Goodput and PDR analysis of AODV, OLSR and DYMO protocols for vehicular networks using CAVENET. Int. J. Grid Util. Comput. **2**(2), 130–138 (2011)

Cooperative Communications Framework for Industrial Applications

Mudit Agrawal, Nandkishor Joshi and Neetesh Purohit

Abstract The communication network of any industry plays very crucial role in its efficient operations and hence there is an increasing demand for building a dedicated wireless network which should be used within the industry by various entities. In this work, a cooperative communication-based unified framework has been developed that allows the spectrum sharing among the employees, machines, operators, managers, etc. Using the concepts of cognitive and cooperative communications with relays, all communications are carried over the same spectrum in very efficient way. Simulation is carried out by deploying both stationary as well as mobile nodes and the overall performance is analyzed in terms of the outage characteristics and channel capacities of both primary (fixed nodes) and secondary (mobile nodes) communications. A protocol stack with three layers has been proposed at the end for supporting the prototype development.

Keywords Cognitive radio (CR) · Cooperative communication (CC) · DF relaying protocol · Rayleigh fading channels · Outage probability · Channel capacity

1 Introduction

Industrial areas rely on human interaction and large-scale wired infrastructures for communication between various devices and sectors. However, there are several limitations like terrain, area of coverage, etc., associated with landline communications. Using the concept of maximum frequency reuse and the capabilities of small evolved NodeBs (SeNB) as described in [1], a pervasive wireless infrastructure can be developed to cater to both wireless sensor network (WSN) and human communication

M. Agrawal · N. Joshi (✉) · N. Purohit
IIIT-Allahabad, Allahabad, India
e-mail: nandkishor@iiita.ac.in

M. Agrawal
e-mail: muditgrwl32@gmail.com

N. Purohit
e-mail: np@iiita.ac.in

171

traffic. Since we propose dealing with both machine to machine (M2M) and human to human (H2H) communications over long-term evolution (LTE), the various traffic issues attributed to varying loads of uplink and downlink, periodic and bursty nature and mobility of the nodes was studied as in [2], while considering cooperative communications for improved performance, efficiency and simplified deployment (as compared to MIMO systems) as in [3].

The essence of this paper is how primary and secondary flows will take place concurrently under the framework of cooperative cognitive communications. In this paper, the author has studied the performance of the system, expressing the perspective of the underlay in the context that knowledge of interactive channels is important for implementing underlay theory in hardware [5]. The concept of decode-and-forward (DF) relaying in case of mobile terminals and no direct line-of-sight is studied as in [6] while [7] studies the behaviour of outage of various such protocols used to tackle multipath fading. Formulae for estimating the outage, the capacity of the channel employed by the secondary user and the scheme for controlling power of the secondary users are also inspired by this literature. The papers [8–10] discuss the marvels of Internet of Things (IoT), the various technologies required to implement the solution as a whole (e.g. cloud computing, M2M and WSN) and various present-day cognitive IoT frameworks studied across various literature but also highlight several real-world scenarios.

The objective of this paper is to design and develop a communications scheme which caters to both the primary (M2M) and secondary (H2H/M2H) communications while maximizing the use of the available resources under the constraint of the underlay mode of operation of cognitive radio. It involves designing a layout of nodes with proper channel modelling and power allocation for CR operation, supporting cooperative relaying protocols. The main purpose of performing these tasks is to study the metrics of performances of such a network and understand the feasibility of deployment of such a set-up and view it with the perspective of an IoT inspired solution. The wonderful thing about IoT is the sheer amount of flexibility it provides in terms of making a solution application-specific.

Section 2 explains the proposed framework; simulation results and discussion are presented in Sect. 3 and Sect. 4 presents the conclusion and future scope of this paper.

2   The Proposed Framework

As shown in Fig. 1, the machines are considered as the primary users and the human/vehicle are considered as secondary users. Both primary and secondary communications follow DF relaying protocol in which nodes of similar nature act as relays between the source and destination. The secondary communication runs in parallel with primary communication. It is communicated by correcting the power used by the primary relay without obstructing the total power available in the system. Now, the system can be called tolerant for a particular interference level. Thus, the power of the secondary transmitter is calculated with respect to this constraint.
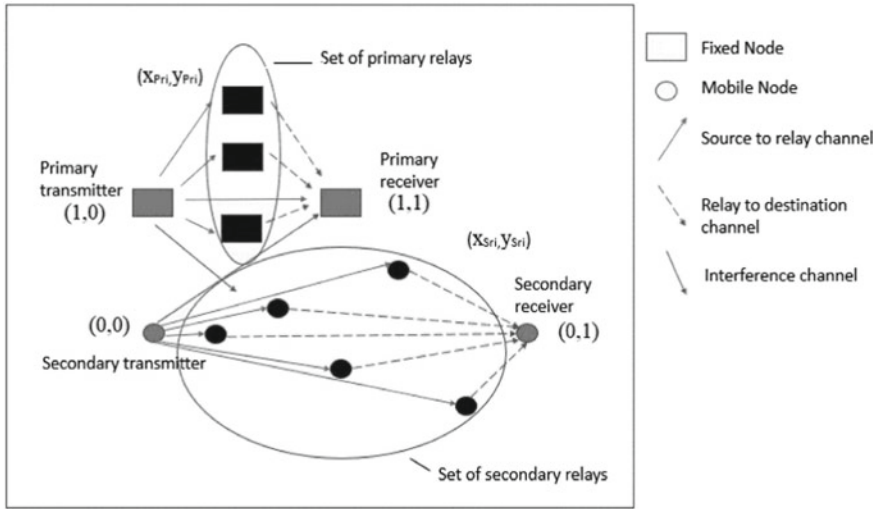
Fig. 1 The schematic diagram of the communication system

Power available for secondary relay is obtained from three powers, primary transmitter power, primary relay and secondary transmitter. In this paper, entities of the primary network are fixed in nature. Hence the fading channels between the source and relays and the relays and destinations are modelled according to the Rayleigh fading distribution. Since, the entities of the secondary network are movable, the fading channels are modelled according to the double Rayleigh fading distribution.

The total power available for the system is denoted by-

$$P_{Total} = P_{tr} + P_r + P_{Str} + P_{Sr} \tag{1}$$

where P_{tr} is the power available at the primary transmitter, P_r is the power available at the primary relay, P_{Str} is the power available at the secondary transmitter and P_{Sr} is the power available at the secondary relay.

Figures 2 and 3 represent the primary and secondary relaying infrastructures, respectively. P_t and P_d are the primary transmitter and receiver, respectively. P_{Ri} represents the relays in the primary network with M relays. $d_{Pt,PRi}$ and $h_{Pt;PRi}$ depict the distance and channel coefficients (modelled by flat Rayleigh fading distribution) between the primary transmitter and the ith relay and $i = 0, 1, 2,…,$ M. All units in the secondary network are mobile and there is no direct line-of-sight for communication to the primary receiver, secondary transmitter and secondary receiver.

The channel coefficients between the secondary transmitter and secondary relay and those between the relay and the destination, i.e. $h_{St,SRj}$ and $h_{SRj,Sd}$, respectively are modelled by the double Rayleigh fading distribution. $h_{Sj,P}$ and $h_{P,SRj}$ are the interference channels between the primary and secondary networks and are modelled as flat Rayleigh fading channels due to stationary nature of $P.d_{St;SRj}$ and $d_{SRj;Sd}$ are the
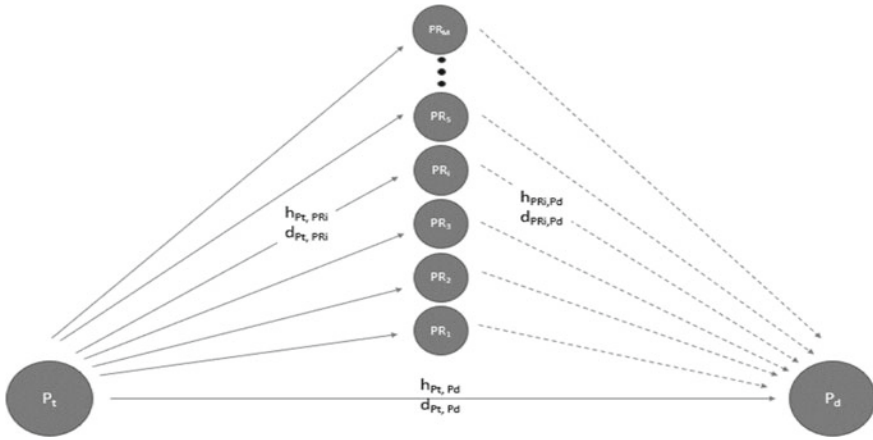
Fig. 2 Primary relay network model
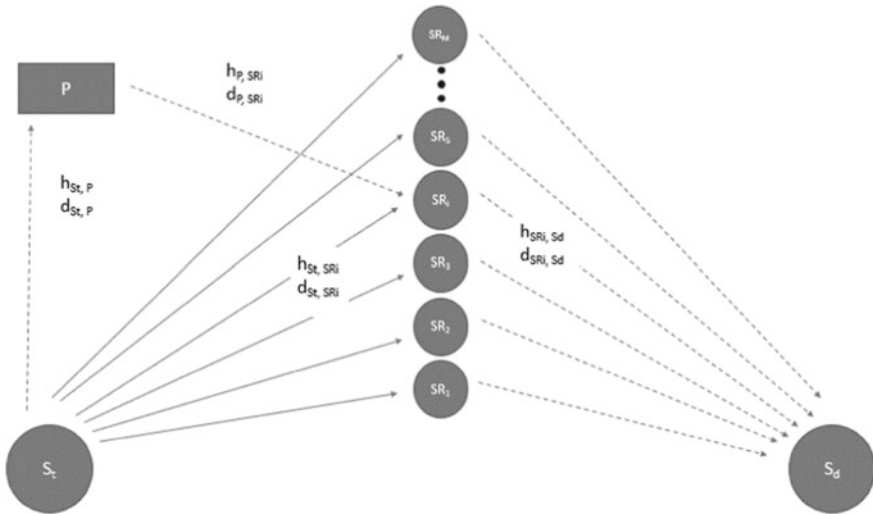
Fig. 3 Secondary relay network model

distances between the transmitter and relay and relay and destination, respectively, while $d_{St,P}$ and $d_{P,SRj}$ are the distances between the primary transmitter and the secondary transmitter and secondary relay, respectively. If the system has m relays then $j = 0, 1, 2, \ldots, m$.

In the first hop, the transmitter transmits messages to the relay and in the second hop, after the partial relay selection by the DF protocol, the message is sent to the destination. Outage ($P_{out,primary}$) is calculated when the received signal to noise ratio at the destination falls below a predefined threshold. Channel capacity (C_P) is estimated

by the formula [4].

$$C_p = B \times \log_2\left(1 + \Upsilon_{e2e, primary}\right) \tag{2}$$

where B is the bandwidth and $\Upsilon_{e2e;primary}$ is the end-to-end SNR for primary communication and is determined by the following equation[5]:

$$\Upsilon_{e2e, primary} = \min\left(\Upsilon_{pt; PRi}; \Upsilon_{PRi; Pd}\right) \tag{3}$$

$\Upsilon_{Pt;PRi}$ and $\Upsilon_{PRi;Pd}$ are the instantaneous SNR at the ith relay and Pd, respectively [5].

$$P_{out; Primary} = P_r[\Upsilon_{e2e, primary} < \Upsilon_{th}] \tag{4}$$

where Υ_{th} is the threshold of SNR for outage of primary communication.

If I_{th} is the interference tolerable at the primary communications, then the power of the secondary transmitter can be modelled as

$$P_{st} = I_{th} / \left|h_{Pr;St}\right|^2 \tag{5}$$

where $h_{Pr;St}$ is the channel coefficient between the primary receiver and the secondary transmitter. Apart from this interference power constraint, the SU transmitter must also transmit at a powerless than its peak power, so as to maintain the peak power constraint as well. The power for the secondary relay is the remaining power available in the system.

Now, DF relaying is employed through partial relay selection to determine the end-to-end SNR for the secondary flow [5].

$$\Upsilon_{e2e, primary} = \min\left(\Upsilon_{pt; PRi}; \Upsilon_{PRi; Pd}\right) \tag{6}$$

where $\Upsilon_{St;SRj}$ and $\Upsilon_{SRj;Sd}$ are instantaneous signal to noise ratios at the relay and the destination, respectively. Now, outage ($P_{out;Secondary}$) is calculated whenever the end-to-end SNR degrades below the desired level or the interference at the primary receiver (I_{Pr}) exceeds I_{th} [6].

$$I_{Pr} = \left|h_{Pr;St}\right|^2 x P_{st} \tag{7}$$

$$P_{out; Primary} = P_r\left[\Upsilon_{e2e, primary} \langle \Upsilon_{th} \cap I_{Pr} \rangle I_{th}\right] \tag{8}$$

Channel capacity of secondary is calculated by using the end-to-end SNR as well as the interference caused by the primary transmission at the secondary relay (I_{Sr}) [5].

$$I_{Sr} = \left|h_{P;SRj}\right|^2 x P_t \tag{9}$$

$$C_p = \frac{B \times log_2\left(1 + \Upsilon_{e2e,primary}\right)}{I_{Sr} + \sigma^2} \tag{10}$$

where σ^2 is the variance of AWGN.

3   Simulation Results and Discussion

By Monte Carlo simulation of the system using MATLAB, there are able to give graphical results for the channel capacities and outage probabilities of both the primary and secondary networks. The parameters for simulation are also provided in this section. Figure 1 also depicts the system model deployed in the software. The primary relays (three nodes) are assumed to be at equal distances (0.4, 0.6, 0.8 units) from the receiver and transmitter. The secondary relays (five nodes) are at a distance of $d_{St;Srj}$ and $(1 - d_{St;Srj})$ from the secondary transmitter and the receiver, respectively. (x_{Pri}, y_{Pri}) and (x_{Srj}, y_{Srj}) are the coordinates of the i_{th} primary and j_{th} secondary relays, respectively. The power of the primary transmitter is taken as a fraction of the total power available, which takes the values 0.5, 0.6, 0.7, 0.8 and the power of the primary relays is fixed as 0.1 of the total power. The transmission powers of the secondary transmitter and relays are calculated according to the Eqs. (1) and (5), respectively. The model is simulated for 10^5 times and the values of outage probabilities for both the systems were calculated according to Eqs. (4) and (8).

Graphs of Figs. 4 and 5 compare the outage probabilities and channel capacities of both primary and secondary communications concurrently. It is seen that the outage probability of the secondary network improves when the total power at the primary transmitter is less as this allows for higher power allocation to the secondary entities. There is a steep decline in the outage probability of the primary network with increased power being distributed to the primary transmitter. This, in turn, increases the power at the disposal of primary relays, allowing for better decoding of incoming messages and thus higher probability of correct retransmission to the destination. Power constraints say that the secondary will be less than 0.01 with current allocation and five relays. This can be further improved when scaled to a larger number of relays. Channel capacity in case of primary is directly proportional to the end-to-end SNR of primary communications while in case of secondary is inversely proportional to the interference caused at the secondary transmitter by the primary communication. Simply put, as power at the primary transmitter's disposal increases, the interference at the secondary nodes adversely affects the channel capacity. Also, when it comes to the primary network, there is a trade-off between the number of relay nodes and the end-to-end SNR as in any partial DF relaying system, and provides better results when scaled to a larger number of nodes.

The respective capacities of the channels of both primary and secondary nature are calculated by the Eqs. (2) and (10). With increase in the effective power available at the disposal of the primary transmitter, the channel capacity of the secondary channel
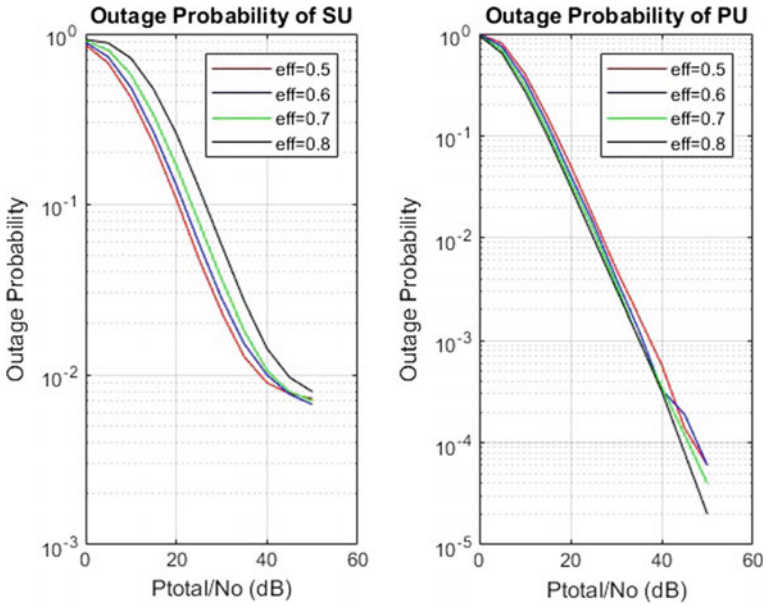
Fig. 4 Comparison of outage probabilities
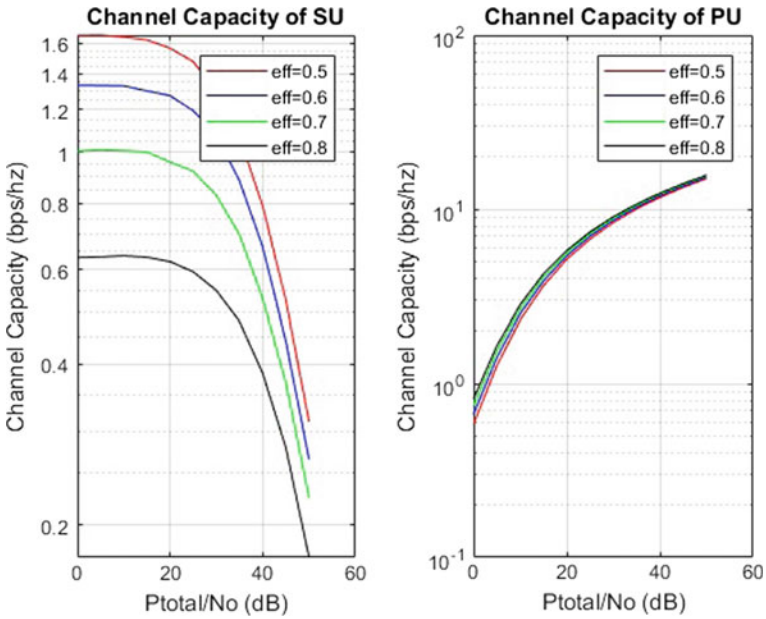
Fig. 5 Comparison of channel capacities

reduces because of the violation of the condition of keeping the interference at the primary receiver in check. Also noteworthy is that in the graph of channel capacity of the primary user, the interpolations overlap at higher values of power. This can be increased by increasing the total number of relays (Table 1).

Since the fundamentals of cognitive and cooperative communication define how the nodes will communicate among themselves, the operation of the solution can be envisioned as a three-layer stack in which the physical and the application layers are joined together by a session layer that manages the interconnections between the referred nodes shown in Fig. 6. It is to take note here that we do not wish to replace any existing communications framework but rather deploy a system that will go hand in hand in the wireless frequency spectrum, in harmony with the existing technologies. Through the application interface available via the application layer protocol, the users can request a countless number of services like data query, visualizations, location services, configuration services, etc. The devices in communicating among each other for the execution of the required service will be informed via the sessions

Table 1 Summary of the simulation parameters

Parameter	Value
Number of simulations	10^5
Ptotal	0:5:50 dB
Pt	eff $\times$ P_{total}
eff	{0.5, 0.6, 0.7, 0.8}
Pr	0.1 $\times$ P_{total}
Ith	10 dBm
B	1
N0	1
Υ_{th}	3
σ^2	-110 dBm
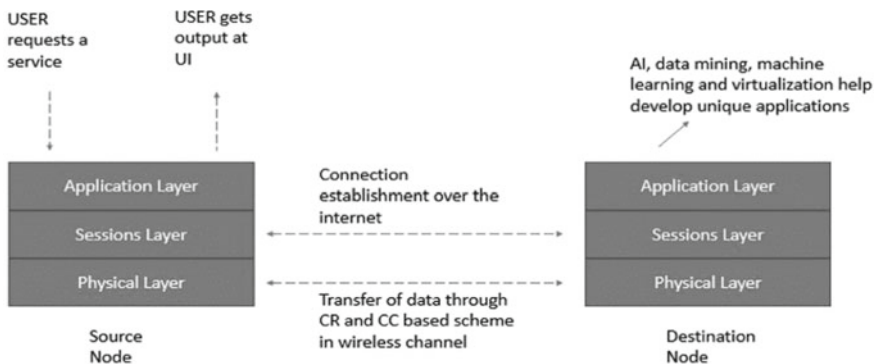
Fig. 6 The proposed three-layer protocol stack

layer about their enquiry and then data is shared according to the schemes of the underlying layer.

4 Conclusion and Future Scope

A cooperative communication-based unified framework has been presented considering all machines to be fixed nodes and human/vehicle as mobile nodes. Furthermore, a protocol stack perspective of the solution was given in which the overall stack was simplified to three layers with the unique physical layer proposals and highly flexible application layer(s) available in the market. Many open issues and challenges are present at each layer, which leave many hurdles (scheduling design, increased overhead in cooperative M2M communications, varying interference between M2M and H2H users, increased latency between the end nodes and channel estimation and interference calculation techniques, etc.) to be crossed before reaching to a prolific and comprehensive solution. More features may be added to this framework towards making it intelligent and self-sustaining ecosystem for industrial applications and development.

References

1. Hassan, F., Roy, A., Saxena, N.: Convergence of WSN and cognitive cellular network using maximum frequency reuse. IET Commun. **11**(5), 664–673 (2017)
2. Raymond, J.W., Olwal, T.O., Kurien, A.M.: Cooperative communications in machine to machine (M2M): solutions, challenges and future work. IEEE Access **6**, 9750–9766 (2018)
3. AbuAli, N.: Enabling technologies of energy efficient cooperative M2M networks: benefits and challenges. In: 38th Annual IEEE Conference on Local Computer Networks—Workshops, Sydney, NSW, 2013, pp. 987–993 (2013)
4. Goldsmith, A.: Wireless Communications. Cambridge University Press, U.K., Cambridge (2004)
5. Kaushik, A., Sharma, S.K., Chatzinotas, S., Ottersten, B., Jondral, F.K.: On the performance analysis of underlay cognitive radio systems: a deployment perspective. IEEE Trans. Cognit. Commun. Netw. **2**(3), 273–287 (2016)
6. Duy, T.T., Alexandropoulos, G.C., Tung, V.T., Son, V.N., Duong, T.Q.: Outage performance of cognitive cooperative networks with relay selection over double-Rayleigh fading channels. IET Commun. **10**(1), 57–64 (2016)
7. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans. Inf. Theory **50**(12), 3062–3080 (2004)
8. Khan, A.A., Rehmani, M.H., Rachedi, A.: Cognitive-radio-based internet of things: applications, architectures, spectrum related functionalities, and future research directions. IEEE Wirel. Commun. **24**(3), 17–25 (2017)
9. Khan, A.A., Rehmani, M.H., Rachedi, A.: When cognitive radio meets the internet of things? In: 2016 International Wireless Communications and Mobile Computing Conference (IWCMC), Paphos, 2016, pp. 469–474 (2016)
10. Shah, M.A., Zhang, S., Maple, C.: Cognitive radio networks for internet of things: applications, challenges and future. In: 2013 19th International Conference on Automation and Computing, London, 2013, pp. 1–6 (2013)

p-Cycles as Their Own *m*-Cycles for Fault Detection and Localization in Elastic Optical Networks

Deo Chandra Jaiswal and Rachna Asthana

Abstract In the new generation, the requirement of transmission of information is an unavoidable need for an exponentially growing traffic demand which is provided by elastic optical networks. A small disruption in these networks interrupts an enormous amount of transferring of data among common as well as prime users. Network survivability has, therefore, become a necessity for uninterrupted network services. In order to achieve the network survivability, preconfigured protection cycle (*p*-cycle) structures due to their fast restoration speed as well as the capacity efficiency are utilized with additional resources to cater the backup protection. *p*-Cycles protect span, with the presumption of having the knowledge of failed span in real time. In case of failed span is not known the fault detection and localization is required before providing protection which is done by monitoring cycles (*m*-cycles). It is investigated that *p*-cycle structures which are used for protection can also be made able to function as *m*-cycles for fault detection and localization. An Integer Linear Programming (ILP) model is developed to minimize total spare capacity considering *p*-cycles also performing the function of *m*-cycles. Numerical results reveal that there is no or fewer extra cost required in case of combined *p*-cycles with *m*-cycles compared to pure *p*-cycles. The reason for no extra cost in combined *p*-cycles with *m*-cycles is that here some *p*-cycles are performing the dual role of protection as well as detection and localization of failed span of the network.

Keywords Elastic optical network · Fault detection and localization · *m*-Cycles · *p*-Cycles · Survivability

D. C. Jaiswal (✉) · R. Asthana
Department of Electronics Engineering HBTU, Kanpur, Uttar Pradesh, India
e-mail: deohbti@gmail.com

R. Asthana
e-mail: rachnaasthana@rediffmail.com

1 Introduction

In order to fulfill the rapidly increasing traffic demand, the elastic optical networks (EONs) are indispensable. This is due to its many desirable properties including flexible data rate, flexible spectrum allocation and efficient spectrum utilization [1–3]. The conventional method of working of wavelength division multiplexing (WDM) networks are fixed spectral band, i.e., International Telecommunication Union (ITU)-T grid, the capacity of a fixed spectral band (e.g., 50 GHz) is underused as traffic demand is lesser, leading to a large amount of capacity wastage [4]. The optical transmission and networking technologies become more flexible, efficient, and scalable with an advanced utilization of capacity wastage [3]. Migration from the fixed ITU-T grid to flexible grid technique and division of 50 GHz spectral band into multiple 12.5 GHz spectral bands enhances the elasticity of WDM optical networks category [5]. 12.5 GHz spectral band referred to as frequency slots. The EONs are based on the OFDM modulation technique [6]. OFDM is a multiplexing technique which is used to transmit high-speed single carrier data stream by dividing several orthogonal low speed multi-carriers subdata streams [2]. Fixed bandwidth channel exists which is the limitation of WDM, can be split into multiple subchannels. By combining these two properties of OFDM multiplexing and migration from fixed bandwidth to flexible bandwidth, a flexible bandwidth channel can be formed as per the requirement of transmitted data rate in EONs [7]. Increasing traffic demand, efficient spectrum utilization resolves the issue of capacity efficient solutions. In EONs, the constraints of the requirement of the same group of frequency slots must be used on all the links crossed by the route of demand pair is referred as spectrum continuity and the requirement that all frequency slots allocated to the route of demand pair must be adjacent is referred as spectrum contiguousness [5]. Due to fine granularity, there is an increase in the number of the channel in EONs that was left as capacity wastage in WDM networking [8, 9]. They (EONs) are beneficial due to their better spectrum utilization, flexibility in bandwidth assignment, and increase the number of channels [1]. In EONs, the channel with the desired spectral band is formed by combining a group of contiguous slots in the state of using the channel of a preset fixed spectral band. It improves the elasticity of spectral band allocation which reduces capacity wastage resulting in improved overall spectrum frequency usage [10]. Network failures for a small-time duration affect the tremendous data loss in affording service of data transmission. Survivability is an imperative requirement against network failures though requires additional resources [11, 12]. For the abovementioned improved EONs, survivability is the main consideration for reliable communication [13]. When survivability is on focus, more redundant resources are needed for providing backup protection. Preconfigured protection cycles (*p*-cycles) structure are deployed with additional redundant resources to achieve survivability requirement. Fault detection and localization by monitoring cycles (*m*-cycles) are needed before providing *p*-cycle protection. The *m*-Cycle concept was studied in the Ph.D. thesis by Hongquing Zeng [14]. Here *p*-cycles have been used as *m*-cycles for

self-fault detection and localization in elastic optical networks [15]. *p*-Cycle structure of a network with properties of *m*-cycle without significant added cost is not initially obvious. *m*-Cycles must cover all the spans of the network as only the on cycle span while *p*-cycles also protect straddling links which makes it more efficient [16–18]. Straddling span relationship increases the efficiency of *p*-cycle as it protects on cycle spans as well as straddling spans. The feature of *p*-cycle to protect both on cycle spans and straddling spans make it more beneficial, but the basic issue of reconciliation of two concepts full span protection and failure localization under the same design model arises while considering the minimization of the total cost. Self-failure isolating survivable network is found as the modern classification when the above concept becomes feasible and practical which may make *p*-cycle protection in elastic optical networks more advantageous. The basic structure of EON mainly consists of bandwidth variable transponders (BVTs), bandwidth variable cross connects (BV-WXCs) and erbium-doped fiber amplifiers (EDFAs) [19].

The rest of the paper is prepared as follows. Section 2 discusses the *p*-cycle concept in EONs. Section 3 explains *p*-cycles function as their own *m*-cycle. Section 4 provides an ILP optimization model for optimized spare capacity considering *p*-cycles combined with *m*-cycle in EONs. In Sect. 5 simulations and results are shown. Finally, the conclusions are summarized in Sect. 6.

2 *p*-Cycle Concept in Elastic Optical Networks

p-Cycle concept in elastic optical networks is discussed in this section presented in Fig. 1. Three *p*-cycles (1-2-3-4-1), (1-6-7-4-1) and (1-2-3-4-7-6-1) are shown here.

Fig. 1 *p*-Cycle protection concept [20]
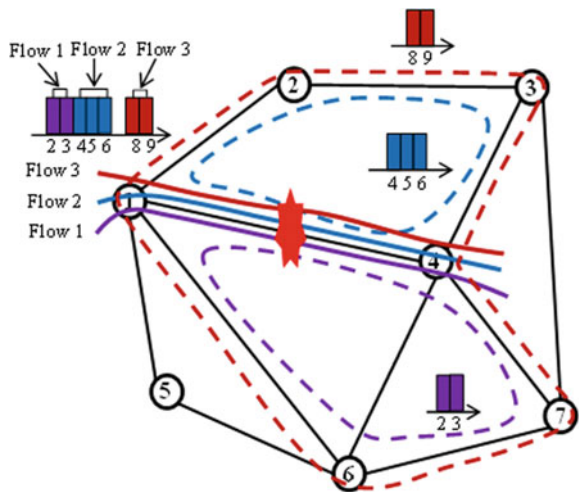
The span (1-4) is traversed by three working flows. The allotted FSs for these working flows are (FS 2-3, Flow 1), (FS 4-6, Flow 2), and (FS 8-9, Flow 3) respectively.

If a failure occurs in span (1-4) three *p*-cycles mentioned above collectively protect working flow through it [21]. The span (1-4) is an on cycle span of two *p*-cycles (1-6-7-4-1) and (1-2-3-4-1). Particularly the *p*-cycles (1-6-7-4-1) and (1-2-3-4-1) protect working flows (FS 2-3, Flow 1) and (FS 4-6, Flow 2) of span (1-4) as an on cycle span respectively. The span (1-4) is also straddling(chord) span of the *p*-cycle (1-2-3-4-7-6-1). The *p*-cycle (1-2-3-4-7-6-1) protects the working flow (FS 8-9, Flow 3) of span (1-4) as straddling span providing two protection path (1-2-3-4) and (1-6-7-4). Due to *spectrum contiguity* constraint, only one protection path can be used to protect the working flow (FS 8-9, Flow 3) of span (1-4) because working flow cannot be split into different routes [20]. The *spectrum contiguity* requirement creates a difference between these *p*-cycles and the orthodox *p*-cycles.

3   *m*-Cycle Concept in Elastic Optical Networks

Monitoring cycles (*m*-cycles) are those cycles in which all nodes and links of the given network are covered at least by one cycle. In each monitoring cycle, one node is allotted with a transceiver thus a supervisory channel for dedicated loop is formed.

The *m*-cycles are those cycles which are covering all the spans with no one span in the network can be left uncovered by at least one cycle and the same set of cycles cannot be used for two different spans. If any span fails then it can be uniquely identified by abovementioned properties. The unique identification of failed spans can be recognized by the different set of cycles by displaying an alarm state. The assumption about this is anyone node of each cycle has one monitor signal. The *m*-cycle is used for fault detection and localization which is needed before protection and restoration [14, 15, 22]. *m*-Cycle concept is introduced in this section presented in Fig. 2. Here four *m*-cycles i.e. (1-2-3-1), (1-3-4-1), (1-4-5-1) and (2-3-4-5-2) are required for monitoring five node and eight span network. With the *m*-cycle concept, the combination of four cycles can monitor the eight spans. Specifically, we use

Fig. 2 Network covering *m*-cycles

Links	*m*-cycles			
	1	**2**	**3**	**4**
1–2	1	0	0	0
1–3	1	1	0	0
1–4	0	1	1	0
1–5	0	0	1	0
2–3	1	0	0	1
2–5	0	0	0	1
3–4	0	1	0	1
4–5	0	0	1	1
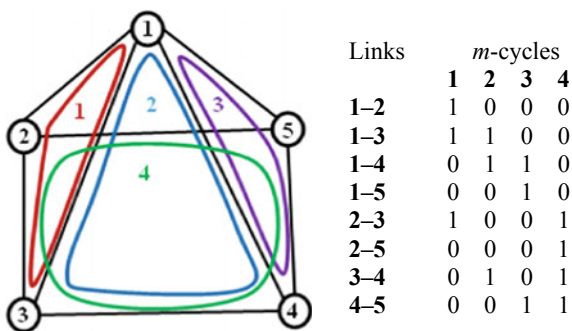
cycle (1-2-3-1), cycles (1-2-3-1) and (1-3-4-1), cycles (1-3-4-1) and (1-4-5-1), cycle (1-4-5-1), cycles (1-2-3-1) and (2-3-4-5-2), cycle (2-3-4-5-2), cycles (1-3-4-1) and (2-3-4-5-2), and cycles (1-4-5-1) and (2-3-4-5-2) to monitor the span 1-2, 1-3, 1-4, 1-5, 2-3, 2-5, 3-4, and 4-5 respectively. Monitoring cycle set for each span is uniquely identified that can be used for fault detection and localization.

4 ILP Optimization Model

An ILP model can be used for the optimization of network design. In this section, an ILP model is developed to minimize total spare capacity in elastic optical networks considering *p*-cycles combined with *m*-cycles. Following sets, parameters and variables are used by such an ILP model.

Sets:
S: Set of network spans indexed by j.
N: Set of network nodes indexed by n.
D: Set of network demand pairs indexed by r.
P: Set of network *p*-cycles indexed by p.
P_i: Set of eligible *p*-cycles for ith link failure recovery.
D_i: Set of demand pairs whose working paths include link i.

Parameters:
d_r: The number of frequency slots (demand) required for demand pair r.
π_j^p: A binary data that equals 1 if *p*-cycle p crosses span j; 0 otherwise.
α: A weight factor

Variables:
$x_i^{p,r}$: equal to 1 if failure of any link i of working path between demand pair r can be recovered by an eligible ring of that link; 0, otherwise.
n_p: Number of spare capacity for ring p.
s_j: Spare capacity reserved on span j.
c: The maximum index of frequency slots assigned on network links.
The objective of ILP is to Minimize:

$$\sum_{j \in S} s_j + \alpha \cdot c \tag{1}$$

Constraints:

$$\sum_{p \in P_i} x_i^{p,r} = 1 \quad \forall j \in S, \quad \forall r \in D_i \tag{2}$$

$$\sum_{\forall r \in D_i} d_r \cdot x_i^{p,r} = n_p \quad \forall j \in S, \quad \forall p \in P_i \tag{3}$$

$$s_j = \sum_{p \in P} \pi_j^p \cdot n_p \quad \forall j \in S \tag{4}$$

$$\sum_{p \in P : \pi_i^p = 1 \text{ and } \pi_j^p = 1} n_p \geq 1, \quad \forall \{i, j\} \in S^2, i \neq j \tag{5}$$

$$\sum_{p \in P : \pi_j^p = 1} n_p \geq 1, \quad \forall j \in S \tag{6}$$

$$s_j <= c \quad \forall j \in S \tag{7}$$

$$\sum_{r \in D_i} d_r <= c \quad \forall j \in S \tag{8}$$

Objective (1) is to minimize the total spare capacity required for *p*-cycle formation in the entire network. Constraint (2) ensures that failure of any link *i* of working path between demand pair *r* can be recovered by only one eligible *p*-cycle due to the requirement of *spectrum contiguouity*. Constraint (3) ensures that cycle *p* has sufficient capacity for protection that can restore the working path between demand pair *r* upon failure of any link *i*. Constraint (4) calculates the total number of frequency slots on the links are reserved for making all the *p*-cycles. Constraints (5) and (6) ensures that the requirement of *m*-cycle cover that the same set of cover cycles cannot be used for two different spans and that at least one cycle covers each span. Constraint (7) says that the total number of frequency slots on any link *j* should always be lesser than the maximal FS index. Constraint (8) ensures that total working demands on link *j* by eligible demand pair should always be lesser than the maximal FS index.

5 Simulations and Results

The simulations for optimized results have been performed on IBM ILOG CPLEX OPTIMIZATION STUDIO 12.6.3.0 simulation software on an Intel Core i5 equipped with 2.4 GHz CPU and 4 GB RAM. Performance is evaluated with pure *p*-cycle and combined *p*-cycle with *m*-cycle by running simulations for three test networks, including (1) the 6 node 8 span network shown in Fig. 3a, (2) the 8 node 11 span network shown in Fig. 3b and (3) the 14 node 21 span network shown in Fig. 3c. Simulation results are presented in Table 1.

Comparing the total spare capacity with pure *p*-cycle and combined *p*-cycle with *m*-cycle, performance indicates there is no extra cost is needed for fault detection and localization due to *p*-cycles are selected for protection include *m*-cycles in case of full as well as some partial transmission. It makes sense of monotonic increment of total spare capacity with increasing network demands whereas cycle cover for monitoring is purely network topology dependent. The main benefaction of this work gives the answer of the feasibility of integrating *m*-cycle functionality into *p*-cycle design.
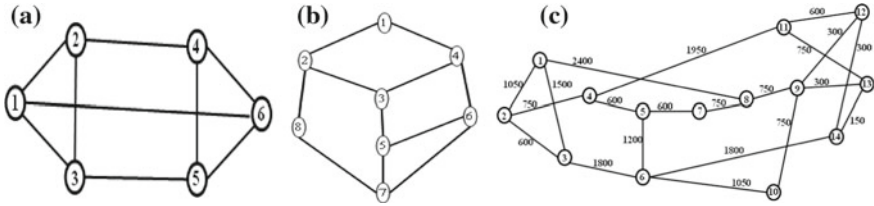
Fig. 3 Test networks used in simulations, **a** 6 node 8 span network, **b** 8 node 11 span network, **c** 14 node 21 span network

Table 1 Combined *p*-cycle/*m*-cycle simulations for 3 different networks

Networks topology	Total spare capacity		
	Pure *p*-cycle design	Combined design cost	% Increase
6n9s	46	50	8.69%
8n11s	54	54	0%
14n19s	66	66	0%

6 Conclusion

In elastic optical networks, fault detection and localization is done by monitoring cycles (*m*-cycles). We applied *m*-cycle to elastic optical networks with *p*-cycle protection and developed ILP optimization models to minimize total spare capacity. Numerical results indicate that without or some extra cost the combined *p*-cycles with *m*-cycles can be used for fault diagnosis (detection and localization) and protection both. Some *p*-cycle structures can be assigned the task of monitoring any loss of light for various fault localization as an *m*-cycles and also used for protection. It opens a new class of survivability scheme in elastic optical networks that combine *p/m*-cycles.

References

1. Wu, J., Liu, Y., Yu, C., Wu, Y.: Survivable routing and spectrum allocation algorithm based on *p*-cycle protection in elastic optical networks. J. Light Electron Opt. **125**(16), 4446–4451 (2014)
2. Zhang, G., et al.: A survey on OFDM-based elastic core optical networking. IEEE Commun. Surv. Tutor. **15**(1), 65–87 (2013)
3. Jinno, M., Takara, H., Kozicki, B., Tsukishima, Y., Sone, Y., Matsuoka, S.: Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies. IEEE Commun. Mag. **47**(11), 66–73 (2009)
4. Keiser, G.: Optical Fiber Communications, 4th edn. Tata McGraw-Hill (2010)
5. Wei, Y., Shen, G., Bose, S.K.: Applying ring cover technique to elastic optical networks. In: Proceedings of ACP (2013)

6. Christodoulopoulos, K., Tomkos, I., Varvarigos, E.: Elastic bandwidth allocation in flexible OFDM-based optical networks. IEEE/OSA J. Lightwave Technol. **29**(9), 1354–1366 (2011)
7. Nag, A., Tornatore, M., Mukherjee, B.: Optical network design with mixed line rates and multiple modulation formats. J. Light wave Technol. **28**(4), 466–475 (2010)
8. Ramaswami, R., Sivarajan, K.N.: Optical Networks: A Practical Perspective, 2nd edn. Morgan Kaufmann (2009)
9. Wang, Y., Cao, X., Hu, Q., Pan, Y.: Towards elastic and fine-granular bandwidth allocation in spectrum-sliced optical networks. J. Opt. Commun. Netw. **4**(11), 906–917 (2012)
10. Chen, X., Ji, F., Zhu, Z.: Service availability oriented *p*-cycle protection design in elastic optical networks. J. Opt. Commun. Netw. **6**(10), 901–910 (2014)
11. Zhou, D., Subramaniam, S.: Survivability in optical networks. Network **14**(6), 16–23 (2000)
12. Ramamurthy, S., Sahasrabuddhe, L., Mukherjee, B.: Survivable WDM mesh networks. J. Lightwave Technol. **21**(4), 870–883 (2003)
13. Ruan, L., Xiao, N.: Survivable multipath routing and spectrum allocation in OFDM-based flexible optical networks. J. Opt. Commun. Netw. **5**(3), 172–182 (2013)
14. Zeng, H.: Monitoring-cycle based fault detection and localization in mesh all optical networks. Ph.D. Dissertation, Carleton University, Ottawa, Canada (2007)
15. Grover, W.D., Grue, A.: Self-fault isolation in transparent *p*-cycle networks: *p*-cycles as their own *m*-cycles. IEEE Commun. Lett. **11**(12) (2007)
16. Grover, W.D., Stamatelakis, D.: Cycle-oriented distributed preconfiguration: ring-like speed with mesh-like capacity for self-planning network restoration. In: Proceedings of IEEE International Conference on Communications, Atlanta, Georgia, USA, pp. 537–543 (1998)
17. Kiaei, M.S., Assi, C., Jaumard, B.: A survey on *p*-cycle protection method. IEEE Commun. Surv. Tutor. **11**(3), 53–70 (2009)
18. Asthana, R., Singh, Y.N., Grover, W.D.: *p*-Cycles: an overview. IEEE Commun. Surv. Tutor. **12**(1), 97–111, First Quarter 2010
19. Zhang, Xiaoning, Tan, Shufang: Green optical networks with connection availability guarantee (Optical Society of America publishing (OSA)). Opt. Express **23**(13), 17456–17466 (2015)
20. Wei, Y., Xu, K., Jiang, Y., Zhao, H., Shen, G.: Optimal design for *p*-cycle-protected elastic optical networks. Photon Netw. Commun. **29**(3), 257–268 (2015)
21. Asthana, R.: Study of *p*-cycle based protection in optical networks and removal of its shortcomings. Ph.D. Thesis, Indian Institute of Technology, Kanpur (2007)
22. Eshoul, A., Mouftah, H.: Survivability approaches using *p*-cycles in WDM mesh networks under static traffic. IEEE/ACM Trans. Netw. **17**(2), 671–683 (2009)

Analysis of Modified Swastika Shaped Slotted (MSSS) Microstrip Antenna for Multiband and Ultra-wideband Applications

Devesh Tiwari, Mohd. Gulman Siddiqui, A. K. Saroj, J. A. Ansari, Neelesh Agrawal and Mukesh Kumar

Abstract In this paper, a modified swastika shaped slotted (MSSS) microstrip antenna is designed for multiband and ultra-wideband (UWB) applications. The analysis of the proposed microstrip antenna has been analyzed using HFSS simulation software. The multiband behavior of antenna has been achieved for UWB, Ku-band, and K-band applications. The proposed antenna covers almost entire UWB band (3.1–10.6 GHz) useful for commercial purposes such as Wi-Fi (2.4 GHz), WiMAX (3.3–3.8 GHz), and WLAN (5.15–5.825 GHz) and also useful for Ku-band (12–18 GHz) and K-band (18–27 GHz) applications. The antenna resonates at different frequencies 6.65 GHz, 13.87 GHz, 16.99 GHz, 19.52 GHz, and 27.91 GHz having return loss of -33.785 dB, -12.401 dB, -17.869 dB, -19.313 dB, and -13.267 dB, respectively. Its radiation characteristics show that it has better UWB and multiband performances for wireless applications.

1 Introduction

The various advantages of the microstrip patch antenna include small size, low profile, conformal to any shapes, and ease of fabrication which is compatible for numerous applications [1]. The attractive features of microstrip antennas for wireless applications have got the attention of researchers and scientists [2]. For wireless communication systems, the antenna should have high gain, compact in size multiband [3, 4], and Ultra-wideband [5] frequency along with better radiation characteristics. In 2002, the frequency spectrum from 3.1 to 10.6 GHz was allocated for UWB applications by the Federal Communication Commission (FCC) [5]. UWB antenna is a key component of UWB wireless communication system. UWB communication system provides a huge frequency band, the permitted power spectral density of the

D. Tiwari (✉) · Mohd. Gulman Siddiqui · A. K. Saroj · J. A. Ansari · N. Agrawal · M. Kumar
Department of Electronics and Communication, University of Allahabad, Allahabad, UP, India
e-mail: deveshcdpm@gmail.com

189

UWB signal is rather limited to avoid interference with other systems [6]. The evolution of UWB systems has been attracted much attention for wireless communication systems. The UWB systems have several advantages such as high data rates, low power level, low cost, and ease of fabrication. Various shapes of monopole antennas rectangular or circular printed monopole [6] have been reported for a compact UWB antenna. The microstrip antennas have drawn attention in recent years for their large bandwidths, simple topologies, omnidirectional radiation patterns, low cost, and ease of fabrication [2]. But there are some narrow bands which are used for other communication systems such as Wi-Fi (2.4 GHz), WiMAX (3.3–3.8 GHz), and WLAN (5.15–5.825 GHz) which cause interference with the UWB systems [7–9].

Multiband behavior of antenna can be achieved by loading of slots [3] and notches [4] on the patch or by using fractals. The loading of slots or notches on the ground plane of MSA is known defected ground structure (DGS). DGS on the patch is useful for WLAN/WiMAX applications [7, 8]. Recently various types of fractal geometries have been introduced to modify the MSA structure. The main feature of fractal geometries is to reduce the overall geometrical size and increase its effective length. Fractal structures are uneven shapes which can be separated into subparts and every subpart is a small copy of the overall shape [10, 11]. Some popular fractal geometries are—Sierpinski carpet, Sierpinski gasket (triangle), Koch curve, Minkowski curve, Cohen-Minkowski curve, etc. Miniaturization and space filling are the two main characteristics of fractal antennas, provide multiband performance along with improved gain and directivity.

In this paper, a modified swastika shaped slot loaded microstrip antenna is proposed and analyzed using HFSS simulation software. The antenna is mainly composed of a metallic patch with good radiation performance for wireless applications.

2 Basic Microstrip Antenna

A simple microstrip antenna has two radiating metallic (copper) structure one is on the top side (patch) of the substrate and other on the bottom side (ground) as shown in Fig. 1a. The patch can be of any shape, but generally, the regular shape is preferred
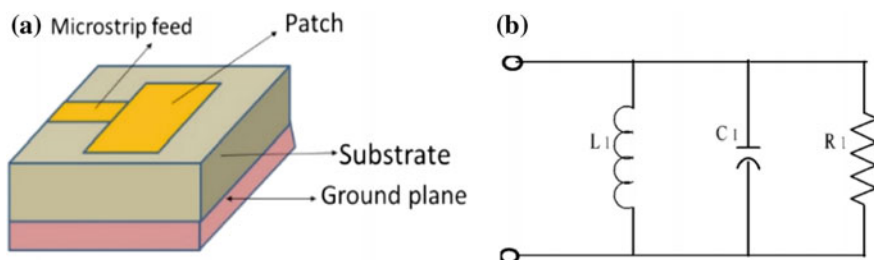
Fig. 1 Simple microstrip antenna. **a** Primary structure. **b** Equivalent circuit
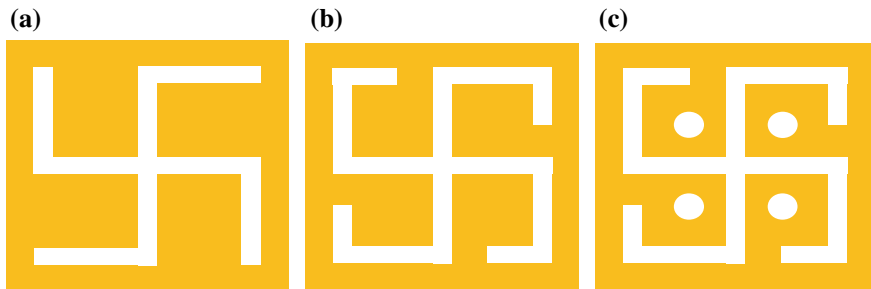
Fig. 2 The stepwise analysis of the proposed modified swastika shaped slotted antenna: **a** Design 1, **b** Design 2, **c** Design 3

for simplified analysis of microstrip antenna. In this design, FR4 dielectric substrate is used having dielectric constant of 4.4 and loss tangent of 0.02.

Figure 1, shows a behavior simple microstrip antenna and its equivalent circuit. A simple microstrip antenna behaves as a parallel RLC circuit.

The equivalent circuit of MSA consists of a parallel combination of resistance (R_1), inductance(L_1), and capacitance (C_1) which is shown in Fig. 1b.

Since

$$C_1 = \frac{LW\varepsilon_0\varepsilon_e}{2h} \cos^2(\pi Y_0/L) \tag{1}$$

$$R_1 = Q/\omega_r^2 C_1 \tag{2}$$

$$L_1 = 1/\omega_r^2 C_1 \tag{3}$$

The stepwise analysis of proposed swastika antenna is shown in Fig. 2. In the first step, a simple swastika shaped slot is loaded microstrip antenna. Further, in the next step, every four arms of swastika additionally added and finally circular slot of fixed radius in every four arms is loaded on the patch.

3   Description of MSSS Microstrip Antenna

In this microstrip antenna, a square dielectric substrate is sandwiched by two square metallic plates, one patch and another ground plane. Here a metallic patch of dimension $30 \times 30 \times 1.6$ mm^2 is placed on FR4 glass epoxy substrate with dielectric constant 4.4 and loss tangent 0.02. Further, a metallic ground plane of dimension 50 $\times$ 50 mm^2 is placed on the bottom of the dielectric substrate. The detailed analysis of antenna dimensions is explained in Table 1. The analysis of the proposed modified swastika shaped slotted antenna is shown in Fig. 2 has the first step a simple swastika

Table 1 Dimensions of the proposed MSSS antenna

Symbols	Parameters	Dimensions (mm)
L_P	Length and width of the patch	30
W_P	Length and width of the patch	30
L_1	Length of slots (plus shaped) loaded on a patch	24
L_2	Length of slots (swastika arm) loaded on a patch	11
L_3	Length of slots (rotated swastika arm) loaded on a patch	7
R	Radius of the circular slot	1.5
L_F	Length of the feed line	10
W_F	Width of the feed line	2
t	Width of slots and feed line	2
L_S	Length of the dielectric substrate	50
W_S	Width of the dielectric substrate	50
h	Thickness of substrate	1.6
L_G	Length of the ground plane	50
W_G	Width of the ground plane	50

shaped slot is loaded on the patch. The pulse-shaped slot of length 24 mm and the arm of the swastika is 11 mm is initially loaded on the patch. Further in second design additional rotated arm of swastika slot of length 7 mm is loaded. Finally, the circular slot of radius 1.5 mm is loaded on the center of each four side swastika arm. Here the width of swastika arm is taken as 2 mm. A simple microstrip line feeding is used of dimension 10×2 mm^2 on the top of the substrate.

Figure 3 shows the current distribution of the proposed MSS slotted antenna. It shows the current is maximum at the edges of the patch and the corner of slots.
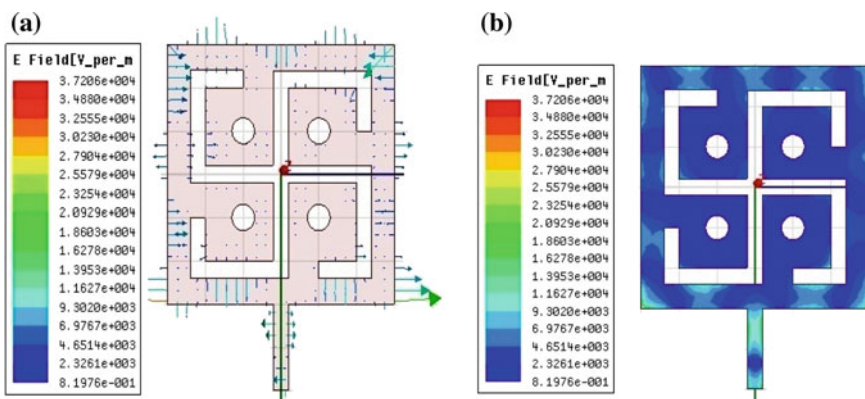
Fig. 3 The current distribution plot of the proposed microstrip antenna. **a** Vector electric field, **b** magnitude electric field

Figure 4, shows the top view, side view, and bottom view of proposed MSSS microstrip antenna.
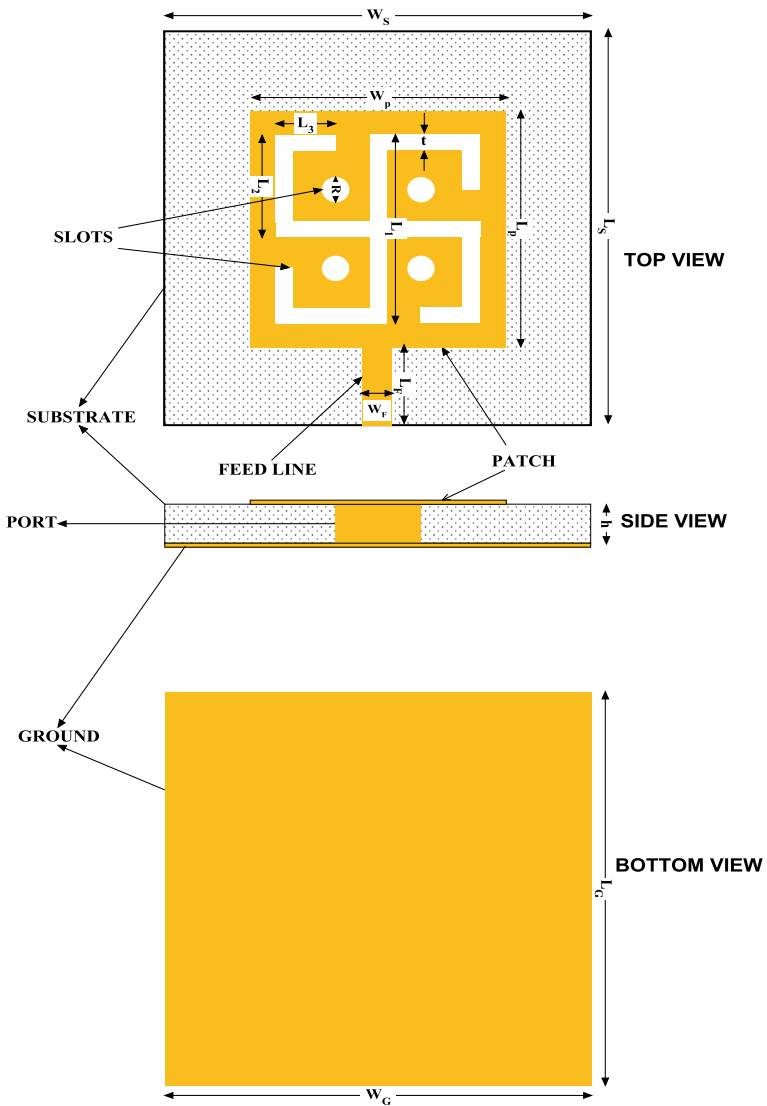
Fig. 4 Top, side, and bottom view of proposed MSSS microstrip antenna

4 Results and Discussion

In this design, a modified swastika shaped antenna is simulated by ANSOFT HFSS software using finite element method (FEM) solver. The result obtained from this antenna can be used for multiband operation at 6.655, 13.88, 16.99, and 27.91 GHz and ultra-wideband at 6.655 GHz resonant frequency. It is observed that the UWB application is obtained from 1.97 to 10.16 GHz frequency band which is near the range provided by FCC for UWB, i.e., 3.1 to 10.6 GHz and for Ku-band [12–18 GHz] and K-band [18–27 GHz] application. Hence obtained results are best suited for the commercial purpose.

Figure 5, shows the S_{11} parameter versus frequency plot of the proposed antenna. Table 2, described the comparative analysis of S_{11} parameter (Return Loss) at a different resonant frequency of the antenna. Initially, there are four bands available at different resonant frequencies but after modification in the pattern, the number of the frequency bands is increased with better return loss and bandwidth. Figure 6, shows the comparative VSWR versus frequency plot of the designed microstrip antenna. The antenna is properly matched for VSWR value which lies between 1 and 2. Figure 7 shows the radiation pattern of the proposed microstrip antenna. It is observed that at lower frequency range better radiation pattern is obtained. As the frequency increases higher number of modes are generated which creates back lobes in radiation pattern.
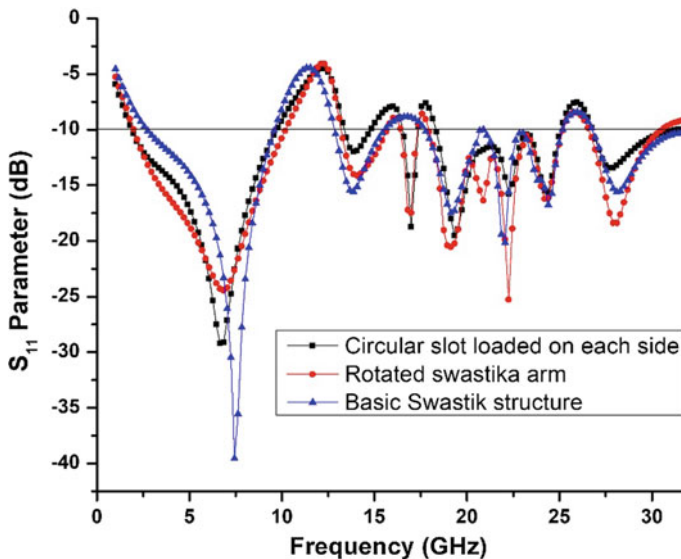
Fig. 5 Comparative plot of the S_{11} parameter versus frequency plot

Table 2 Comparative analysis of S_{11} parameter of the antenna at a different resonant frequencies

Design	Design 1		Design 2		Design 3	
Parameters	Resonant frequency (GHz)	Maximum return loss (dB)	Resonant frequency (GHz)	Maximum return loss (dB)	Resonant frequency (GHz)	Maximum return loss (dB)
fr1	7.435	−39.553	6.85	−24.447	6.65	−33.785
fr2	13.870	−15.583	14.07	−14.121	13.87	−12.401
fr3	–	–	16.99	−17.489	16.99	−17.869
fr4	19.330	−17.310	22.25	−25.271	19.52	−19.313
fr5	28.100	−15.911	27.91	−18.396	27.91	−13.267
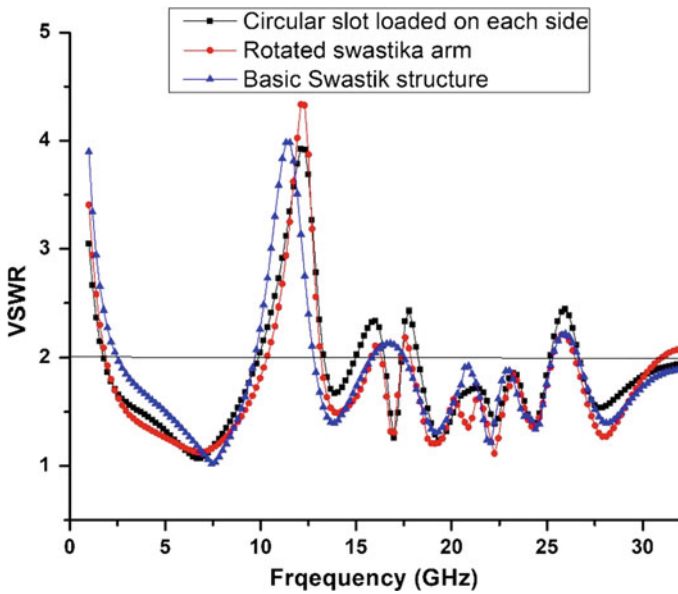
Fig. 6 Comparative VSWR versus frequency plot for a different configuration

Figure 8 shows gain versus frequency plot of the proposed antenna. It shows the gain at each frequency from 2 to 32 GHz. Table 3 shows the gain of the proposed antenna as 6.39 dB, 5.68 dB, 5.53 dB, 5.97 dB, and 5.15 dB at the frequency of 6.74 GHz, 14.34 GHz, 17.01 GHz, 18.85 GHz, and 27.88 GHz, respectively. The achieved radiation efficiency is approximately 72%.
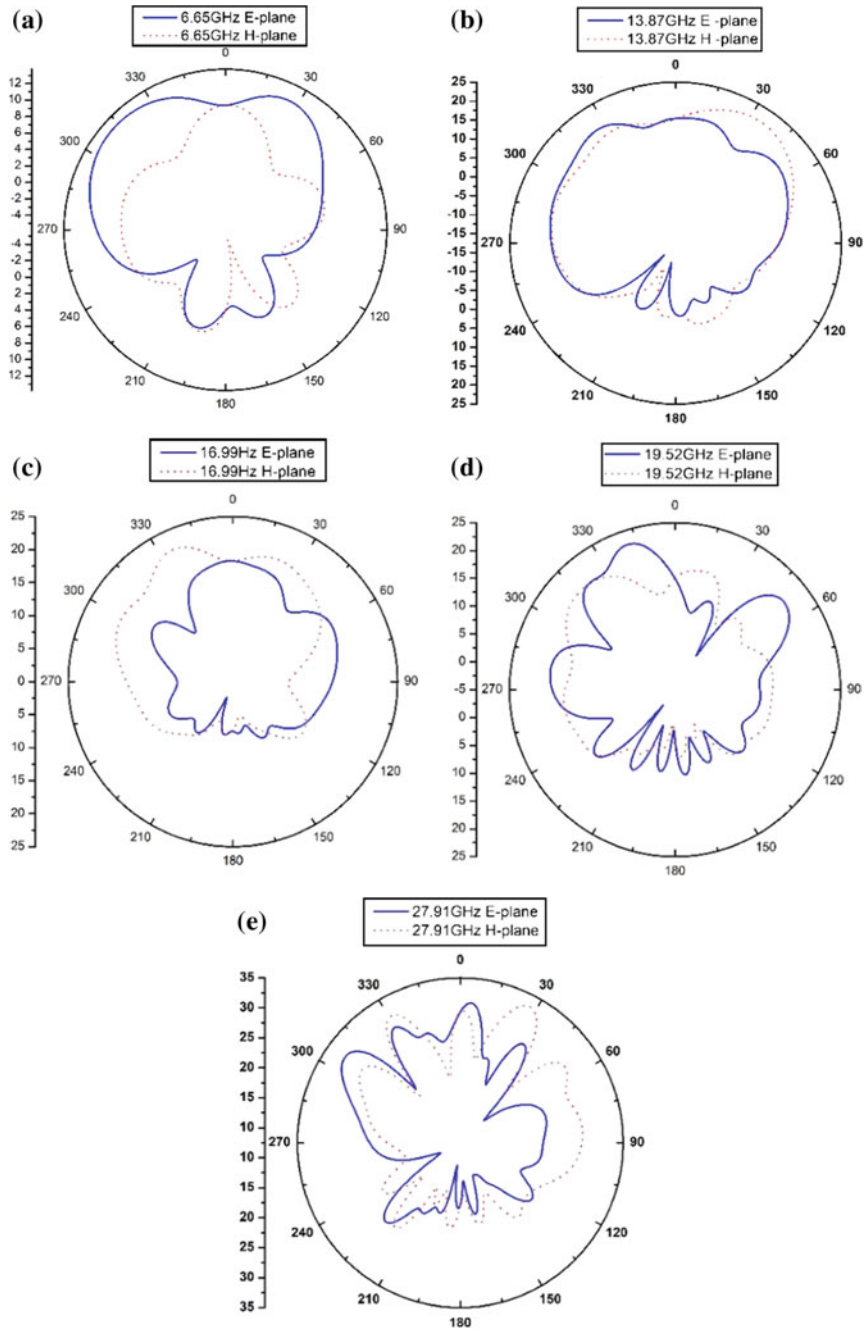
Fig. 7 Radiation pattern of proposed MSSS antenna at **a** 6.65 GHz, **b** 13.87 GHz, **c** 16.99 GHz, **d** 19.52 GHz, **e** 27.91 GHz
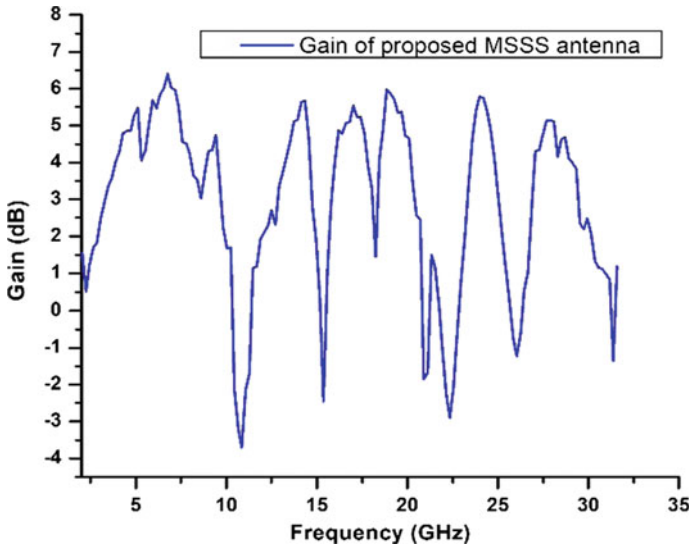
Fig. 8 Gain versus frequency plot of proposed MSSS antenna

Table 3 Analysis of gain of the proposed MSSS antenna at different frequencies

Frequency (GHz)	Gain (dB) of proposed antenna
6.74	6.39
14.34	5.68
17.01	5.53
18.85	5.97
27.88	5.15

5 Conclusions

The modified swastika shaped slotted (MSSS) microstrip antenna is designed for multiband and ultra-wideband applications which covers entire UWB band (3.1–10.6 GHz) useful for commercial purposes such as Wi-Fi (2.4 GHz), WiMAX (3.3–3.8 GHz), and WLAN (5.15–5.825 GHz) as well as Ku-band (12–18 GHz) and K-band (18–27 GHz) applications. The multiband performance of antenna at different frequencies 6.65, 13.87, 16.99, 19.52, and 27.91 GHz is achieved. Its radiation characteristics are simulated by HFSS simulation software and the results show that it has better UWB and multiband performances.

References

1. Bahl, I.J., Bhartia, P.: Microstrip Antennas, Dedham. Artech House, MA (1980)
2. Carver, K.R., Mink, J.W.: Microstrip antenna technology. IEEE Trans. Antennas Propag. **AP-29**, 2–24 (1981)
3. Tiwari, D., Ansari, J.A., Siddiqui, M.G., Saroj, A.K.: Analysis of modified square sierpinski gasket fractal microstrip antenna for wireless communications. Int. J. Electron. Commun. (AEÜ) **94**(2018), 377–385 (2018)
4. Siddiqui, M.G., Saroj, A.K., Tiwari, D., Ansari, J.A.: Multi-band fractaled triangular microstrip antenna for wireless applications. Prog. Electromagn. Res. M **65**, 51–60 (2018)
5. Ahmed, O., Sebak, A.R.: A printed monopole antenna with two steps and a circular slot for UWB applications. IEEE Antenna Wirel. Propag. Lett. **7**, 411–413 (2008)
6. Federal Communications Commission, Washington, DC, USA, "Federal Communications Commission revision of Part 15 of the commission's rules regarding ultra-wideband transmission system from 3.1 to 10.6 GHz," ET-Docket, pp. 98–153 (2002)
7. Kaiser, T., Zheng, F., Dimitrov, E.: An overview of ultra-wide-band systems with MIMO. Proc. IEEE **97**(2), 285–312 (2009)
8. Ansari, J.A., Mishra, A., Yadav, N.P., Singh, P., Vishvakarma, B.R.: Compact tripple U-shaped slot loaded circular disk patch antenna for bluetooth and WLAN application. Int. J. Microw. Opt. Technol. **6**, 88–96 (2011)
9. Singhal, S., Singh, A.K.: Modified theta-shaped monopole antenna with defected ground structure for UWB applications. Microw. Opt. Technol. Lett. **57**(7), 1625–1632 (2015)
10. Anguera, J., Boada, L., Puente, C., Borja, C., Soler, J.: Stacked H-shaped microstrip patch antenna. IEEE Trans. Antennas Propag. **52**(4) (2004)
11. Puente, C., Romeu, J., Pous, R., Cardma, A.: On the behavior of the Sierpinski multiband fractal antenna. IEEE Trans. Antennas Propag. **46**, 517–521 (1998)

A Compact Star Shaped Fractal Antenna for Multiband Applications

Shweta Sharma and Som Pal Gangwar

Abstract In this paper, the design and analysis of a fractal-based patch antenna is described for the multiband applications. The proposed structure is designed on a single substrate. On one side of the substrate, star shaped fractal geometry is printed and on the other side of the substrate, a defected ground structure is created. The antenna is simulated using HFSS software. To verify the simulated results, a prototype of the structure is fabricated using FR4 substrate. The proposed antenna exhibits triple band characteristics. It has three resonant frequencies at 2.25 GHz, 3.61 GHz, and 5.16 GHz with the return loss of -12.25 dB, -32.5 dB, and -35.75 dB respectively. The simulated and measured results are compared and these are close enough. The proposed antenna can be used for triple band applications in S-band (2–4 GHz) and C-band (4–8 GHz).

Keywords Star shape · Fractal antenna · Triple band · S-band · C-band · Self similarity

1 Introduction

In the last few decades, with the development of wireless communication, multiband antennas with small size and stable far field characteristics are in great demand [1, 2]. Currently, a lot of research is going on to get multiband features in a single antenna. Multiband antennas have two important advantages: (i) multiple frequency bands can be obtained in a single antenna; (ii) good signal to noise ratio can achieve [3]. Many techniques have been used in the past to get dual-band characteristics in antenna. One of them is to use different types of slot like cross slot, a circular slot, a square slot, and an offset circular slot [4–7]. The word fractal has derived from the

S. Sharma (✉) · S. P. Gangwar
Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur 228118, UP, India
e-mail: shweta.aish23@gmail.com

S. P. Gangwar
e-mail: gangwar_sp@rediffmail.com

Latin word "fractus" which means any uneven, extremely irregular curves or design that repeat themselves. The geometry of fractal is important because the effective length of the fractal antenna can be increased while keeping the shape same. The fractal antenna can be designed by an iterative mathematical process called iterative function scheme (IFS). There are many features of fractal antennas which make them useful for many applications. Some of them are: self similarity, never ending pattern, small dimension, and space-filling ability. There are numerous applications of fractal antennas [8–10].

This article presents the star shaped fractal antenna with defected ground structure (DGS) for triple band applications. Microstrip line has been used to feed the proposed antenna. This antenna structure is designed for three frequency bands. It also provides a stable far field characteristics. For better representation, present article is divided into four different sections: (i) introduction; (ii) antenna geometry; (iii) results and discussion; and (iv) conclusion.

2 Design of Antenna Geometry

The geometry of the proposed fractal antenna is shown in Fig. 1. The substrate used for fabrication of proposed structure is of FR4 epoxy material having dielectric constant 4.4 and height 1.6 mm. Figure 1 shows the designed fractal geometry with three iterations of star shape. The dumbbell shaped microstrip line is used for feeding the structure. The advantage of microstrip line feeding is that it eliminates the spurious feed radiation. DGS is used for miniaturization of structure and also to enhance its

Fig. 1 Geometry of the proposed antenna
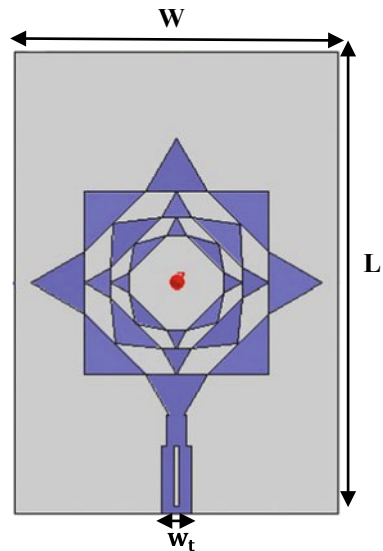
performance. The simulation analysis has been done using HFSS software which is a finite element method based solver for electromagnetic structures.

The basic or first iteration of the designed antenna structure consists of a single star shape with microstrip feedline. Insertion of star structure at the center of first iterated structure resulted into second iteration of the antenna structure. This central star is scaled down by a ratio of 0.33. Then, addition of another star structure at the center of second iteration structure derives the third iterative structure (Fig. 2 and Table 1).

(a) Top view   (b) Bottom view
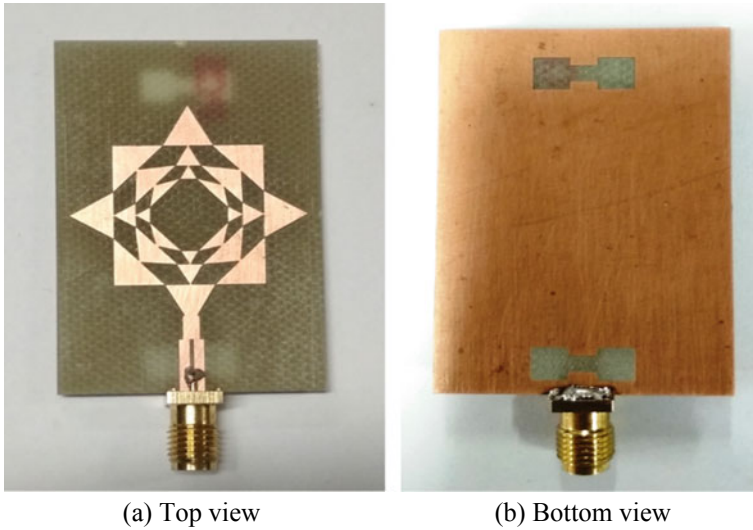
Fig. 2 Fabricated antenna

Table 1 Various parameters of the proposed antenna

Parameter	Symbol	Value (in mm)
Length of substrate	L	31.6
Width of substrate	W	45.4
Width of feed	W_t	3
Thickness of substrate	t	1.6

3 Results and Discussion

3.1 First Iteration Structure

The simulation analysis of star shaped fractal antenna with three iterations has been studied by using HFSS software. The simulated results have been verified practically by using VNA. The geometry of the first iteration of the designed antenna (the star shape antenna with microstrip line feeding) is shown in Fig. 3.

The simulated results, i.e., return loss versus frequency curve, VSWR versus frequency curve and radiation pattern are demonstrated in Fig. 4a–c respectively. First iterated antenna resonates at frequency 4.78 GHz. Figure 4a shows that return loss of first iteration is -12.75 dB at resonant frequency 4.78 GHz. At this resonant frequency, VSWR is 1.59 as shown in Fig. 4b. VSWR value is less than 2 which essentially mean that the S_{11} plot (return loss) is more negative than -10 dB at this frequency. Figure 4c shows the radiation pattern at resonant frequency 4.78 GHz. The radiation pattern is almost omnidirectional with one null at 150°.

3.2 Second Iteration Structure

The geometry of the second iteration of designed antenna is shown in Fig. 5. The second iteration includes two-star structures, one same as first iteration and other added at the center of first iterative star.

Fig. 3 First iteration structure (top view)

(a) Return Loss

(b) VSWR

(c) Radiation Pattern
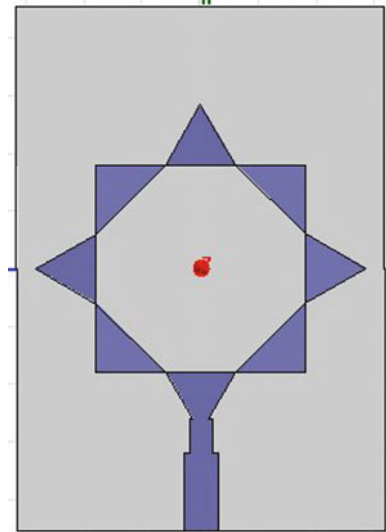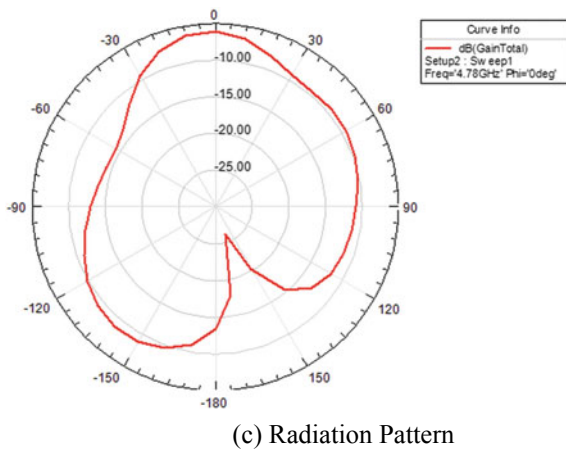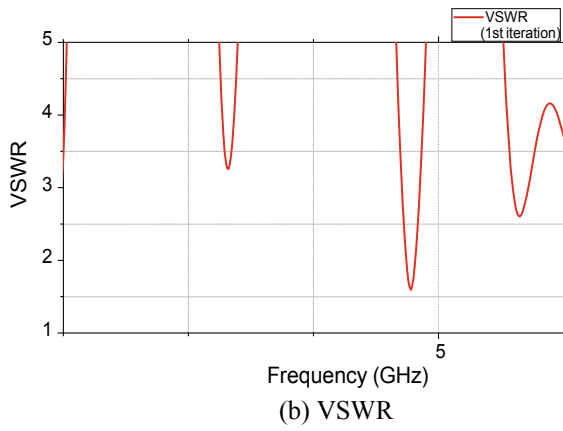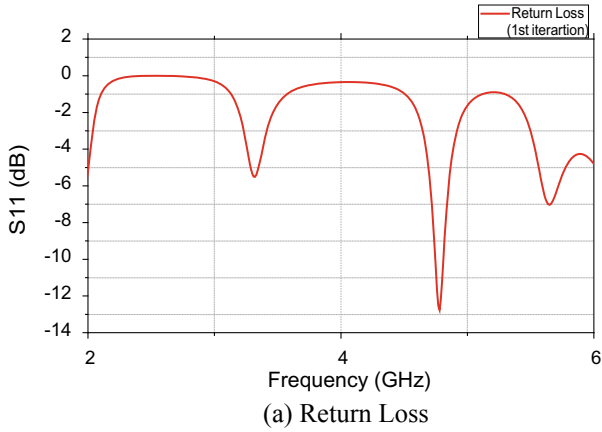
Fig. 4 Simulated results of first iteration structure (**a–c**)

Fig. 5 Second iteration
structure (top view)
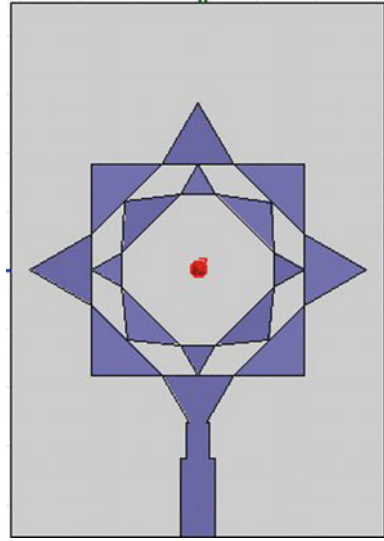
Second iterated design of antenna resonates at 2.27 GHz and 3.61 GHz frequency with a return loss of −10.35 dB and −21.89 dB respectively as shown in Fig. 6a. VSWR is in between 1 and 2 at these two resonant frequencies as shown in Fig. 6b. The radiation pattern for second iteration at 2.27 and 3.61 GHz frequencies is shown in Fig. 6c. There are two colored plots one is red and other is blue. The red color plot gives the radiation information at frequency 2.27 GHz and blue color plot gives the same at 3.61 GHz.

3.3 Third Iteration Structure

The top view of the third iteration geometry of proposed antenna is shown in Fig. 1 and bottom view of the same is shown in Fig. 7. This final (third) design of antenna resonates at 2.25 GHz, 3.63 GHz, and 5.16 GHz frequencies with return loss of − 12.55 dB, −32.69 dB, and −37.628 dB respectively as shown in Fig. 8a.

VSWR is in between 1 and 2 as shown in Fig. 8b. Radiation patterns at these three frequencies are shown in Fig. 8c.

The comparison of resonant frequency, return loss and VSWR of all three iterations are shown in Table 2. Figure 9 shows a comparison between simulated and measured values of return loss of proposed antenna.
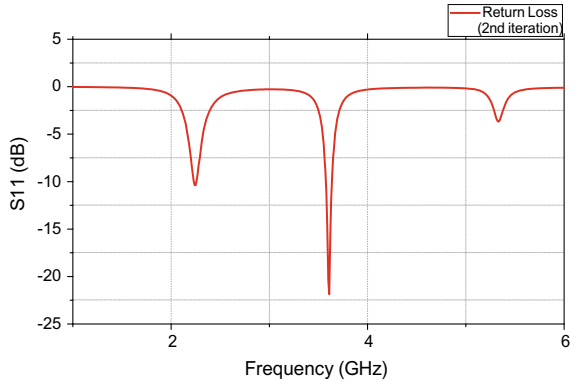
(a) Return Loss

(b) VSWR

(c) Radiation Pattern

Fig. 7 Bottom view of a third iteration structure
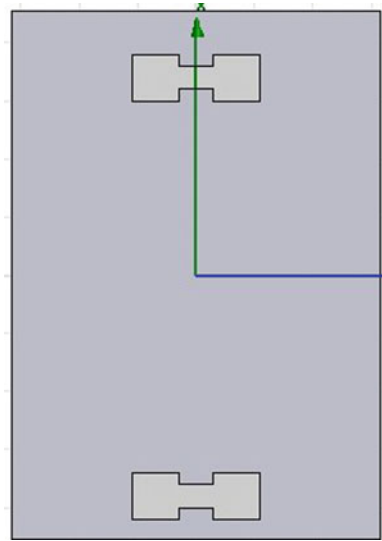
4 Conclusion

The design and radiation performance of a compact star shaped fractal antenna with the defected ground structure for multiband applications is accounted for in this article. The simulation analysis is completed by utilizing HFSS software and measurements are completed by using VNA. The proposed antenna is designed on low-cost and effectively accessible glass epoxy FR-4 substrate. The antenna utilizes a simple rectangular metal strip as a feed element to accomplish impedance matching for star shaped fractal geometry without any external matching circuitry. It is a triple band antenna with three resonant frequencies at 2.25, 3.63, and 5.16 GHz, also provides the stable far field characteristics. The simulated and measured results are compared, these are in great understanding. The antenna can be used in S-band (2–4 GHz) and C-band (4–8 GHz) for triple band applications.

Fig. 8 Simulated result of third iteration structure (**a–c**)
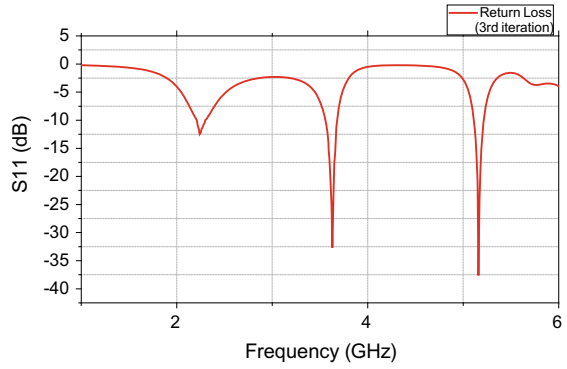
(a) Return Loss
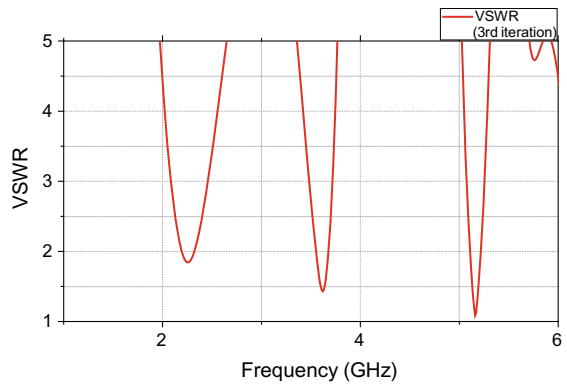
(b) VSWR
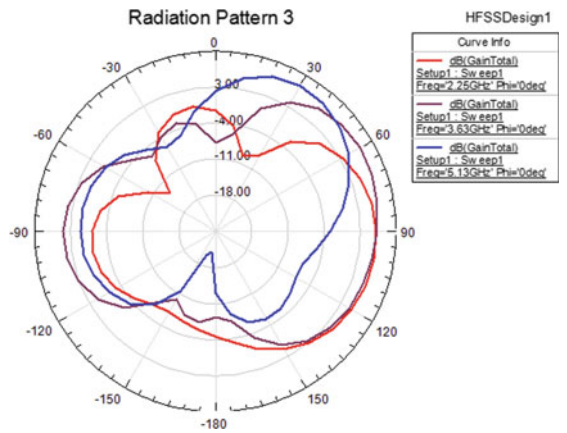
(c) Radiation Pattern

Table 2 Comparison of iterative structures

Iteration	Resonant frequency (GHz)	Return loss (dB)	VSWR
First iteration	4.78	−12.75	$1 \leq$ VSWR ≤ 2
Second iteration	2.27, 3.61	−10.35, −21.89	$1 \leq$ VSWR ≤ 2
Third iteration	2.25, 3.63, 5.16	−12.55, −32.69, −37.628	$1 \leq$ VSWR ≤ 2

Fig. 9 Comparison between simulated and measured return loss of the proposed antenna

References

1. Balanis, C. A.: Antenna Theory, Analysis and Design. Wiley (2005)
2. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: Microstrip Antenna Design Handbook. Artech House (2001)
3. James, J. R., Hall, P.S.: Handbook of Microstrip Antennas. Peter Peregrines (1989)
4. Yang, K.P., Wong, K.L.: Inclined-slot-coupled compact dual-frequency microstrip antenna with cross slot. Electron. Lett. **34**(4), 321–322 (1998)
5. Chen, H.D.: A dual-frequency rectangular microstrip antenna with a circular slot. Microw. Opt. Technol. Lett. **18**(2), 130–132 (1998)
6. Chen, W.S.: Single-feed dual-frequency rectangular microstrip antenna with square slot. Electron. Lett. **34**(3), 231–232 (1998)
7. Puente, C. et al.: An iterative model for fractal antenna: application to the Sierpinski gasket antenna. IEEE Trans. Antennas Propag. **48**(5), 713–719 (2000)
8. Baliarda, C.P., Romeu, J., Cardama, A.: The Koch monopole: a small fractal antenna. IEEE Trans. Antennas Propag. **48**, 1773–1781 (2000)
9. Hwang, K.C.: A modified Sierpinski fractal antenna for multiband application. IEEE Antenna Wirel. Propag. Lett. **6**, 357–360 (2007)
10. Cao, Y., Yuan, B., Wang, G.: A compact multiband open-ended slot antenna for mobile handsets. IEEE Antennas Wirel. Prop. Lett. **10**, 911–914 (2011)

Dual-Band Modified U-Shaped Slot Antenna with Defected Ground Structure for S-Band Applications

Anuradha Gupta and Som Pal Gangwar

Abstract In this article, design and investigation of a modified U-shaped slot antenna for dual-band applications is reported through simulation analysis and measured results. The proposed structure is designed using a single substrate. On one side of the substrate, first a rectangular patch is printed and then modified U-shaped slot is etched. On the other side of the substrate, a defected ground structure is created to enhance the gain. The antenna is simulated using HFSS software. To verify the simulated results, a prototype of the structure is fabricated using FR4 substrate. The proposed antenna exhibits dual-band characteristics. It has two resonant frequencies at 2.59 GHz and 3.76 GHz with return loss of 27.4 dB and 16.9 dB respectively. The difference in simulated and measured results are negligible. The proposed antenna can be used for dual-band applications in S-band (2–4 GHz).

1 Introduction

In this age of wireless communication, miniaturized, lightweight, low-cost, and easy to fabricate antennas are required. Microstrip patch antennas and slot antennas can be used to fulfill these requirements. However, both these antennas have a low value of gain and narrow bandwidth [1]. A slot antenna provides nearly omnidirectional pattern in at least one of its planes. The defected ground structure provides moderate value of gain. Therefore, slot antenna with defected ground structure (DGS) can provide omnidirectional radiation pattern with high gain in comparison to conventional microstrip antenna [2].

A. Gupta (✉) · S. P. Gangwar
Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur 228118, UP, India
e-mail: Anuradha4tech@gmail.com

S. P. Gangwar
e-mail: gangwar_sp@rediffmail.com

Currently, a lot of research is going on to get multiband features in antennas. Multiband antennas have two important advantages: (i) multiple frequency bands can be obtained in a single antenna; (ii) good signal to interference ratio can be achieved [3]. Many techniques have been used in the past to get dual-band characteristics in antenna. One of them is to use different types of slot like cross slot, a circular slot, a square slot, and an offset circular slot [4–7]. A dual-band antenna was obtained by employing an offset microstrip-fed line and a strip close to the radiating edges in the circular slot patch. But its size is large [8]. A helix shaped slot radiator was proposed for dual-band applications, but its size is also large [9]. A cavity-backed slot antenna was proposed for dual-band applications but its design is more complex [10].

This article presents the modified U-shaped slot antenna with DGS for dual-frequency applications [11]. Microstrip line has been used to feed the proposed antenna. This antenna structure is designed for two frequency bands, i.e., 2.52–2.63 GHz and 3.715–3.791 GHz. It provides the stable far field characteristics with the average gain of 3.0 dBi and 1.5 dBi in lower and upper frequency bands, respectively.

Even with a single feed point, the slot antenna can radiate significantly at a different range of frequencies. This concept is utilized in the design of proposed antenna. For better representation, present article is divided into five different sections: (i) introduction; (ii) antenna geometry; (iii) parametric analysis; (iv) results and discussion; and (v) conclusion.

2   Design of Antenna Geometry

Initially, a simple U-shaped slot structure is designed and then modified U-shaped structure has been fabricated. The optimized parameters of both antennas are obtained by parametric analysis. A comparative study has also been performed.

2.1   Simple U-Shaped Slot Antenna

The geometry of simple U-Shaped structure is shown in Fig. 1.

2.2   Modified U-Shaped Slot Antenna

The top and bottom view of the geometry of modified U-shaped structure is shown in Figs. 2 and 3 respectively [12]. The top and bottom view of the fabricated antenna are shown in Figs. 4 and 5 respectively. The structure is designed and fabricated on glass epoxy FR-4 substrate having thickness h, relative permittivity ε_r, and loss tangent tanδ. The optimized parameters of simple U-shaped slot antenna are shown

Fig. 1 Simple U-shaped microstrip patch antenna
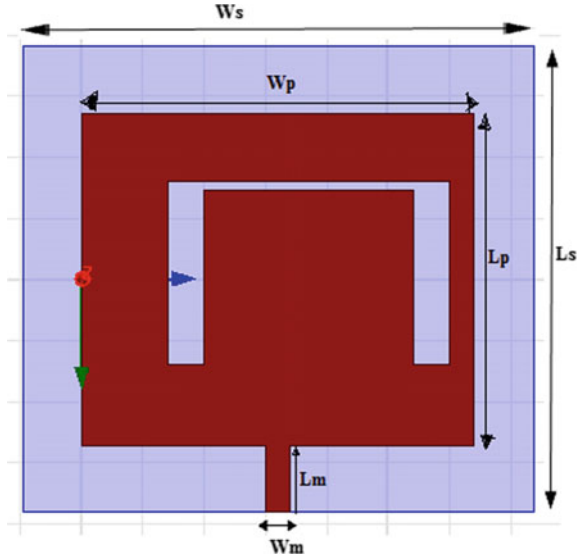
Fig. 2 Top view of modified U-shaped structure
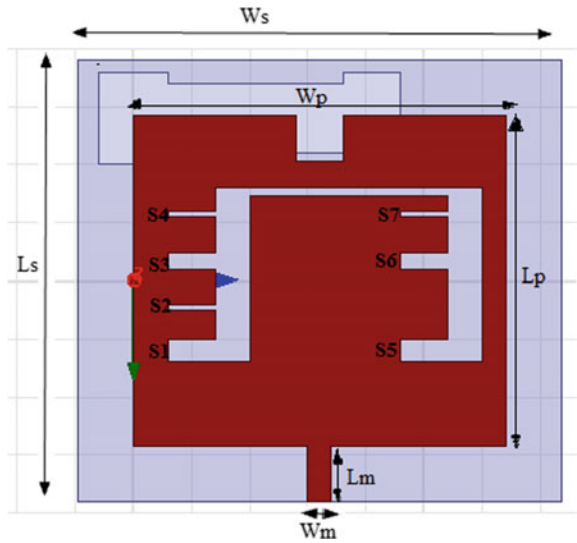
in Table 1. The optimized parameters of modified U-shaped slot antenna are shown in Table 2.

Fig. 3 Bottom view of
modified U-shaped structure
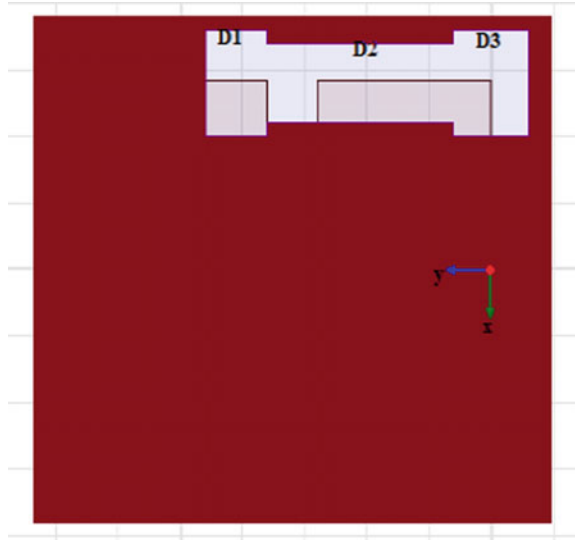
Fig. 4 Top view of the
fabricated (proposed)
antenna
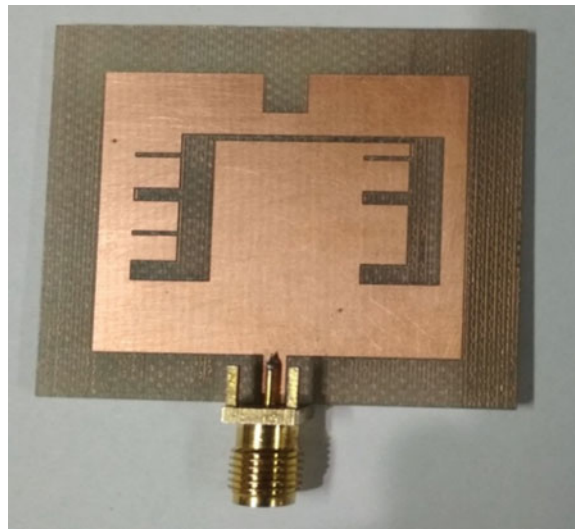
2.3 Fabricated Antenna

Here PCB Prototype Machine is used for design and fabrication of Patch Antenna
with high accuracy, speed, and ease.

Fig. 5 Bottom view of the fabricated (proposed) antenna
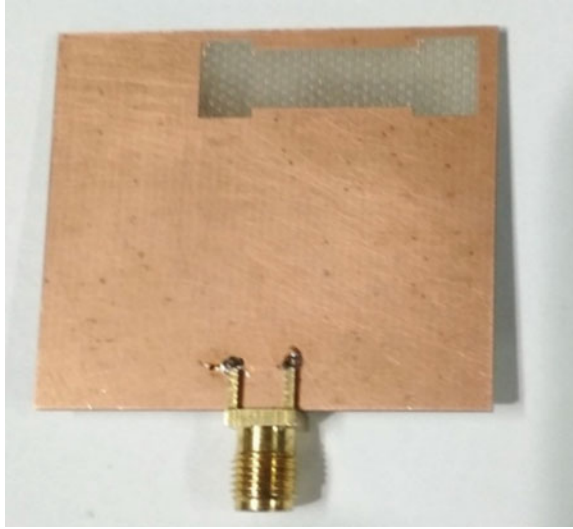
Table 1 Optimised parameters of simple U-shaped slot antenna and modified U-Shaped slot antenna

Antenna parameter	Dimensions
A. For simple U-shaped slot antenna	
Substrate material	FR4
Substrate width (W_s)	41.6 mm
Substrate length (L_s)	38.1 mm
Height of the substrate (h)	1.6 mm
Patch width (Wp)	32 mm
Patch length (L_p)	28.5 mm
Material of the patch	Copper
Feed line (L_m, W_m)	4.8 mm, 2 mm
L_1 (l_1, w_1)	15 mm, 3 mm
L_2 (l_2, w_2)	15 mm, 3 mm
W_2 (l, w)	0.7 mm, 17.18 mm
B. For modified U-shaped slot antenna	
S_1 (l, w)	2 mm, 4 mm
S_2 (l, w)	0.5 mm, 4 mm
S_3 (l, w)	1.5 mm, 4 mm
S_4 (l, w)	0.5 mm, 4 mm
S_5 (l, w)	2 mm, 4 mm

Table 1 (continued)

Antenna parameter	Dimensions
S_6 (l, w)	1.4 mm, 4 mm
S_7 (l, w)	0.5 mm, 4 mm
D_1 (l, w)	8 mm, 5 mm
D_2 (l, w)	6 mm, 17 mm
D_3 (l, w)	8 mm, 6 mm

Table 2 Table for the comparative study of simple U shape and modify U shape

Antenna name	Return loss	Resonant freq	Gain	VSWR
Simple U Shape microstrip patch antenna	−18.69 at 2.8 GHz, −10.5 at 3.9 GHz	2.8 GHz, 3.9 GHz	1.6 dB, 0.59 dB	1.2, 1.8
Modify U Shape microstrip patch antenna	−27.43 at 2.59 GHz, −16.90 at 3.76 GHz	2.59 GHz, 3.76 GHz	3.2 dB, 1.57 dB	1.088, 1.33

2.3.1 Parametric Analysis

A parametric analysis has been performed to get optimized dimensions. Return loss (S_{11}) versus frequency curve for different values of length of patch (L_p) is shown in Fig. 6. It is clearly seen that for $L_p = 28.5$ mm, better results are obtained in terms of return loss. Return loss should be greater than −10 dB at resonant frequencies.
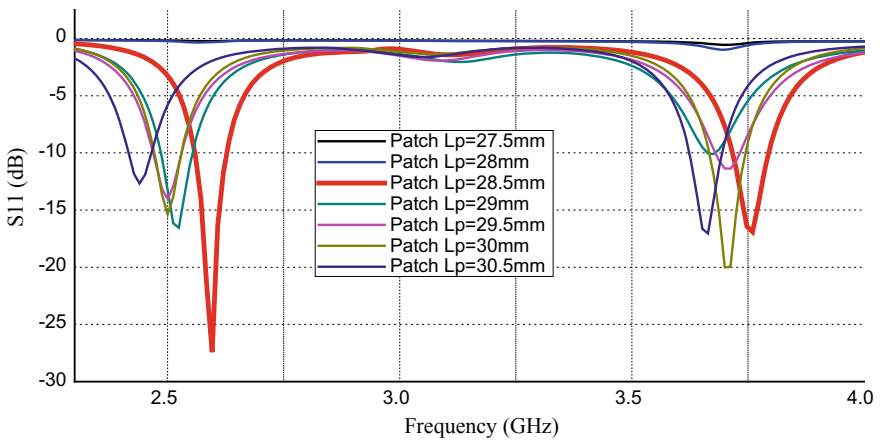
Fig. 6 Return loss with variation in L_p
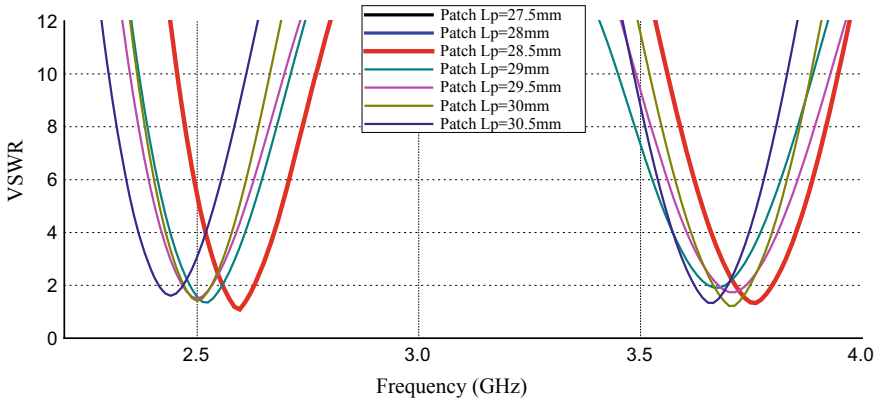
Fig. 7 VSWR with variation in length of patch (L_p)

Antenna structure with $L_p = 28.5$ mm has two resonant frequencies at 2.59 GHz and 3.76 GHz with return loss of -27.4 dB and -16.9 dB respectively.

The VSWR versus frequency curve for different values of length of feed (L_p) is shown in Fig. 7. VSWR should be in between 1 and 2. VSWR is 1.088 at 2.59 GHz and 1.33 at 3.76 GHz i.e. it is in the permissible limit which indicates that there is proper matching between antenna and microstrip feed line.

3 Results and Discussion

3.1 Comparison of Results of Simple U-Shaped and Modified U-Shaped Slot Antenna

Return loss versus frequency graph for simple U-shaped slot antenna and modified U-shaped slot antenna is shown in Fig. 8. For simple U shape structure, return loss is -18.6 dB at 2.8 GHz and -10.59 at 3.9 GHz and for modify U-shaped structure, return loss is -27.4 dB at 2.59 GHz and -16.9 dB at 3.76 GHz.

VSWR (voltage standing wave ratio) versus frequency graph for simple U-shaped slot antenna and modified U-shaped slot antenna is shown in Fig. 9. For simple U-shaped slot antenna, VSWR is 1.2 at 2.8 GHz and 1.8 at 3.9 GHz, and for modify U-shaped slot antenna, VSWR is 1.088 at 2.59 GHz and 1.33 at 3.76 GHz.

Gain versus frequency curve of simple U-shaped slot antenna and modified U-shaped slot antenna is shown in Fig. 10. The gain for simple U-shaped slot antenna is 1.67 dB at 2.8 GHz and for the proposed antenna, i.e., modified U-shaped slot antenna it is 3.07 dB at 2.59 GHz and 1.57 dB at 3.76 GHz. The gain is positive at both resonant frequencies. It is observed from Fig. 10 that the gain has been enhanced for modified U-shaped slot antenna with DGS.
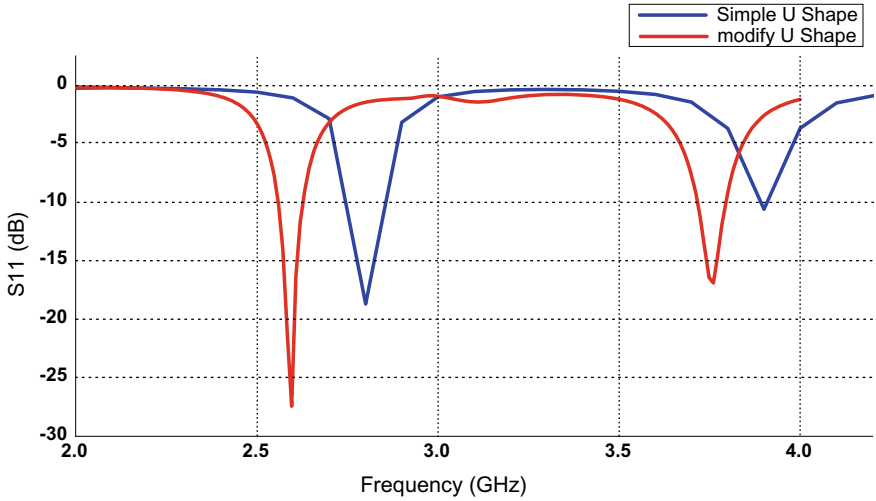
Fig. 8 Return loss versus frequency curve
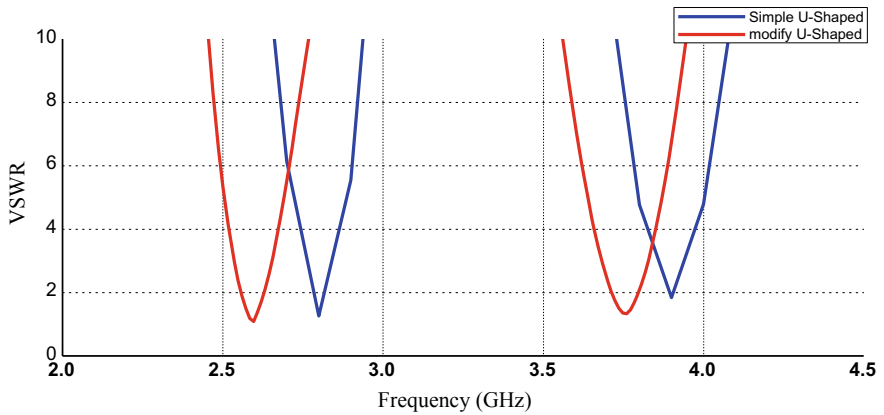
Fig. 9 VSWR versus frequency curve

The term radiation pattern refers to the angular dependence of the strength of the electric field of the antenna. Here the value of outer most of the circle is shown to theta value and inside the circle shows to radiation intensity value. Phi value basically shows to E and H plane, when phi is 0° then it shows to E plane and when phi is 90° then it shows to H plane. The radiation pattern is drawn at both resonant frequencies, i.e., at 2.59 GHz and 3.76 GHz frequency in E and H plane shows in Figs. 11 and 12. The radiation pattern is highly directional at both resonant frequencies.
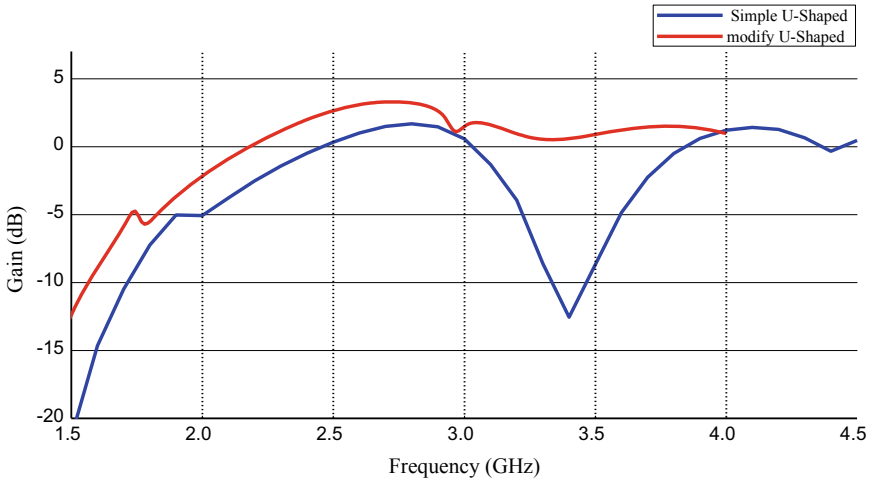
Fig. 10 Gain versus frequency curve

Fig. 11 Radiation pattern at
2.59 GHz
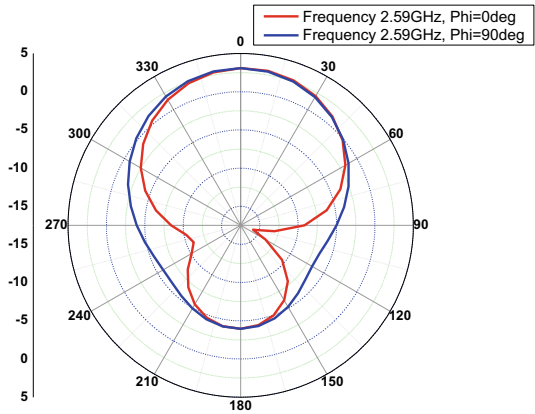
Radiation efficiency versus frequency curve for the proposed antenna is shown in
Fig. 13. Radiation efficiency is 52% at 2.59 GHz and 35% at 3.76 GHz which is not
very promising.
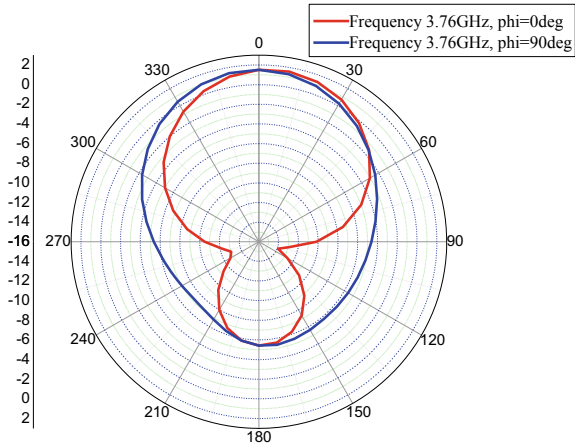
Fig. 12 Radiation pattern at 3.76 GHz
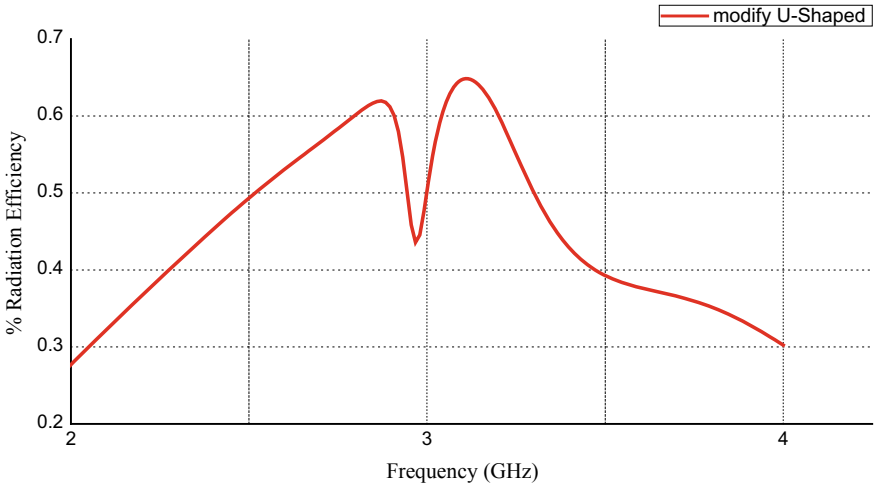
Fig. 13 Radiation efficiency versus frequency curve

3.1.1 Comparison of Simulated and Measured Results of Modified U-Shaped Slot Antenna

To verify the simulated results, the proposed antenna is fabricated and various parameters are measured. Figure 14 shows return loss (simulated and measured) versus frequency curve for the proposed antenna. The simulated and measured curves are identical in nature which confirms the validity of the proposed work.
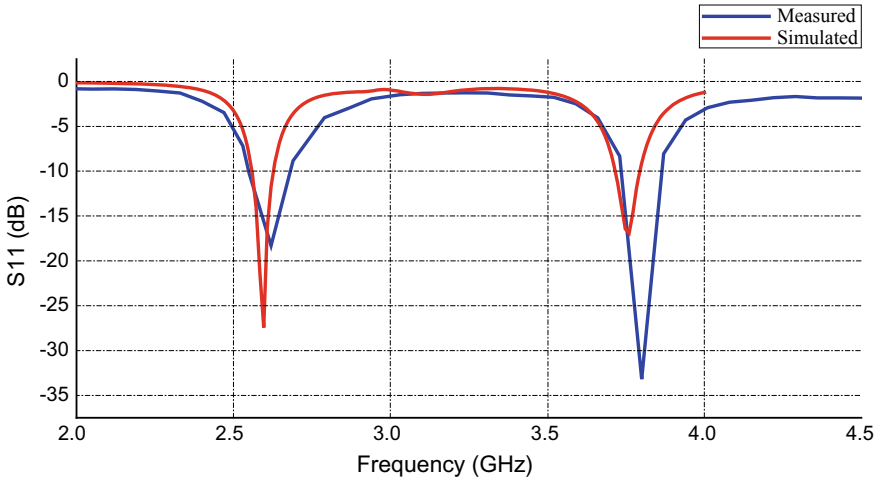
Fig. 14 Return loss (simulated and measured) versus frequency curve

For proposed antenna structure VSWR (simulated and measured) versus frequency curve is shown in Fig. 15. The simulated and measured values are close enough.
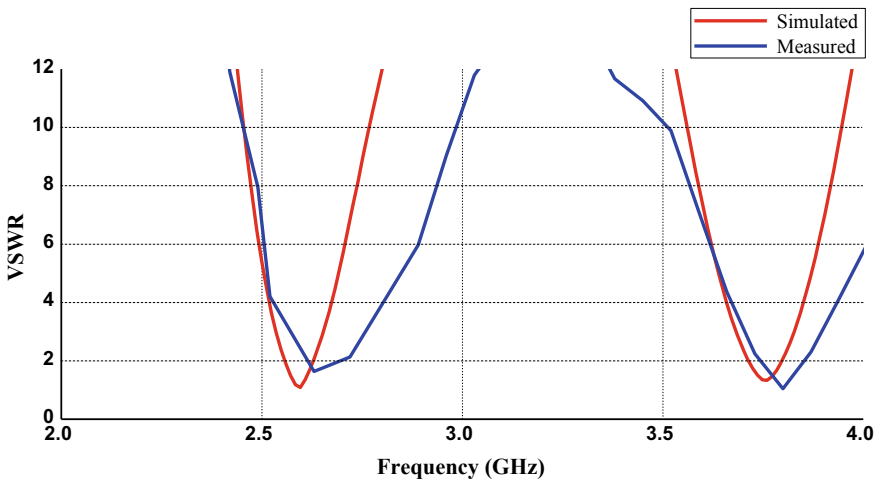
Fig. 15 VSWR (simulated and measured) versus frequency curve

4 Conclusion

The design and radiation performance of a compact modified U-shaped slot antenna for dual-band applications is reported in this article. The simulation analysis is carried out by using HFSS software and measurements are taken by using VNA. The proposed antenna is designed on low-cost and easily available glass epoxy FR-4 substrate. This antenna structure has dual-band characteristics in frequency bands 2.52–2.63 GHz and 3.715–3.791 GHz. The proposed antenna is a dual-band antenna with two resonant frequencies at 2.59 GHz and 3.76 GHz. It provides a stable far field characteristics. The antenna utilizes only a simple rectangular metal strip as a feed element to achieve impedance matching for modified U-shaped slot with DGS without any external matching circuitry. The simulated and measured results are compared and these are in good agreement. The antenna can be used for dual-band applications in frequency range 2–4 GHz (S-band).

References

1. Balanis, C.A.: Antenna Theory, Analysis and Design. Wiley & Sons (2005)
2. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: Microstrip Antenna Design Handbook. Artech House (2001)
3. James, J.R., Hall, P.S.: Handbook of Microstrip Antennas. Peter Peregrines (1989)
4. Yang, K.P., Wong, K.L.: Inclined-slot-coupled compact dual-frequency microstrip antenna with cross slot. Electron. Lett. **34**(4), 321–322 (1998)
5. Chen, H.D.: A dual-frequency rectangular microstrip antenna with a circular slot. Microw. Opt. Technol. Lett. **18**(2), 130–132 (1998)
6. Chen, W.S.: Single-feed dual-frequency rectangular microstrip antenna with square slot. Electron. Lett. **34**(3), 231–232 (1998)
7. Lu, J.H., Wong, W.L.: Compact dual-frequency circular microstrip antenna with an offset circular slot. Microw. Opt. Technol. Lett. **22**(4), 254–256 (1999)
8. Tiang, J.J., Islam, M.T., Misran, N., Mandeep, J.S.: Circular microstrip slot antenna for dual-frequency RFID application. Prog. Electromagn. Res. **120**, 499–512 (2011)
9. Yang, Y.H., Guo, J-L., Sun, B-H., Huang, Y-H.: Dual-band slot helix antenna for global positioning satellite applications. IEEE Trans. Antennas Propag. **64**(12), 5146–5152 (Dec. 2016)
10. Hung, C., Chiu, T.: Design of dual-band cavity-backed slot antenna loaded with spurline. IET Microw. Antennas Propag. **10**(9), 939–946 (2016)
11. Sofi, M.A., Saxena, J., Muzaffar, K.: Design and simulation of a novel dual band microstrip patch antenna with defected ground structure for Wlan/Wimax. Int. J. Electron. Electr. Eng. **7**(5), ISSN 0974-2174 (2014)
12. Sang-Hyuk, W., et al.: Wideband microstrip patch antenna with U-shaped parasitic elements. IEEE Trans. Antennas Propag. **55**, 1196–1199 (2007)

Localization of Sensor Nodes in WSN Using Area Between a Node and Two Beacons

Prateek Raj Gautam ⓘ**, Sunil Kumar** ⓘ**, Akshay Verma** ⓘ**, Tarique Rashid and Arvind Kumar** ⓘ

Abstract This paper presents a localization scheme for sensor nodes in Wireless Sensor Networks ($WSNs$). The scheme is capable of localizing the nodes in the region with the help of a few beacon nodes. The nodes don't have to exchange information among them, which saves time and reduce energy consumption at nodes. Nodes determine their location only based on the information received by them from three closest beacons nodes. The beacon nodes can vary transmit power P_t to increase or decrease coverage radius.

Keywords Sensor localization · Area between nodes · Wireless sensor network (WSN) · Area localization

1 Introduction

In the recent times, the use of Wireless Sensor Networks (WSNs) has touched many areas such as industrial application [7, 12], biomedical [2, 15], and household applications [16] because of their ease and low-cost [14, 20]. WSN have been used for IoT, gas leakage detection, fire detection, security monitoring, health care, and indoor localization [4, 5, 13, 17, 18, 21] WSN uses hundreds of randomly deployed sensors to gather information of various physical phenomenon like temperature, rain, humidity; but the information gathered by WSN is useful only if the location of sensing source is known, then only we can take necessary action in the desired region. Hence, localization is a very important task in WSN and each node must know its location, so that, the information conveyed by it can be used to take proper action.

P. R. Gautam (✉) · S. Kumar · A. Verma · T. Rashid · A. Kumar
Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India
e-mail: prateekrajgautam@gmail.com; rel1601@mnnit.ac.in
URL: http://mnnit.ac.in

1.1 Related Work

Recent localization is based on received signal strength indicator (RSSI) [6, 19] to estimate distance between receiver and transmitter, special antenna array and digital compass can be used for angle estimation, time of arrival (TOA), angle of arrival (AOA), time difference of arrival (TDOA), and triangulation to find the position of unknown nodes with the help of GPS enabled nodes [1, 3, 9–11].

1.1.1   Ripple Localization Algorithm (RLA) [8]

In [8] (RLA), beacon nodes can vary their transmit power so that the radius of nodes receiving beacons is changed. The beacon signal also contains the value of current radius (R_i), increment in radius (d_r), minimum radius (r_{min}), and maximum radius (R_{max}). A typical beacon consists of six parameters

$$[R_i, R_{min}, R_{max}, d_r, x, y]$$

(x, y) represents the location coordinate of beacon node transmitting. The beacon transmission starts with $R_i = R_{min}$ and it is increased up to R_{max} in steps of d_r. As shown in Fig. 1, any unknown node that receives a beacon signal can estimates its radial distance $R_{estimated}$ can be given by (1) as

$$R_{estimated} = \left(R_i - \frac{d_r}{2} \right). \tag{1}$$

Maximum error in radial estimation is given by (2) as

$$\varepsilon_r \leq \pm \frac{d_r}{2}. \tag{2}$$

1.2 Novelty

This paper presents a localization process that is capable of localizing sensor nodes without increasing the cost of the node or the energy consumption at the nodes. Localization is performed with the help of three beacon nodes that know their location with the help of GPS or manual deployment. Minimum of three nodes are required for the accurate location of an unknown node but here we have not used centroid calculation or applied linear algebra for estimation. In the proposed scheme any node select two beacon nodes and area between these three nodes is calculated based on our calculation two possible node points are determined, then any third beacon is selected and distance of the node from the beacon is compared with the distance
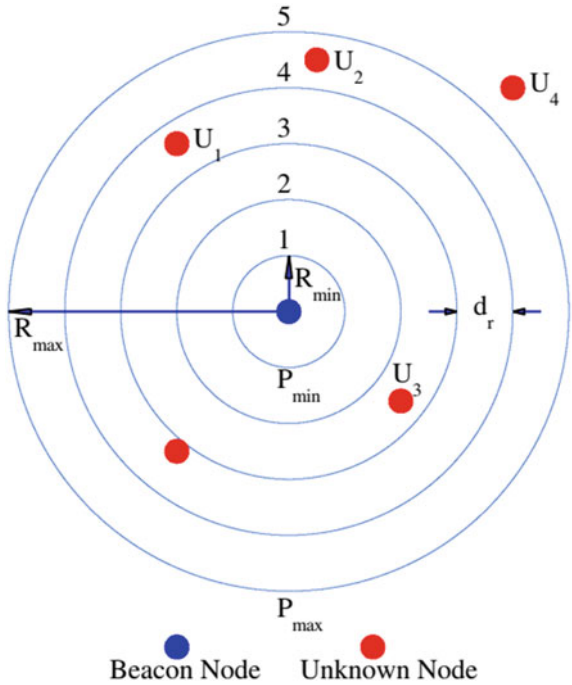
of two determined node points, whose distance matches is selected as the estimated
node point. Nodes don't have to transmit so its energy consumption is reduced as
compared to other localization processes and no special hardware addition is required
at the node thereby keeping cost and size of node same.

2 Proposed Scheme: Localization Using Area Between a Node and Two Beacons

In the proposed scheme, assuming all nodes lie on same XY plane, beacon nodes
transmit beacon packets as in RLA [8], any node receiving beacons from at least three
beacon nodes can estimate its location. Beacon nodes know their location either by
GPS module (requires GPS module on nodes) or by manual placement and they may
also serve as regular nodes for sensing and routing. For our ease, the proposed scheme
will be refereed as AREA localization or simply AREA. First node select two closest
nodes, as shown in Fig. 2, node n1 select two beacon nodes b1 and b2. Node n1 can
estimate distance b1n1, b2n1 and also the distance between two beacons b1b2 from
the coordinate transmitted by them. The line connection b1b2 will be referred as
baseline from here. Now if we drop a perpendicular from n1 on this baseline (b1b2),
the perpendicular meets at point N. We can estimate the coordinate of n1 if we know

n1 B

a=b1n1
b=n1b2
c=b1b2
d=n1N
e=n1b3

b1 α
A

β b2
C

N

n2 D

b3 E
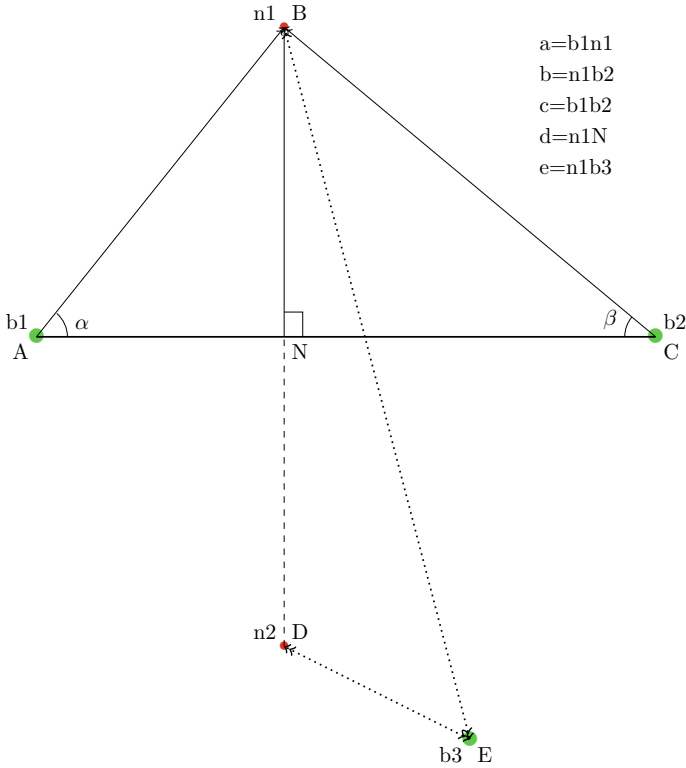
Fig. 2 Schematic diagram of localization using the area between a node and two beacons

the distance n1-N, then after we can determine angle α and β. For this purpose area of triangle, n1b1b2 is calculated as

$$area = \frac{1}{2}base \times altitude = \frac{1}{2}b1b2 \times n1N. \tag{3}$$

The area of triangle can also be estimated using heron's formula, if all the sides of the triangle are known as

$$\sqrt{s(s-a)(s-b)(s-c)}. \tag{4}$$

where a, b, c represents sides of triangle and $s = \frac{a+b+c}{2}$. In Fig. 2 a=b1n2, b=n1b2, and c=b1b2. Let n1N=d so from (3) and (4)

$$n1N = \frac{2 \times \sqrt{s(s-a)(s-b)(s-c)}}{b1b2}. \tag{5}$$

$$d = \frac{2 \times \sqrt{s(s-a)(s-b)(s-c)}}{c}. \tag{6}$$

Now α and β can be estimated as

$$\alpha = \sin^{-1}\left(\frac{d}{a}\right). \tag{7}$$

$$\beta = \sin^{-1}\left(\frac{d}{b}\right). \tag{8}$$

and b1N and b2N and be determined as

$$b1N = a\cos\alpha. \tag{9}$$

$$b2N = b\cos\beta. \tag{10}$$

Now all the variables in this triangle b1n1b2 are known hence coordinate of n1 can be determined using vector algebra by (11), where (x_A, y_A) represents coordinate of b1 at point A, $\overrightarrow{AN}$ is vector in direction b1 to b2 with magnitude b1N and $\overrightarrow{NB}$ is vector normal to direction of b1b2 with magnitude n1N. But this normal vector can also be in direction $\overrightarrow{ND}$ leading us to point D.

$$(x_A, y_A) + \overrightarrow{AN} + \overrightarrow{NB} \tag{11}$$

Solving this we end up two possibilities in node n1 one at point B and another at point D located equally opposite to the baseline.

So now, we select another beacon node, b3 in the region and distance of node from b3 is compared with two node points say n1 and n2, the point whose distance matches with the distance of beacon received is selected as estimated node point, means $|\overrightarrow{EB}|$ and $|\overrightarrow{ED}|$ is compared with distance estimated from node from beacon b3 located at E to select one point as estimated location.

3   Results and Performance Analysis

A 100m×100m is selected and 100 nodes are deployed randomly, Minimum of three beacon nodes are manually placed in the region at the location given in Table 1. However, optimum number of beacon nodes and their placement has not been studied in this paper, this will be extended in future work. Simulation parameters are shown in Table 1.

Table 1 Simulation parameters

Parameter	Symbols	Value
Area (m^2)	A	100×100
Total nodes	n	100
Ripple step	d_r	10
Starting ripple radius	R_{min}	d_r
Max ripple radius	R_{max}	100
Total number of beacon nodes		10%
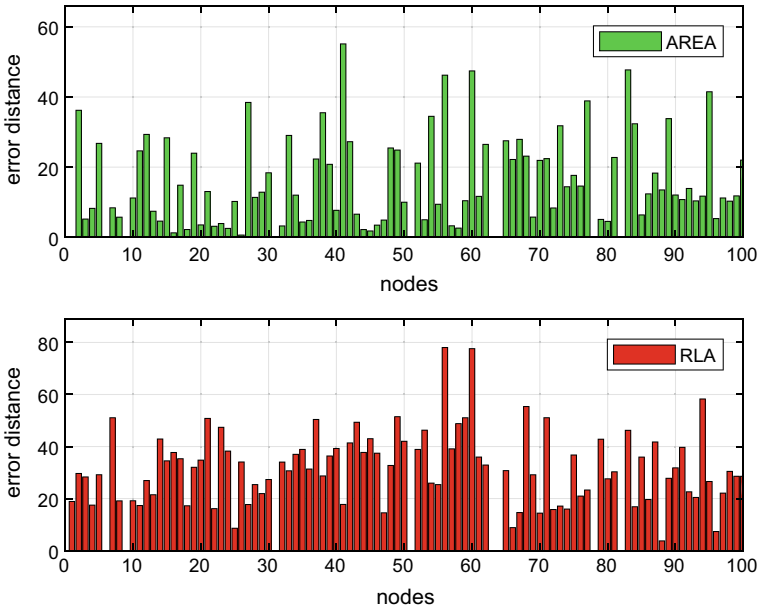
Fig. 3 Error distance in estimation via AREA and RLA

3.1 Simulation Results and Discussion

Figure 3 shows the error distance of all the nodes in case of AREA localization and RLA.

Figure 4 shows a comparison between CDF of error distance between RLA and AREA localization. Although the maximum error is approximately the same error via AREA localization is less in comparison to RLA. Hence, AREA localization gives better accuracy in comparison to RLA.
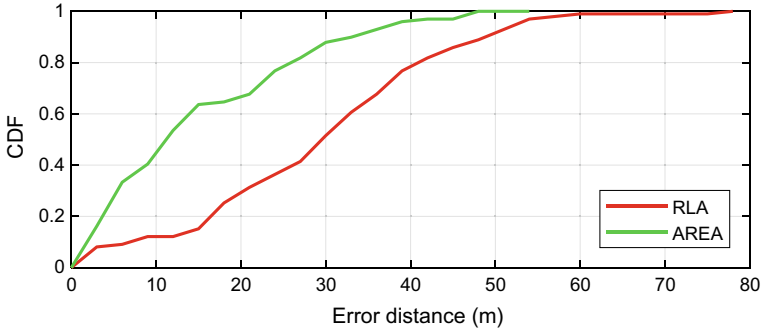
Fig. 4 Comparison via CDF of error distance

4 Conclusion and Future Scope

The simulation results verify the proposed scheme as the localization process have been performed with accuracy better than RLA. The maximum error distance observed in both the schemes is almost same but AREA localization performs much better than RLA. The number of beacons and ripple transmission is exactly the same as used in RLA [8] only beacon processing algorithm has been changed.

Future work will be extended in direction to reduce the number of beacon nodes and their optimum placement so that minimum beacon nodes can localize the desired region. Extended work will be published soon.

References

1. Chelouah, L., Semchedine, F., Bouallouche-Medjkoune, L.: Localization protocols for mobile wireless sensor networks: a survey. Comput. Electr. Eng. (2017)
2. Chiang, P.Y., Chao, P.C.P., Tarng, D.C., Yang, C.Y.: A novel wireless photoplethysmography blood-flow volume sensor for assessing arteriovenous fistula of hemodialysis patients. IEEE Trans. Ind. Electron. **64**(12), 9626–9635 (2017). https://doi.org/10.1109/TIE.2017.2701780
3. Chowdhury, T.J., Elkin, C., Devabhaktuni, V., Rawat, D.B., Oluoch, J.: Advances on localization techniques for wireless sensor networks: a survey. Comput. Netw. **110**, 284–305 (2016). https://doi.org/10.1016/j.comnet.2016.10.006
4. Chraim, F., Erol, Y.B., Pister, K.: Wireless gas leak detection and localization. IEEE Trans. Ind. Inform. **12**(2), 768–779 (2016). https://doi.org/10.1109/TII.2015.2397879
5. Chraim, F., Erol, Y.B., Pister, K.: Wireless gas leak detection and localization. IEEE Trans. Ind. Inform. **12**(2), 768–779 (2016)
6. Cui, W., Zhang, L., Li, B., Guo, J., Meng, W., Wang, H., Xie, L.: Received-signal-strength based indoor positioning using random vector functional link network. IEEE Trans. Ind. Inform. **PP**(99), 1–1 (2017). https://doi.org/10.1109/TII.2017.2760915
7. Derr, K., Manic, M.: Wireless sensor networks-node localization for various industry problems. IEEE Trans. Ind. Inform. **11**(3), 752–762 (2015). https://doi.org/10.1109/TII.2015.2396007

8. Farooq-I-Azam, M., Ni, Q., Ansari, E.A.: Intelligent energy efficient localization using variable range beacons in industrial wireless sensor networks. IEEE Trans. Ind. Inform. **12**(6), 2206–2216 (2016). https://doi.org/10.1109/TII.2016.2606084

9. Han, G., Jiang, J., Zhang, C., Duong, T.Q., Guizani, M., Karagiannidis, G.K.: A survey on mobile anchor node assisted localization in wireless sensor networks. IEEE Commun. Surv. Tutor. **18**(3), 2220–2243 (2016). https://doi.org/10.1109/COMST.2016.2544751

10. Han, G., Xu, H., Duong, T.Q., Jiang, J., Hara, T.: Localization algorithms of wireless sensor networks: a survey. Telecommun. Syst. **52**(4), 2419–2436 (2013). https://doi.org/10.1007/s11235-011-9564-7

11. Haute, T.V., Verbeke, B., Poorter, E.D., Moerman, I.: Optimizing time-of-arrival localization solutions for challenging industrial environments. IEEE Trans. Ind. Inform. **13**(3), 1430–1439 (2017). https://doi.org/10.1109/TII.2016.2550531

12. Iqbal, Z., Kim, K., Lee, H.N.: A cooperative wireless sensor network for indoor industrial monitoring. IEEE Trans. Ind. Inform. **13**(2), 482–491 (2017). https://doi.org/10.1109/TII.2016.2613504

13. Pak, J.M., Ahn, C.K., Shmaliy, Y.S., Lim, M.T.: Improving reliability of particle filter-based localization in wireless sensor networks via hybrid particle/fir filtering. IEEE Trans. Ind. Inform. **11**(5), 1089–1098 (2015)

14. Sheng, Z., Mahapatra, C., Zhu, C., Leung, V.C.M.: Recent advances in industrial wireless sensor networks toward efficient management in IoT. IEEE Access **3**, 622–637 (2015). https://doi.org/10.1109/ACCESS.2015.2435000

15. Wan, L., Han, G., Shu, L., Chan, S., Zhu, T.: The application of DOA estimation approach in patient tracking systems with high patient density. IEEE Trans. Ind. Inform. **12**(6), 2353–2364 (2016). https://doi.org/10.1109/TII.2016.2569416

16. Wang, J.C., Lin, C.H., Siahaan, E., Chen, B.W., Chuang, H.L.: Mixed sound event verification on wireless sensor network for home automation. IEEE Trans. Ind. Inform. **10**(1), 803–812 (2014). https://doi.org/10.1109/TII.2013.2262283

17. Wu, D., Chatzigeorgiou, D., Youcef-Toumi, K., Ben-Mansour, R.: Node localization in robotic sensor networks for pipeline inspection. IEEE Trans. Ind. Inform. **12**(2), 809–819 (2016). https://doi.org/10.1109/TII.2015.2469636

18. Wu, D., Chatzigeorgiou, D., Youcef-Toumi, K., Ben-Mansour, R.: Node localization in robotic sensor networks for pipeline inspection. IEEE Trans. Ind. Inform. **12**(2), 809–819 (2016)

19. Yaghoubi, F., Abbasfar, A.A., Maham, B.: Energy-efficient rssi-based localization for wireless sensor networks. IEEE Commun. Lett. **18**(6), 973–976 (2014). https://doi.org/10.1109/LCOMM.2014.2320939

20. Yan, J., Zhou, M., Ding, Z.: Recent advances in energy-efficient routing protocols for wireless sensor networks: a review. IEEE Access **4**, 5673–5686 (2016). https://doi.org/10.1109/ACCESS.2016.2598719

21. Yang, G., Xie, L., Mäntysalo, M., Zhou, X., Pang, Z., Da Xu, L., Kao-Walter, S., Chen, Q., Zheng, L.R.: A health-iot platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box. IEEE Trans. Ind. Inform. **10**(4), 2180–2191 (2014)

A Compact Rectangular Patch Antenna with Defected Ground Structure for Multiband Applications

Yogesh Kumar Mishra and Som Pal Gangwar

Abstract This paper investigates the design and radiation performance of rectangular patch antenna using a Defective Ground Structure (DGS) for multiband applications. The proposed structure is designed using FR-4 substrate. On one side of the substrate, a rectangular patch is printed whereas, on the other side of the substrate, dumbbell-shaped DGS is created. The antenna is simulated using HFSS software. To verify the simulated results, a prototype of the structure is fabricated, and measurements are taken. The multiband operation is achieved by introducing DGS of dumbbell shape. The proposed design has five resonant frequencies at 2.638, 3.84, 6.72, 10.67, and 13.78 GHz. The proposed antenna can be used for multiband applications.

Keywords Multiband antenna · Defective ground structure · FR-4 epoxy · Patch antenna

1   Introduction

The microstrip antennas have many unique and attractive properties which include low profile, lightweight, compact, ease of fabrication and integration with solid-state devices. These features attracted researchers to adapt the design of these antennas for various wireless applications where size and weight are a constraint. These properties contribute to the successful use of microstrip antennas not only in the military, such as aircraft, missiles, and spacecraft, but also in the commercial areas, such as mobile satellite system, terrestrial cellular communications, broadcast satellite system, and global positioning system [1–3].

Y. K. Mishra (✉) · S. P. Gangwar
Department of Electronics Engineering, Kamla Nehru Institute of Technology, Sultanpur 228118, UP, India
e-mail: Ykmishra1991@gmail.com

S. P. Gangwar
e-mail: gangwar_sp@rediffmail.com

Over the last two decades, mobile wireless service is increasing. Due to the increase of wireless standards, multiband antenna designs have become very important for wireless communications to fulfill the requirement of modern handheld communication devices. Multiband antennas can integrate more than one communication standard in a single compact module. Such an antenna can work at several frequencies simultaneously [4–10]. DGS is one of the methods which is used to enhance the bandwidth of the patch antennas and also reduces its size [11–14]. This paper presents the design of multiband rectangular patch antenna using a defective ground structure. Simulated results of the multiband antenna have been compiled. The proposed antenna is fabricated,and measurements are taken using Vector Network Analyzer (VNA). For better representation, present article is divided into five different sections: (i) introduction): (ii) design of antenna geometry; (iii) design of DGS; (iv) results and discussion; and (v) conclusion.

2 Design of Patch Antenna Geometry

A typical rectangular patch antenna is designed using FR-4 substrate having a dielectric constant of 4.3 and height of 1.6 mm. The antenna has been designed using standard design equation [1].

The effective dielectric constant is given by [1]:

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

Where

ϵ_r Dielectric constant of substrate
h Height of dielectric substrate
W Width of the patch

The dimensions of the patch along its length have to be extended on each end by a distance ΔL due to fringing, which is given empirically [1] as

$$\Delta L = 0.412h\frac{\left(\epsilon_{reff} + 0.3\right)\left(\frac{W}{h} + 0.264\right)}{\left(\epsilon_{reff} - 0.258\right)\left(\frac{W}{h} + 0.8\right)} \tag{2}$$

The effective length of the patch L_{eff} now becomes

$$L_{eff} = L + 2\Delta L \tag{3}$$

For a given resonant frequency f_0 the effective length is given by [1] as

$$Leff = \frac{c}{2f_o\sqrt{\epsilon_{reff}}} \tag{4}$$

For a rectangular microstrip patch antenna, the resonance frequency for any TM_{mn} mode is given by [1]

$$f_0 = \frac{c}{2\sqrt{\epsilon_{reff}}} \left[\left(\frac{m}{L}\right)^2 + \left(\frac{n}{W}\right)^2 \right]^{\frac{1}{2}} \qquad (5)$$

where m and n are modes along L and W, respectively.

For efficient radiation, the width W is given by [1]

$$W = \frac{c}{2 f_o \sqrt{\frac{\epsilon_r + 1}{2}}} \qquad (6)$$

Probe feeding often referred to as a coaxial feeding is used to feed the antenna; Fig. 1 shows the top and bottom views of the proposed antenna designed on HFSS software. It shows two schematic views of the designed antenna. Table 1 shows all the dimensions of the proposed antenna.

Fig. 1 Geometry of the proposed antenna

(a) Top view　　　　　　(b) Bottom view
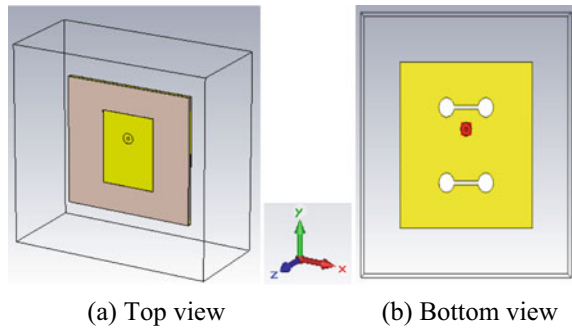
Table 1 Design parameters

Dimensions	Value (mm)
The radius of DGS circles	2.97
Dumbbell width	1.1
Length of the ground plane	50
Length of substrate	50
Length of patch	28.6
The width of the ground plane	50
Width of substrate	50
Width of patch	20.8
Feed location offset	4.77

3 Design of Multiband Antenna Using DGS

The defective ground structure is a method in which the ground of patch is made defected by different etching shapes of slots (of a specific dimension). In this method, the ground plane metal of a microstrip (or stripline, or coplanar waveguide) circuit is intentionally modified to enhance its performance. The DGS helps in shifting the resonant frequency to get the desired frequency [11–14] (Table 2).

3.1 Patch Antenna with Dumbbell-Shaped DGS

Initially, four circular slots are etched on the ground plane to see the effect on the resonant frequency and return loss. Then to introduce more resonant frequencies slots are also etched between the upper two circles and lower two circles. Finally, a dumbbell shape comes into existence. Figure 2 shows the top view and bottom view of the fabricated structure. The antenna has been fabricated using standard photolithography procedure.

Table 2 Comparison of simulated and measured return loss

S. no.	Simulated		Measured	
	Frequency (GHz)	Return loss (dB)	Frequency (GHz)	Return loss (dB)
1	2.638	−26.40	2.68	−12.514
2	3.84	−15.94	3.94	−18.68
3	6.72	−14.36	7.02	−15.5
4	10.67	−29.14	11.01	−24.83
5	13.78	−30.94	14.3	−22.5

Fig. 2 A fabricated antenna (proposed structure)

(a) Top view   (b) Bottom view
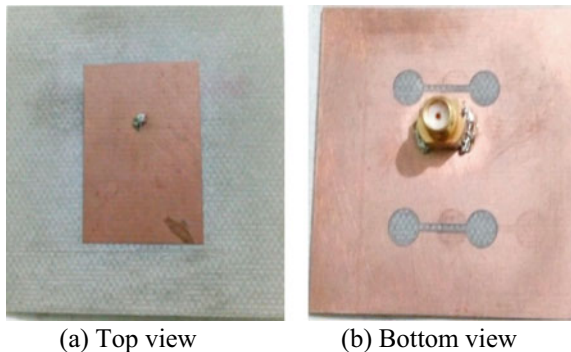
4 Results and Discussion

4.1 Return Loss

Figure 3 represents the comparison graph between the measured and simulated return loss (S_{11}) of the proposed antenna. It is observed from the figure that the proposed antenna provides multiband operation at resonance frequencies 2.638, 3.842, 6.72, 10.67, and 13.78 GHz.

It is observed from this figure that both the curves are closely matching. There is little variation which may be due to several factors including fabrication error.

4.2 VSWR

Figure 4 represents the comparison graph between measured and simulated VSWR of the proposed antenna. It is observed that both the results are closely matched.

Table 3 compares the simulated and measured VSWR values for the proposed fabricated antenna.

It is observed from the results that the measured VSWR values for the resonant frequencies are close to unity, which is much below the standard value of 2. So, it is concluded that the proposed antenna is well suitable for the multiband applications in the various areas.

Fig. 3 Comparison of measured and simulated return loss
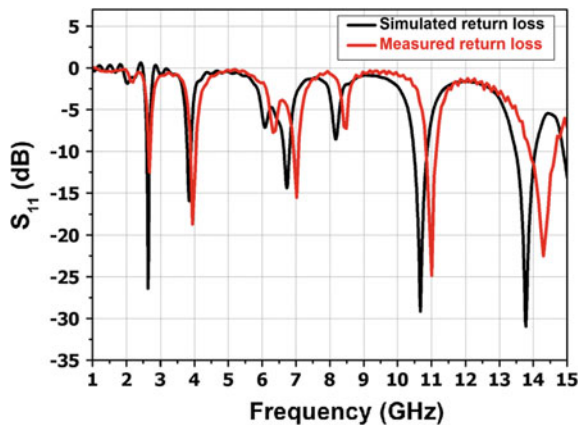
Fig. 4 Comparison of
measured and simulated
VSWR
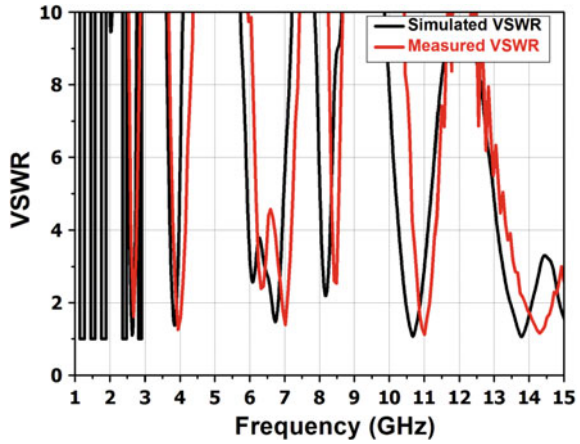
Table 3 Comparison of measured and simulated VSWR

S. no.	Simulated		Measured	
	Frequency (GHz)	VSWR	Frequency (GHz)	VSWR
1	2.638	1.1	2.68	1.62
2	3.84	1.3	3.94	1.26
3	6.72	1.4	7.02	1.4
4	10.67	1.07	11.01	1.12
5	13.78	1.05	14.3	1.16

4.3   Radiation Pattern

Figure 5 shows far-field radiation pattern of the proposed antenna at the frequency
of 2.638 GHz. The radiation pattern is bidirectional.

Figure 6 shows far-field radiation pattern of the proposed antenna at the frequency
of 3.84 GHz. It is shown that at 3.84 GHz, the main lobe occurs at 180°. The angular
width is 80.8° and side lobe level is −10.3 dB.

Figure 7 shows far-field radiation pattern of the proposed antenna at the frequency
of 6.73 GHz. It is shown that at 6.73 GHz, the main lobe occurs at 45°. The angular
width is 67.3° and side lobe level is −6.6 dB.

Figure 8 shows far-field radiation pattern of the proposed antenna at the frequency
of 10.674 GHz. It is shown that at 10.674 GHz, the main lobe occurs at 180°. The
angular width is 54.4° and side lobe level is −6.8 dB.

Figure 9 shows far-field radiation pattern of the proposed antenna at the frequency
of 13.782 GHz. It is shown here that the main lobe occurs at 31°. The angular width
is 35°, and the side lobe level is −7.6 dB.

Farfield Directivity Abs (Phi=0)

farfield (f=2.638) [1]

Frequency = 2.638
Main lobe magnitude = -6.6 dBi
Main lobe direction = 180.0 deg.
Angular width (3 dB) = 111.9 deg.

Theta / Degree vs. dBi
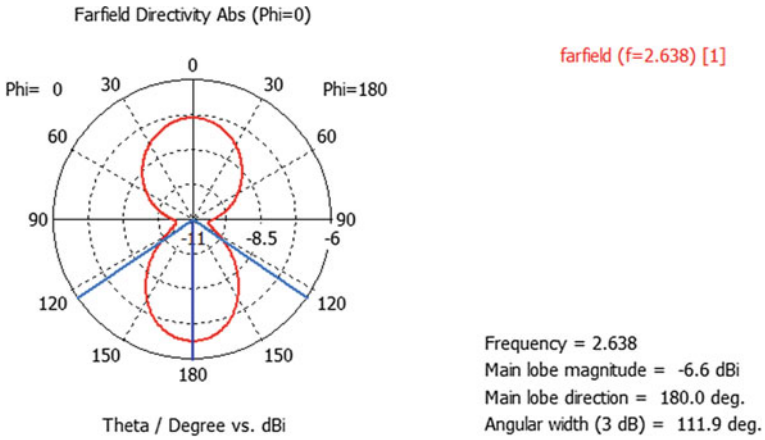
Fig. 5 Radiation pattern at 2.638 GHz

Farfield Directivity Abs (Phi=0)

farfield (f=3.84) [1]

Frequency = 3.84
Main lobe magnitude = 6.9 dBi
Main lobe direction = 180.0 deg.
Angular width (3 dB) = 80.8 deg.
Side lobe level = -10.3 dB

Theta / Degree vs. dBi
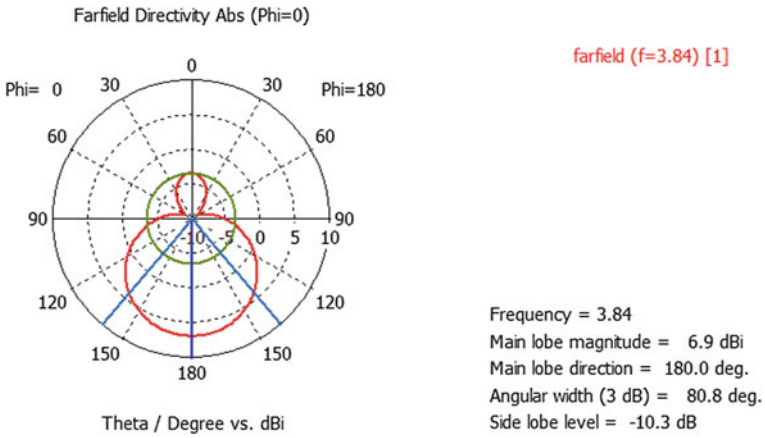
Fig. 6 Radiation pattern at 3.84 GHz

4.4 Gain

Figure 10 shows the gain of the proposed antenna. It is observed from the figure that the gain of the antenna is varying between 1.4 and 6 dB.

Farfield Directivity Abs (Phi=0)

farfield (f=6.73) [1]

Frequency = 6.73
Main lobe magnitude =   3.0 dBi
Main lobe direction =   45.0 deg.
Angular width (3 dB) =   67.3 deg.
Side lobe level =   -6.6 dB

Theta / Degree vs. dBi
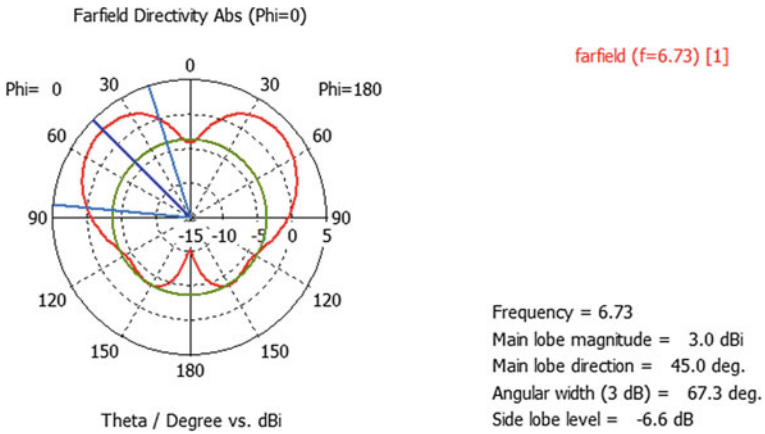
Fig. 7   Radiation pattern at 6.73 GHz

Farfield Directivity Abs (Phi=0)

farfield (f=10.674) [1]

Frequency = 10.674
Main lobe magnitude =   9.1 dBi
Main lobe direction =   180.0 deg.
Angular width (3 dB) =   54.4 deg.
Side lobe level =   -6.8 dB

Theta / Degree vs. dBi
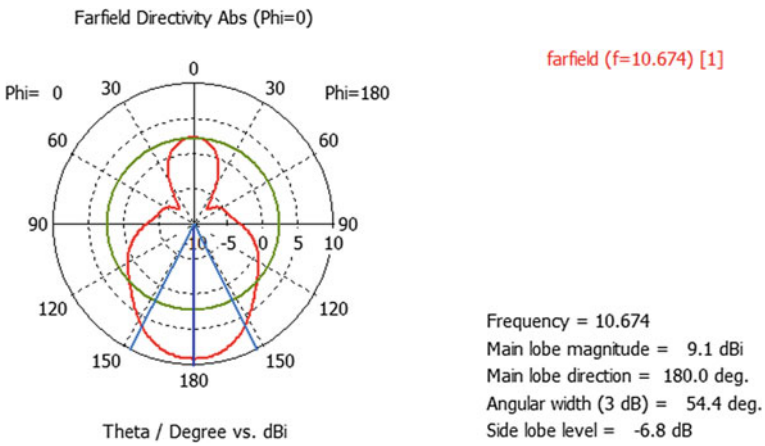
Fig. 8   Radiation pattern at 10.674 GHz

4.5 Efficiency

Figure 11 shows the radiation efficiency and total efficiency of the proposed antenna. It is observed that the radiation efficiency is varying between 33 and 55%. The efficiency is at its lowest at 2.638 GHz.

Farfield Directivity Abs (Phi=0)

farfield (f=13.782) [1]

Frequency = 13.782
Main lobe magnitude = 3.7 dBi
Main lobe direction = 31.0 deg.
Angular width (3 dB) = 35.0 deg.
Side lobe level = -7.6 dB

Theta / Degree vs. dBi
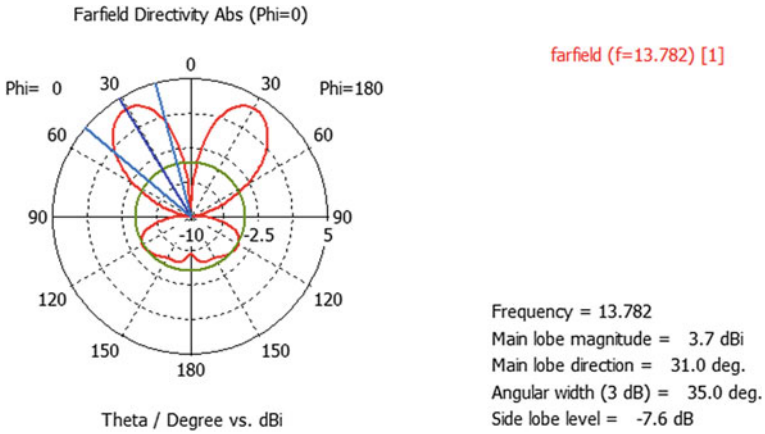
Fig. 9 Radiation pattern at 13.782 GHz

Fig. 10 Gain of the proposed antenna
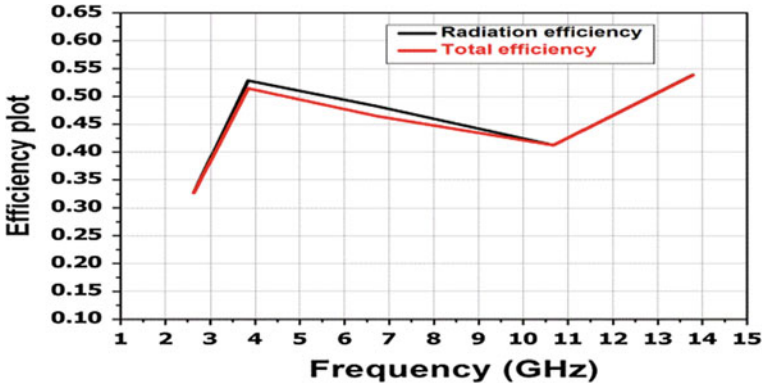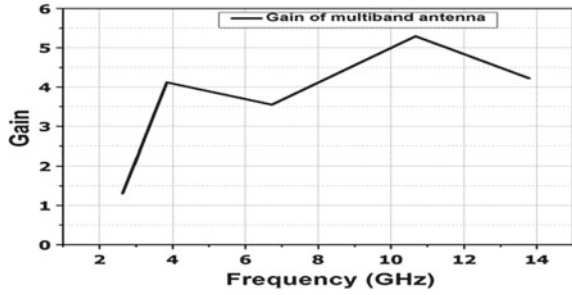
Fig. 11 Radiation efficiency of the proposed antenna

Table 4 Application area of resonating frequencies

S. no.	Resonant frequency (GHz)	Application area
1	2.638	Fixed mobile and aeronautical mobile communication
2	3.84	Amateur radio and mobile communication
3	6.72	Fixed satellite (Earth-to-satellite and satellite-to-earth) communication
4	10.67	X-band Satellite Communication and Fixed satellite (Earth-to-satellite and satellite-to-earth) communication
5	13.78	Space research and radio location

5   Conclusion

In this paper design and radiation performance of rectangular patch antenna using dumbbell shape, DGS is reported through simulated analysis and measured results. The proposed antenna is a multiband antenna and resonates at five different frequencies, i.e., 2.638, 3.842, 6.72, 10.67, and 13.78 GHz. The return loss, VSWR, radiation efficiency, gain and radiation patterns are described in detail. The results show that the proposed antenna is quite suitable for multiband applications. Some of these possible applications are summarized in Table 4.

References

1. Balanis, C.A.: Antenna Theory, Analysis and Design. Wiley (2005)
2. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: Microstrip Antenna Design Handbook. Artech House (2001)
3. James, J.R., Hall, P.S.: Handbook of Microstrip Antennas. Peter Peregrines (1989)
4. Yang, K.P., Wong, K.L.: Inclined-slot-coupled compact dual-frequency microstrip antenna with a cross slot. Electron. Lett. **34**(4), 321–322 (1998)
5. Chen, H.D.: A dual-frequency rectangular microstrip antenna with a circular slot. Microwave Opt. Technol. Lett. **18**(2), 130–132 (1998)
6. Chen, W.S.: Single-feed dual-frequency rectangular microstrip antenna with square slot. Electron. Lett. **34**(3), 231–232 (1998)
7. Lu, J.H., Wong, W.L.: Compact dual-frequency circular microstrip antenna with a circular offset slot. Microw. Optical Technol. Lett. **22**(4), 254–256 (1999)
8. Tiang, J.J., Islam, M.T., Misran, N., Mandeep, J.S.: Circular microstrip slot antenna for dual-frequency RFID application. Prog. Electromag. Res. **120**, 499–512 (2011)
9. Yang, Y.H., Guo, J.L., Sun, B.H., Huang, Y.H.: Dual-band slot helix antenna for global positioning satellite applications. IEEE Trans. Antennas Propag. **64**(12), 5146–5152 (2016)
10. Hung, C., Chiu, T.: Design of dual-band cavity-backed slot antenna loaded with spur line. IET Microw. Antennas Propag. **10**(9), 939–946 (2016)
11. Gajera, H.R., Anoop, C.N., Naik, G., Archana, S.P., Nandini, R., Pushpitha, B.K., Ravi Kumar, M.D.: The microstrip rectangular microstrip patch antenna (RMPA) with the defected ground plane forhiperlan. Int. J. Electron. Commun. Technol. IJECT, **2**(3), 172–175 (2011)
12. Breed, G.: An introduction to defected ground structures inmicrostripcircuits. In: Tutorials on High-Frequency Electronics, Summit Technical Media, Nov. 2008. LLC, pp. 50–52 (2008)

13. Geng, J.P., Li, J., Jin, R.H., Ye, S., Liang, X., Li, M.: The development of curved microstripantenna with defected groundstructure. Prog. Electromagnet. Res. PIER, pp. 53–73 (2009)
14. Bansal, R., Jain, A., Kumar, M., Meena, R.S.: Design, analysis and modeling of dual-band microstrip loop antenna using defective ground plane. Int. J. Electr. Robot. Electron. Commun. Eng. **6**(9), 72–76 (2012)

Optimizing Resource Allocation of MIMO-OFDM in 4G and Beyond Systems

Arun Kumar Singh, Neelam Srivastava and Saurabh Dixit

Abstract The blend of Multiple-Input–Multiple-Output (MIMO) and Orthogonal Frequency-Division Multiplexing (OFDM) provides a formidable air interface for fourth-generation (4G) technology. In beyond 4G standards, the derivative technologies of OFDM along with MIMO are poised to function as the air interface. The task of optimizing the performance metrics for resource allocation in MIMO-OFDM is a crucial aspect in 4G systems like Long Term Evolution (LTE) or LTE Advanced (LTE-A). As the evolution of LTE-Advanced continues and opens the vistas to fifth-generation (5G) standard, several performance metrics need to be optimized. The prime metrics which characterize a wireless system are the ergodic capacity and Spectral Efficiency (SE) of the system. There exists a precarious trade-off between SE and Energy Efficiency (EE) which needs to be judiciously balanced. Other aspects of resource optimization include enhancement of the throughput at the user end, reduction of cost per bit. This paper explores the critical aspects of potential enhancement and discusses ways to optimize resource allocation. Specific attention is given to the provisioning of the Cyclic Prefix (CP) in OFDM and use of Multiuser MIMO (MU-MIMO). Although CP is an overhead, but it is required to overcome the effects of Inter-Symbol Interference (ISI). However, adaptive CP length adjustment can improve the spectral efficiency. It is demonstrated that MU-MIMO yields a higher energy efficiency as compared to point-to-point MIMO. Therefore, MU-MIMO is ideally suited to yield a higher throughput in 4G and beyond systems.

Keywords Cyclic prefix · LTE advanced · MIMO · MU-MIMO · OFDM

A. K. Singh (✉) · N. Srivastava
Rajkiya Engineering College, Kannauj, India
e-mail: arun@reck.ac.in

N. Srivastava
e-mail: neelam.srivastava@ietlucknow.ac.in

S. Dixit
Central Institute of Plastics Engineering and Technology, Lucknow, India
e-mail: saurabh2911@ieee.org

1 Introduction

In the fourth-generation (4G) networks, there has been a ubiquitous proliferation of Multiple-Input–Multiple-Output (MIMO) along with Orthogonal Frequency-Division Multiplexing (OFDM) as the radio channel. In the upcoming fifth-generation (5G) networks, MIMO is expected to be further scaled up to massive MIMO. For the multiplexing technique, the derivative technologies of OFDM like Filter-Bank-Multi-Carrier (FBMC) and Generalized Frequency-Division Multiplexing (GFDM) are potential candidates. To elevate the performance of the wireless network, it is imperative to deploy adaptive transmission and dynamic resource allocation. The Base Station (BS) dons the responsibility for choosing the best method for bit loading in MIMO-OFDM Multiple Access Control (MAC) scheme. The BS also allocates the subcarrier to be used by the symbol. In [1] the author(s) opine that the BS exploits the Channel State Information (CSI) to select the transmit antenna from among the multiple users. In [2] the author(s) venture to design a robust technique modeled around convex optimization theory. By means of this technique, they have tried to minimize the total transmit power required for MIMO-OFDM communications. The author(s) have vigorously worked toward the analysis of the individual user rate constraints. The merit of this strategy is that it can be applied to downlink as well as uplink as it requires only linear transmit and receives processing. The complexity of the system is deflated by dividing the entire process into M number of individual subprocess. The process of division is achieved by applying the Lagrangian dual objective function which is used for the sum power. Thus, the whole process of convex optimization is accomplished by applying the Lagrangian dual function to chip the problem into M number of individual subproblems. Here, M is equal to the number of subcarriers. This approach drastically scales down the complexity of the system to vary linearly with M instead of varying exponentially with M. As the number of subcarriers is usually large for multi-carrier systems, savings can be augmented by this approach. To dynamically harmonize the step sizes of the Lagrange multiplier, super gradient of the dual function is used. The convergence such a manner style of the step sizes dictates the process of dynamic adjustment. This adjustment is carried out so that the convergence of the algorithm is hastened.

In [3] the author(s) have diagnosed the compatible energy-conserving schemes for cellular networks for a realistic power model. They have attempted to realize their objective by highlighting the issue of general overlay assumptions about the underlying hardware. There is a significant departure in their approach as the focus shifts to diminish the supply power of the cellular BS instead of trying to disparage the Radio Frequency (RF) transmission power. Hence, to furnish a tangible mechanism for current cellular systems like 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE), the author(s) present a new strategy. In their approach, they attempt to achieve antenna adaptation by deploying power control and sleep modes algorithm. The application of power control blended with sleep mode shrinks the BS supply power consumption of MIMO-OFDM. This decrease in power is more prominent for multiuser MIMO (MU-MIMO). In [4] the author(s) strongly promote

the use of MU-MIMO scheme as they opine that it will boost the multiple access capacity. The multiple access capacity varies directly to the number of BS antennas employed. In contrast, MU-MIMO is more robust against most of the propagation limitations which hinder single user (SU) MIMO communications. The loss in channel rank in SU-MIMO is successfully countered in MU-MIMO. Similarly, the issue of antenna correlation is addressed aptly in MU-MIMO. In SU-MIMO, line-of-sight (LOS) propagation causes harsh decline in the performance of Spatial Multiplexing (SM) schemes. However, in MU-MIMO scheme, LOS propagation ceases to pose a problem. Therefore, MU-MIMO facilitates the SM gain at the BS while at the same time obliterates the need for multiple antenna terminals. In a nutshell, MU-MIMO lays the foundation for the deployment of small and cost-effective terminals in a way which allows the control of such that intelligence and cost on the infrastructure side. In [5] the author(s) propose adaptive modulation and coding (AMC) to boost the capacity of wireless network. Wireless and mobile communication systems are characterized by time-varying channel conditions and time-varying system capacity. To circumvent the ill-effects of such time-varying channel conditions, the wireless systems should acquire a formidable degree of tractability to adapt on many levels to achieve desired performance. The instances of such flexibility include information rate adaption, power control, code adaptation, bandwidth adaptation, antenna adaptation, and protocol adaptation. AMC is a kind of information rate adaptation and is used presently in High-Speed Downlink Packet Access (HSDPA) and IEEE 802.16 standard.

The rest of the paper is organized as follows: in Sect. 2 a survey of various schemes for the design of Cyclic Prefix (CP) is put forth. Section 3 highlights the capacity advantage and Energy Efficiency (EE) of MU-MIMO as compared to point-to-point MIMO and results presented. Section 4 draws the conclusion and presents future direction of the work.

2 Design of Cyclic Prefix

Figure 1 represents a generic OFDM scheme where Cyclic Prefix (CP) is inserted at the beginning of the symbol to nullify the hazardous effects of Inter-Symbol Interference (ISI). The CP in OFDM is used as a guard band to combat the ill-effects of inter-OFDM symbol interference. To prevent Inter-Carrier Interference (ICI), it is necessary to maintain orthogonality among subcarriers; hence, a CP is added at the head of each symbol, and the length of CP is kept higher than the expected length Of Channel Impulse Response (CIR). However, CP being an overhead decreases the Spectral Efficiency (SE) of the system. It wastes channel resources for relatively short channel impulse response. CP may not render itself as necessarily optimal to eliminate interference. Hence, several approaches have been inculcated to cope up with this problem.

In [6], the author(s) have addressed the problem of designing the CP length for OFDM systems in a capacity-optimal way. The optimal and simplified metrics which
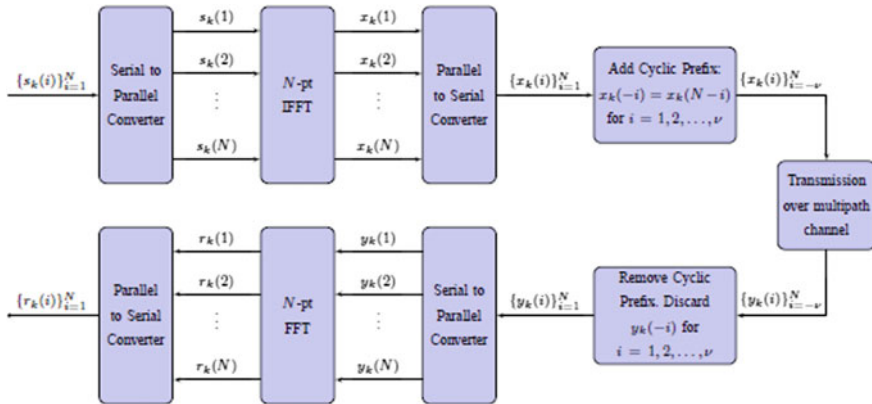
Fig. 1 An OFDM schematic

render themselves suitable to maximize capacity are highlighted. After applying those metrics, the author(s) propose practical resource algorithms like bit loading and OFDM sub-channel assignment along with CP-length adaptation. In [7], the author(s) devise a technique for determining optimal windows that trade-off interference suppression with intercarrier and interframe interference. The technique proposed by the author(s) assumes that channel statistics are known. Thus, this novel technique optimized windows for various levels of interference, delay spreads, frame length, and cyclic prefix lengths can be produced. In [8], the author(s) use a Discrete Multi-tone (DMT) system where a cyclic prefix serves as a tool to divide the channels into independent, subcarriers. However, the duration of the cyclic prefix has to be as long as the channel impulse response. In practical designs, the cyclic prefix is usually fixed. As a result, distortion might occur at, the channel output, if the channel impulse response is longer than the cyclic prefix. The distortion may be so severe that it dominates other noise. A precoder is used at the transmitter to ensure that distortion does not exist at the receiver. In [9], the author(s) deploy the approach of Per-Tone Equalization (PTEQ) for MIMO-OFDM systems.

3 Capacity and Energy Efficiency in MU-MIMO

In a wireless fading channel, capacity serves as a standard performance metric [10]. The capacity of a wireless channel gives an estimate of maximum data rates that can be transmitted with a small error rate. According to Shannon, capacity is defined as the mutual information of the channel maximized over all possible input distributions [11]. It is expressed by the formula:

$$C = B \log_2(1 + \gamma) \tag{1}$$

where B is the channel bandwidth, γ is the Signal-To-Noise Ratio (SNR). In Fig. 2 the capacity of MIMO is depicted over the 16 Quadrature Amplitude Modulation (QAM) for a two transmitter, two receiver scheme. The capacity is displayed for multiplexing, Additive White Gaussian Noise (AWGN), and Rayleigh channel. With a higher Degree of Freedom (DoF), there is a capacity advantage. In Fig. 3 the sum needed power for different schemes are compared. It is seen that the sum needed power for Zero forcing Dirty Paper coding (ZF-DPC) and Zero forcing Beamforming (ZF-BF) are much less as compared to that of traditional Time Division Multiple Access (TDMA). In a wireless fading channel, the capacity is calculated as a mean with respect to time, as described by [12]. The capacity in a probabilistic fading channel is calculated as an average over a distribution. Hence, it is known as the ergodic capacity and is characterized by [13, Eq. 8], assuming independent and identically distributed (i.i.d.) fading. It is expressed by (2) as

$$\widehat{C} \triangleq \int_0^\infty B \log_2(1 + \gamma) p_\gamma(\gamma) \tag{2}$$

where $p_\gamma(\gamma)$ is the Probability Density Function (Pdf) of SNR. To simplify Eq. (2), we use the integral identity in [14, Eq. 4.337.2] such that

$$\widehat{C} = \frac{1}{2 \log_e 2} \exp\left(\frac{1}{4\Omega_L S}\right) \text{Ei}\left(\frac{1}{4\Omega_L S}\right) \tag{3}$$

where $\text{Ei}(\cdot)$ is the exponential integral function.

The expression for channel capacity in MU-MIMO channel can be expressed as
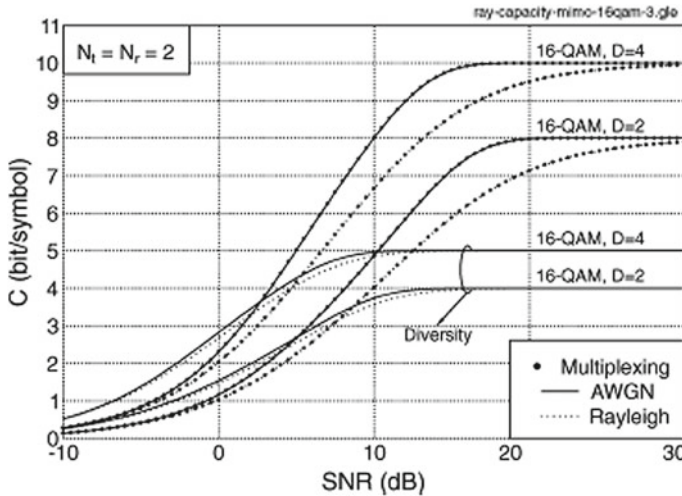
Fig. 2 The capacity of MIMO for 16 QAM
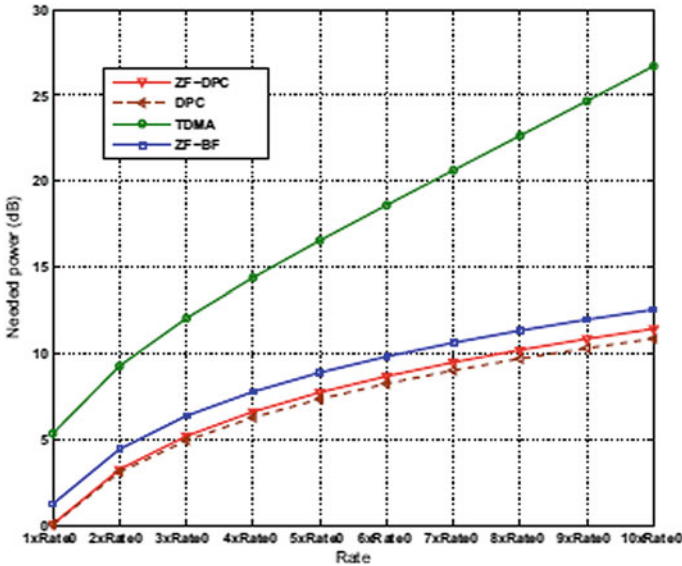
Fig. 3 Sum needed power of different rates

$$C = \sum_{i=0}^{M} \log_2\left(1 + \frac{E_s \lambda_i}{MN_o}\right) \tag{4}$$

Here, λ represents the channel power over the links. For achieving optimum capacity, the channel matrix H should be orthogonal. Hence, when the channel matrix is orthogonal, then the capacity is expressed as

$$C = M \log_2\left(1 + \frac{E_s}{N_o}\right) \tag{5}$$

Equation (5) provides a revealing insight into the capacity for a MIMO channel. The key inference is that the capacity in an orthogonal MIMO channel is M times the capacity of the SISO channel.

The capacity of a point-to-point MIMO channel is given as

$$\widehat{C} = \int_0^\infty \log_2(1 + \gamma) \frac{\gamma^{n_r n_d - 1} \exp\left(\frac{-x}{\bar{\gamma}}\right)}{\bar{\gamma}^{n_r n_d} \gamma(n_r n_d)} d\gamma \tag{6}$$

n_r, n_d are the number of transmitting and receive antennas. $\bar{\gamma}$ is the average SNR, $\Gamma(\cdot)$ is the gamma function, the PDF of (6) is provided by [15]. By employing (5) and (6), the capacity in terms of Energy Efficiency (EE) is plotted in Figs. 4 and 5. Figure 4 represents the capacity advantage and EE of higher number of independent links. As the number of independent links increases, the capacity and EE are augmented. In
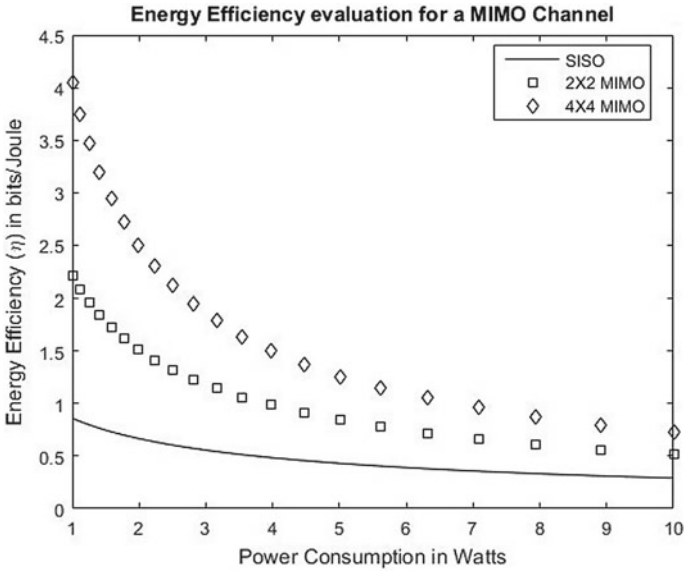
Fig. 4 Energy efficiency in a MIMO system
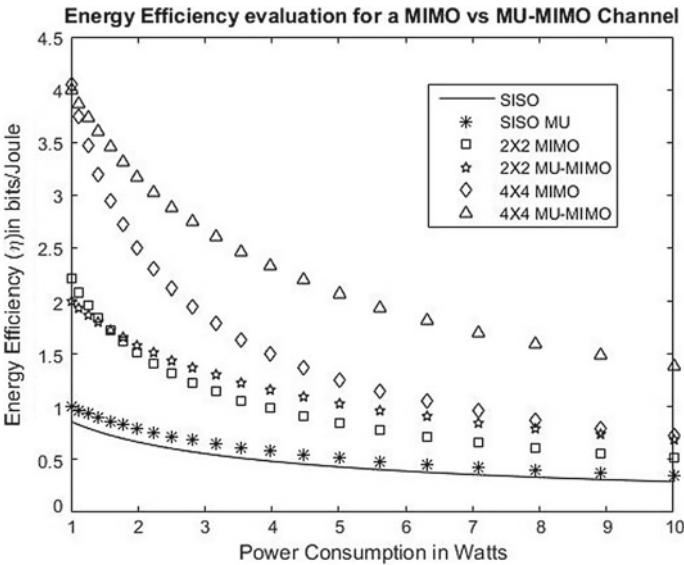
Fig. 5 Energy efficiency in a MU-MIMO system

Fig. 5 the results demonstrate that MU-MIMO yields higher throughput as compared to point-to-point MIMO. Therefore, there is a significant improvement of capacity and EE in MU-MIMO as compared to point-to-point MIMO. These results are the manifestation of the orthogonal channel in MU-MIMO and the use of pre-coding to mitigate the effect of interference.

4   Conclusion

This paper summarizes the various resources provisioning in MIMO-OFDM with a particular emphasis on cyclic prefix and multiuser MIMO. It is inferred that the length of cyclic prefix can be minimized while at the same time orthogonality of the subcarriers be preserved by the use of adaptive algorithms and decoder. Thus, the overhead of including CP can be minimized using algorithms which enable adaptive adjustment of the CP length. The use of multiuser MIMO is imperative to improving the capacity and energy efficiency. MU-MIMO exploits the principle of user diversity and is instrumental in the realization of virtual MIMO system. The results presented demonstrate the higher energy efficiency of MU-MIMO as compared to point-to-point MIMO.

References

1. Maung, S.M., Iwao, S.: Efficient resource allocation for multiuser MIMO-OFDM uplink system to guarantee the proportional data rate fairness among users in a system. In: 1st International Symposium on Access Spaces (ISAS), IEEE-ISAS (2011)
2. Ho, W.W.L., Liang, Y.-C.: Optimal resource allocation for multiuser MIMO-OFDM systems with user rate constraints. IEEE Trans. Veh. Technol. **58**(3) (2008)
3. Holtkamp, H., Auer, G., Bazzi, S., Haas, H.: Minimizing base station power consumption. IEEE J. Sel. Areas Commun. **32**(2), 297–306 (2014). https://doi.org/10.1109/jsac.2014.141210
4. Gesbert, D., Kountouris, M., Heath Jr., R.W., Chae, C.-B., Sälzer, T.: Shifting the MIMO paradigm. IEEE Signal Process. Mag. (2007)
5. Jiang, M., Hanzo, L.: Multi-user MIMO OFDM for next generation wireless systems. Proc. IEEE. **95**(7) (2007)
6. Tonello, A.M, D'Alessandro, S., Lampe, L.: Cyclic prefix design and allocation in bit-loaded OFDM over power line communication channels. IEEE Trans. Commun. (2010)
7. Copeland, A.D., Bliss, D.W., McKellips, A.L.: Optimal windowing in MIMO OFDM for network interference suppression. In: IEEE Conference on Signals, Systems and Computers (2009)
8. Cheong, K.W., Cioffi, J.M.: Precoder for DMT with insufficient cyclic prefix. In: Proceedings of IEEE International Conference on Communications (ICC), pp. 339–343 (1998)
9. Leus, G., Moonen, M.: Per-tone equalization for MIMO OFDM systems. IEEE Trans. Signal Process. **51**, 2965–2975 (2003)
10. Katiyar, H., Bhattacharjee, R.: Average capacity analysis of multi-antenna regenerative cooperative relay in Rayleigh fading channel. ICC (2009)
11. Goldsmith, A.: Wireless Communications. Cambridge University Press (2005)

12. Lee, W.C.Y.: Estimate of channel capacity in Rayleigh fading environment. IEEE Trans. Veh. Technol. **39**(3) (1990)
13. Goldsmith, A.J., Varaiya, P.P.: Capacity of fading channels with channel side information. IEEE Trans. Inf. Theory. **43**(6) (1997)
14. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products, 7th edn. Academic Press (2007)
15. Chalise, B.K., Vandendorpe, L.: Outage probability analysis of a MIMO relay channel with orthogonal space-time block codes. IEEE Commun. Lett. **12**(4) (2008)

Design and Development of 2.1 GHz Horn Antenna

Indra Kumar Singh, Rajiv Kumar Singh and Radhey Lal

Abstract This paper describes the building and simulation of a 2.1 GHz horn antenna which can be widely used in wireless communication systems, electromagnetic sensing, microwave heating, and for calibration of other antennas. The antenna was designed to have the highest gain of 15 dB at 2.1 GHz although it can be used for frequency range in between 1.9 and 2.3 GHz which generally lies in L-band and S-band. This paper describes the calculations and measurement of horn antenna parameters in detail which includes waveguide parameters and flare dimensions. Using the designed value horn antenna is fabricated of thickness 2 mm. The numerical modeling was performed on CST software. The result showed that the highest gain of 15 dB is obtained at 2.1 GHz.

Keywords Horn antenna · Antenna gain · Directivity · Radiation pattern · CST simulation

1 Introduction

Horn antenna is a type of aperture antenna that finds its applications in the wide areas of wireless communication, LAN extensions such as marine communication, satellite communication, microwave heating, electromagnetic heating, phased arrays, feed elements for reflector, and lens antennas. Horn antennas are also widely used as calibration antenna for other antennas [1].

With the knowledge of the fields over the aperture, the radiation field from antenna aperture can be determined. Horn antenna is very popular at higher frequencies and UHF. From Maxwell equations theoretical analysis of the electromagnetic horn

I. K. Singh · R. K. Singh (✉)
Department of Electronics Engineering, Institute of Engineering & Technology, Lucknow, India
e-mail: rajivinbhu@gmail.com

Dr. APJ Abdul Kalam Technical University, Lucknow 226021, India

R. Lal
Council of Science & Technology, Lucknow, UP, India
e-mail: radheylal.cst@gov.in

251

antenna is derived [2]. It is constructed by flaring of waveguide. The flaring is done for impedance matching. The impedance of waveguide is nearly about 50–60 Ω and that of free space is nearly 377 Ω [3]. It can provide high gain and increases the directivity of antenna. It is a directional antenna and generally used for long-distance communication and feed for large parabolic reflector antenna. There are several types of horn antenna such as E-plane sectoral horn, H-plane sectoral horn, pyramidal horn, conical horn, and corrugated horn antenna [3]. The directional pattern can be used so that the power radiated is focused in desired direction. Today, there is wide range of the antenna in the market but sometimes the commercial antennas do not suit the user requirements, either the gain is too small or frequency range does not suit the user requirements. So, only option is to build our own antenna. The advantages of horn antenna are low VSWR, wide bandwidth, robust design, simple in construction, and easy to excite. To fulfill so many practical applications, they are constructed in wide variety of shapes and sizes [4–9].

2   Horn Antenna Design

First step in designing the horn antenna is deciding the gain required for the desired range of frequency for a particular operation. In electromagnetic theory, antenna gain is defined as the ratio of power in the direction of peak radiation to average power in all direction. In other words, antenna gain is the combination of both directivity and efficiency. For a transmitting antenna, antenna gain is defined as how well the antenna is converting the power in the form of electromagnetic waves in desired direction. Similarly, at receiving end, how well the receiving antenna converts the received electromagnetic waves into the electrical power. The gain of antenna is generally expressed in dB. The dipole antenna has a gain which is very less as compared to the horn antenna. So, the aim of this paper is to achieve the gain of 15 dB at 2.1 GHz which is generally used for communication purpose.

The dimensions of waveguide are determined depending on the frequency of use for particular purpose. The rectangular waveguide shown in Fig. 1 has parameters width (a), height (b), and length (l) of the waveguide. Since 2.1 GHz frequency generally lies in S-band, we use WR-430 standard waveguides. For 2.1 GHz horn

Fig. 1 Waveguide dimensions

$$l = \frac{\lambda_g}{2}$$
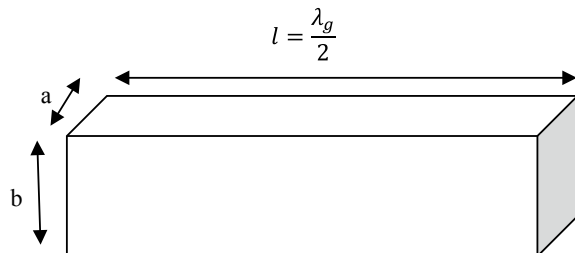
antenna, width has been taken as a $= 4.3'' = 10.922$ cm, height as b $=$ a/2 $= 5.46$ cm, and depth (l) of the waveguide equal to the half of the cutoff wavelength of waveguide.

Waveguide wavelength is given by

$$\lambda_g = \frac{1}{\sqrt{\frac{1}{\lambda_0^2} - \frac{1}{\lambda_c^2}}} \tag{1}$$

where λ_C is the cutoff frequency wavelength for the mode of transmission.

Cutoff wavelength of waveguide in different modes is calculated by

$$(\lambda_c)_{mn} = \frac{1}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} \tag{2}$$

where m and n are integer numbers.

The mode of transmission has been taken as TE_{10} mode (m $= 1$, n $= 0$). Thus $(\lambda_c)_{10} = 2a = 2(10.922) = 21.844$ cm. Wavelength in free space λ_0 is calculated by

$$\lambda_0 = \frac{c}{f} \tag{3}$$

Thus,

$$\lambda_0 = \frac{3 \times 10^{10}}{2.1 \times 10^9} \text{ cm} = 14.2857 \text{ cm}$$

After determining the aperture of waveguide, we have to determine the position and height of the feed antenna. Position of feed antenna is equal to the one-fourth of the waveguide wavelength from the edge of the antenna ($\frac{\lambda_g}{4}$) and the height of feed antenna is equal to the one-fourth of the wavelength in free space ($\frac{\lambda_0}{4}$) (Fig. 2). So, position of feed antenna is $\frac{\lambda_g}{4} = 4.721125$ cm and height of feed antenna is $\frac{\lambda_0}{4} = \frac{14.2857}{4} = 3.5714$ cm.

Fig. 2 Waveguide design
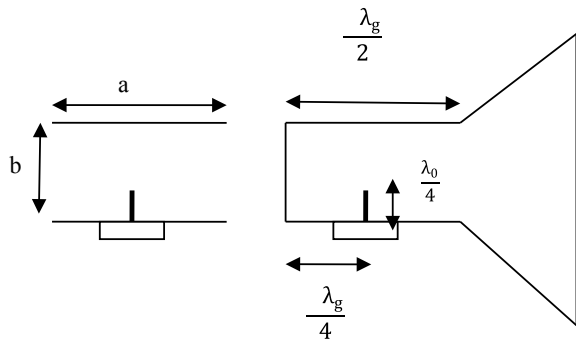
Fig. 3 Pyramidal horn antenna design
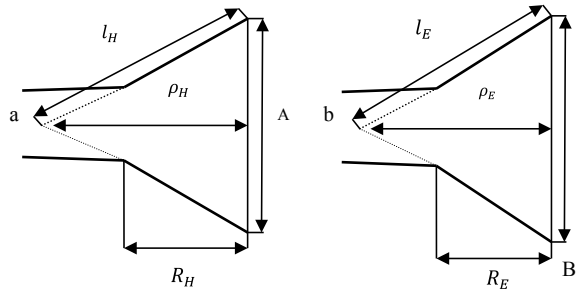
For a pyramidal horn antenna, the gain in terms of aperture area is calculated as

$$G = \varepsilon_{ap} \cdot \frac{4\pi}{\lambda_0^2} \cdot A_{ab} \tag{4}$$

where ε_{ap} = effective aperture coefficient (0.51 for copper material), and A_{ab} = area of the aperture (A × B).

For the desired gain of 15 dB (=31.622) and free space wavelength λ_0 = 14.285 cm = 0.142857(m), the area of aperture A_{ap} (A × B) is 0.100696 (m^2).

Figure 3 shows the flare dimensions and parameters of pyramidal horn antenna. For this antenna following are the design equations:

$$A.B = 0.100696 \, (\text{m}^2) \tag{5}$$

$$\rho_H \left(1 - \frac{a}{A}\right) = \rho_E \left(1 - \frac{b}{B}\right) = R_H = R_E \tag{6}$$

$$A = \sqrt{3\lambda_0 \rho_H} \tag{7}$$

$$B = \sqrt{2\lambda_0 \rho_E} \tag{8}$$

Therefore, from four Eqs. (5), (6), (7), and (8) we can calculate the four unknown values of A, B, ρ_H, and ρ_E. Other dimensions are calculated by the following equations:

$$\frac{\rho_H}{R_H} = \frac{A}{A - a} \tag{9}$$

$$\frac{\rho_E}{R_E} = \frac{B}{B - b} \tag{10}$$

$$l_H = \sqrt{\rho_H^2 + \frac{A^2}{4}} \tag{11}$$

Table 1 Antenna design

A(m)	B(m)	a(m)	b(m)	ρ_H (m)	ρ_E (m)	$R_E = R_H$ (m)	l_H (m)	l_E (m)
0.37	0.27	0.10922	0.05461	0.32	0.26	0.21	0.37	0.29

$$l_E = \sqrt{\rho_E^2 + \frac{B^2}{4}} \tag{12}$$

For the pyramidal horn antenna, the calculated parameters are shown in Table 1.

3 Designing of Horn Antenna

3.1 Experimental Setup

The experimental parameter measurements were performed at the Microwave Laboratory, Department of Electronics Engineering, Institute of Engineering and Technology, Lucknow, India. The fabricated horn antenna and its measurement setup are shown in Fig. 4a, b. The distance between the generator and spectrum analyzer is 6.5 m for generator we have used Klystron power supply, (model no.: NV102), and spectrum analyzer (LG model no.: RS132) has been also used for measurement. Horn

(a)

(b)

Horn antenna

Generator

6.5m

Spectrum analyzer
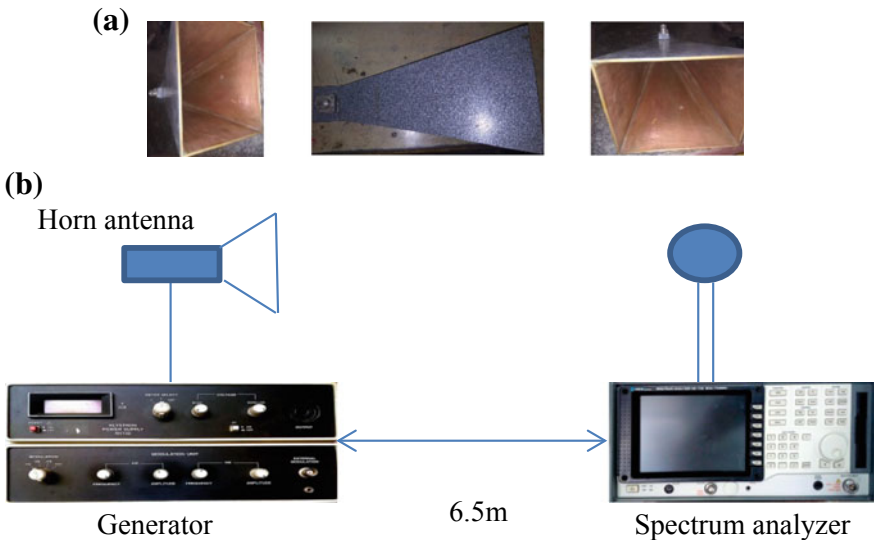
Fig. 4 a Fabricated horn antenna, **b** Measurement setup

Fig. 5 Pyramidal horn antenna design using CST
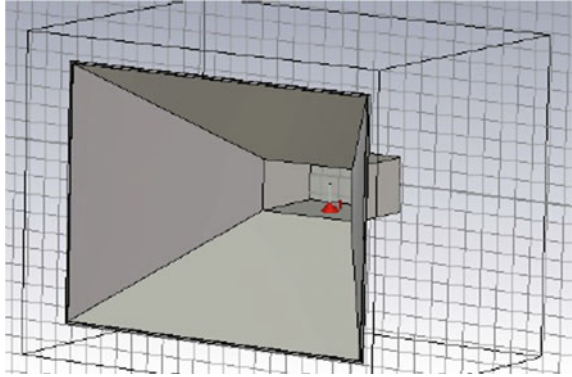
antenna is made up of copper material which is 2 mm thick. The measurement was performed in the frequency range of 1.9–2.3 GHz [1, 3].

3.2 Simulation Modeling

Horn antenna was simulated using the Computer Simulation Technology (CST) software. CST software is very useful for 2D and 3D design and analysis of antenna structures [6]. The pyramidal horn antenna is excited using a waveguide which is fed by supply source. The horn antenna plates are of finite thickness (2 mm). Flare dimensions are the most significant part of horn antenna designing (Fig. 5).

4 Result and Analysis

A horn antenna operating at 2.1 GHz has been fabricated, and simulations using CST software have been done for this antenna. Figure 6 shows S-parameter analysis of the designed antenna which is helpful in measurement of the dissimilarity between impedance of load and metallic transmission line. Here, -10 to -15 dB and higher value of S-parameter has been obtained for the desired operating frequency.

Figure 7 shows the VSWR graph of the designed horn antenna which is associated with impedance matching. Simulated value of VSWR is less than two and hence it is good for signal transmission. Radiation pattern of the designed horn antenna is shown in Fig. 8.

Measurement was performed in the frequency range of 1.6–2.7 GHz for power density and antenna gain. The losses have been taken into account and have been calculated using expression $L = 20log(\lambda/4\pi d)$, where d is separation between the transmitting antenna and receiving antenna.
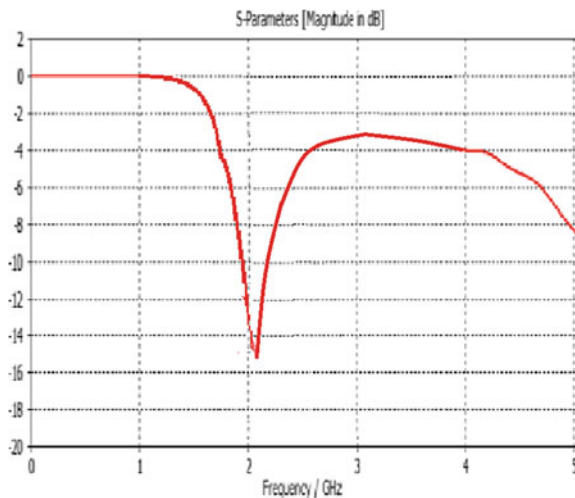
Fig. 6 S-Parameter of horn antenna
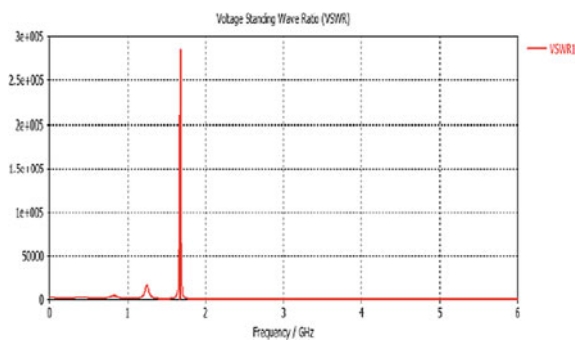
Fig. 7 VSWR of horn antenna
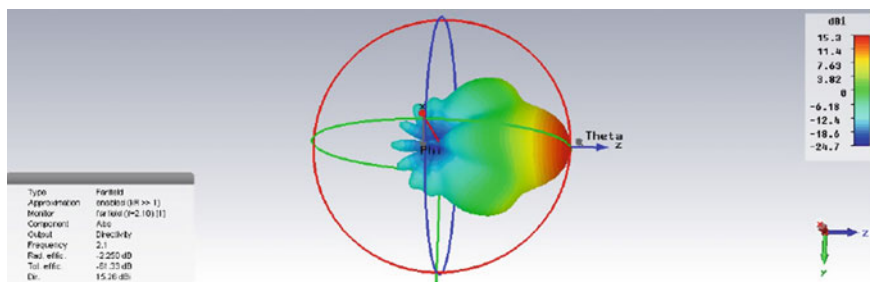
Fig. 8 Radiation pattern of horn antenna

Table 2 Measurement data for the designed horn antenna

Frequency (Ghz)	Received power P_r (dBm)	Signal attenuation L (dB)	Antenna gain G_t (dB)
1.6	−28.10	52.45	4.01
1.7	−29.13	53.95	4.07
1.8	−27.04	54.44	6.67
1.9	−27.58	54.91	6.59
2.0	−27.89	55.36	6.72
2.1	−27.72	56.94	8.48
2.2	−26.86	57.08	9.46
2.3	−27.99	56.57	7.84
2.4	−27.91	55.78	7.13
2.5	−29.48	56.19	5.96
2.6	−32.94	57.30	3.61
2.7	−36.89	57.64	0.00

Antenna gain can be calculated as $G_t = P_r - P_t - G_r - L$ where P_t, P_r, G_r, and L are transmitted power, received power, receiver antenna gain, and transmission loss, respectively. The transmitted power has been set at +15 dBm. Table 2 shows measurement data for the designed horn antenna.

Close observation of Table 2 shows that the highest power density is obtained at 2.1 GHz frequency. Received power (with $G_r = 1$ due to spectrum analyzer) is calculated as $P_r = (P_d/4\pi)(c^2/f^2)G_t$, where P_d is received power density and c is speed of light.

Transmission loss has been taken into account for the calculation of the antenna gain. Cable losses (0.25 dB/m), indoor transmission losses (1 dB/m), and connector losses (0.5 dB each) have been considered.

Figure 9 shows the gain–frequency response of the designed horn antenna. It is observed that the highest gain is achieved near 2.1 GHz, i.e., near the operating frequency. The difference is measured and calculated value of the antenna gain is due to the measurement errors.

Fig. 9 Gain–frequency response of designed horn antenna
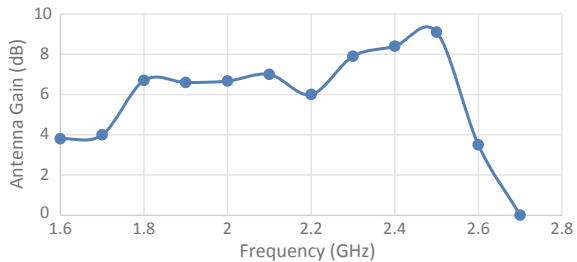
5 Conclusion

The successful implementation and simulation of 2.1 GHz horn antenna are done using CST software and by lab experimental setup. From the experimental and simulation studies, it is clear that signal integrity be intercepted or transmitted depending upon the design consideration of pyramidal horn antenna. The antenna gain of approximately 9.4 dB has been achieved by the measurements. The desired gain of 15 dB could be achieved with precise building of horn antenna and by increasing the thickness of copper metal sheet which is used for manufacturing of antenna. Although antenna is mainly designed for 2.1 GHz, it can be used in the frequency range of 1.9–2.3 GHz with high gain.

References

1. Banjeglav, G., Malarić, K.: 2.4 GHz horn antenna. Trans. Marit. Sci. **01**, 35–40 (2015)
2. Barrow, W.L., Chu, L.J.: Theory of electromagnetic horn. Proc. IRE. **27**, 51–64 (1939). https://doi.org/10.1109/JRPROC.1939.228693
3. Ikram, S., Ahmad, G.: Design & implementation of a standard gain horn antenna for optimized gain and radiation pattern using MathCAD & HFSS. In: Second International Conference on Electrical Engineering, 25–26 March 2008, University of Engineering and Technology, Lahore (Pakistan), 978-1-4244-2293-7/08/$25.00, IEEE (2008)
4. Zang, S.R., Bergmann, J.R.: Analysis of Omni directional dual-reflector antenna and feeding horn using method of moments. IEEE Trans. Antennas Propag. **62**, 1534–1538 (2014). https://doi.org/10.1109/TAP.2013.2296775
5. Abhignya, G., Yogita, B., Abhinay, C., Balaji, B., Murthy, M.B.R.: Design, fabrication and testing of pyramidal horn antenna. Int. J. Eng. Appl. Sci. (IJEAS). **2**, 2394–3661 (2015)
6. Barik, B.R., Kalirasu, A.: Design of a UHF pyramidal horn antenna using CST. Int. J. Pure Appl. Math. **114**, 447–457 (2017). ISSN 1311-8080 (printed version); ISSN 1314-3395 (on-line version). http://www.ijpam.eu
7. Girish, K.: Electrical Engineering Department, IIT Bombay. https://nptel.ac.in/courses/108101092/Week-10-Horn-Antennas-Part-1.pdf. Last accessed 12 Jul 2018
8. Rectangular Waveguide Dimensions. http://www.microwaves101.com/encyclopedia/waveguidedimensions.cfm. Last accessed 5 June 2018
9. Krishnasamy, T., Selvan: Accurate design method for pyramidal horns of any desired gain and aperture phase error. IEEE Antennas Wirel. Propagat. Lett. **7** (2008). https://doi.org/10.1109/lawp.2007.914119

Full-Duplex Wireless Communication in Cognitive Radio Networks: A Survey

Ashish Kumar Rao, Rajiv Kumar Singh and Neelam Srivastava

Abstract Since past few years, as the technology growing in the field of wireless communication with their ubiquitous application, the demand for spectrum resources is also increased. To exploit the demand for spectrum, Cognitive Radio Networks (CRNs) became the most promising solution. Many works have been done in half-duplex CRNs, where at a given time secondary user can only sense the spectrum or transmit the data. Half-duplex CRN is unable to transmit the data during sensing period. This is the major limitation of half-duplex CRN. On the other hand, full-duplex CRNs can simultaneously sense and transmit the data at a given time, which improves the throughput of system as compared to half-duplex CRNs. This property of full-duplex communication motivates for more research in field of energy efficiency, energy harvesting, and throughput maximization in CRNs. In this paper, we have presented the energy efficiency, energy harvesting, and throughput calculations with open issues and challenges of full-duplex communication in cognitive radio networks.

Keywords Cognitive radio network (CRN) · Half-duplex (HD) communication · Full-duplex (FD) communication · Primary user (PU) · Secondary user (SU)

A. K. Rao (✉) · R. K. Singh · N. Srivastava
Department of Electronics Engineering, Institute of Engineering and Technology, Lucknow, India
e-mail: asis151rao@gmail.com

R. K. Singh
e-mail: rajivinbhu@gmail.com

N. Srivastava
e-mail: neelam.srivastava@ietlucknow.ac.in

Dr. APJ Abdul, Kalam Technical University, Lucknow, India

1 Introduction

The tremendous growth in the field of wireless communication technologies in the past few decades enhanced the connectivity between several electronic devices. With the increasing demand of high data rate and better quality of service, service providers move to next-generation technology to provide data-intensive applications. The basic requirements of the consumers are to gain high throughput data communication with better energy efficiency. The subscriptions of wireless services such as television, cellular, broadband, and navigation are increasing day by day. Therefore, it is predicted that global wireless data traffic will increase ten times up to the year 2019, in comparison to the year 2014 [1]. The demand of uninterrupted connectivity and higher data rates motivates the extravagant development in the field of wireless communication and the growth toward next-generation technologies such as femtocells [2, 3], millimeter-wave spectrum technology [4], Multiple-Input–Multiple-Output (MIMO) technology [5, 6], and cognitive radio technology [7–9].

Capacity demand can be achieved with the help of increasing spectral efficiency, spreading of spectrum, and densification of network with increasing small cell in future wireless communication technology. A full-duplex communication system has the potential to double the spectral efficiency by sending and receiving data at the same time and with same frequency resources. However, the transmit antenna and receive antenna introduce the strong self-interference by its transmission process, which makes the decoding of signal difficult. This is the main reason for nonacceptance of FD technology widely until recently. In recent years, the growth of self-interference reduction technique made the FD wireless communication widely acceptable [10]. Many works have been done in the field of FD communication which shows the FD technique can achieve better spectral efficiency in comparison to HD technique [11]. This interesting feature needs more attention to research in the field of FD communication in CRNs. In this paper, our main attention is to present the energy efficiency, energy harvesting, and throughput analysis and discussing research challenges in FD communication.

1.1 Need for Full-Duplex Communication in CRN

Most of the Cognitive Radio Networks (CRNs) exploit the white spaces with half-duplex (HD) radios. The Half-Duplex Cognitive Radios (HD-CRs) perform data communication in two stages. In the first stage, it senses the licensed spectrum and then transmits the data in second stage. In spectrum sensing stage, secondary user (SU) senses the white spaces in license band. The sensing result can be false which results harmful interference to primary user (PU). Therefore, HD-CR user takes a long portion of time for sensing the spectrum and very less portion of time for data communication. One of the big challenges for HD-CR user is, during the data transmission stage, it is unable to detect PU transmission which greatly affects PU activity.

   Data transmission and reception are performed in two different orthogonal channels in HD-CR. This two-channel operation requires more bandwidth in comparison to single-channel operation. It also increases the latencies, as two channels need to be sensed for white space exploitation.

   On the other hand, full-duplex cognitive radio (FD-CR) significantly mitigated the drawback of HD-CR [12–15]. Full-duplex cognitive radio simultaneously senses and transmits the data. The data transmission and reception occur simultaneously over the same channel during a given period of time.

1.2   Contributions of This Survey Article

Many researchers have already given an overview article on selected aspects of Full-Duplex Cognitive Radio Network (FD-CRN). In this paper, we provide the comprehensive state-of-art of FD-CRN along with energy efficiency, energy harvesting, and throughput approach. Introduction and need for full-duplex CRN are discussed in first section. Comparison between half-duplex and full-duplex cognitive radio communication is discussed in second section. In third section, we have discussed the energy efficiency and energy harvesting in FD-CRN. Throughput analysis has been carried out in fourth section. In last section, we discussed the open issue, challenges, and future research direction.

2   Cognitive Radio and Full-Duplex (FD) Communication

2.1   Cognitive Radio Networks (CRNs)

The wireless radio spectrum, ranging from 9 kHz to 3 THz, is divided into chunks of frequency bands. This range of frequency can be divided into two parts: licensed frequency band and unlicensed frequency band. Licensed frequency band is utilized on the basis of licensing and a certain license fee has to be paid to the government regularity authority. It is insured that there will be no other wireless entity using same frequency. Unlicensed frequency band is not sold nationally or internationally and can be utilized for low-cost communication. Due to the limited number of unlicensed bands and large number of users, there are more chances of interferences to occur. International or national regulatory authorities provide the licensed band by static spectrum management policies. According to this polices, fixed spectrum band is provided to licensed holders for long-term basis for a large geographical area. As per the survey [16], it is noticed that utilization of spectrum is varying from 15 to 85%, which shows the most of the spectrums are unused and underutilized.

   Dynamic Spectrum Access (DSA): It is one of the enabling technologies of CR [17] which is widely used for the exploitation of white spaces [7, 18]. CR devices

can identify white spaces using DSA strategies. In this technology, CR devices can sense the white spaces in license band and are able to reconfigure their parameters to move to other white space, when primary user (PU) start communicating in that particular band. During this process, CR user avoids the harmful interference to PU. PUs are licensed users having the exclusive right to use set of licensed frequency band without any interference. On the other hand, CR users or secondary users (SU) are unlicensed users, who do not have the license to use license band. SU can use unlicensed band whenever required without consideration of other user and can use the licensed band opportunistically without causing the interference to primary use (PU).

Spectrum Handoff: Spectrum handoff is one of the most important requirements for uninterrupted communication. Mobility of spectrum depends on spectrum hand-off which is directly related to the quality of service and link maintenance [19]. In CRN, spectrum handoff provides resilient work for SUs. Many works have been done in short-term and long-term handoff performance such as link maintenance probability, number of spectrum handoff, etc. [20].

2.2 Half-Duplex Versus Full-Duplex Communication

Communication can be performed on the basis of data flow capability. The term "duplex" means the two systems communicate with each other simultaneously. In full-duplex (FD) communication, two systems have the capability to communicate with each other simultaneously. The data transmission and reception are performed simultaneously in both the duplex systems. Half-duplex (HD) communication is the most commonly used data flow mode in wireless networks because of its simplicity. The limitation of HD communication systems is that it cannot perform transmission and reception simultaneously. Therefore, in half-duplex (HD), half of the time is utilized for sensing and other half of the time is used for transmission. This process greatly reduces the throughput of the HD system in comparison to FD systems. Furthermore, HD systems are prone to hidden terminal problem [11].

2.3 Network Topology

In FD network, there are mainly three topologies: conventional bidirectional topology, the relay topology, and the base station topology. In conventional bidirectional topology, transmission and reception are performed simultaneously which minimizes the delay and improves the spectral efficiency in comparison to HD network topology. In relay-based topology, the FD relaying can receive and forward data in real time simultaneously on a common carrier [21, 22]. In base station topology, base station supports uplink and downlink data communication on a common carrier simultaneously. In each of these network topologies, multiple antennas are used in a system

Table 1 Comparison between half-duplex CRN and full-duplex CRNs

Parameter	Half-duplex cognitive radio networks	Full-duplex cognitive radio networks
Spectrum sensing (SS)	The duration for spectrum sensing is predefined. This is not continuous in nature	In FD, duration for spectrum sensing is not predefined. This is continuous in nature
Self-interference suppression	Not applicable in HD CRNs	Different approaches are used for self-interference suppression in full-duplex CRNs such as active and passive approach
Secondary transmit power	Increase in the power in HD-CRN increases the throughput also	In full-duplex CRN, there is a trade-off between throughput and power
PR activity	PR activity in half-duplex CRNs monitors primary user with several models	In FD CRN, PU activity becomes more reliable due to continuous spectrum sensing
Security	Less secure in comparison to FD-CRNs	Anti-jamming antennas in FD-CRNs mitigate the impact of various eavesdroppers

and number of antennas can differ per system. The comparison between HD and FD-CRNs is shown in Table 1.

3 Energy Efficiency and Energy Harvesting in FD-CRN

In this section, we find the false alarm probability and the probability of detection for a given system model. We consider the system model used by Fan et al. [23]. In this model of CRN, one primary user pair and one secondary user pair are considered. Primary transmitter (PT) transmits the signal to primary receiver (PR) and secondary transmitter (ST) transmits the signal to secondary receiver (SR) as presented in Fig. 1.

3.1 System Model

The given model presents the RF-Powered Full-Duplex Cognitive Radio Networks (FD-CRNs). In this network model, secondary transmitter (ST) performs the spectrum sensing, and does data transmission when primary user is not active and harvest energy simultaneously.

Antenna A1 performs sensing with time duration τ. After sensing, it decides the presence and absence of primary user; antenna A2 simultaneously transmits the data when primary user is not present, otherwise maintain the silence. Antenna
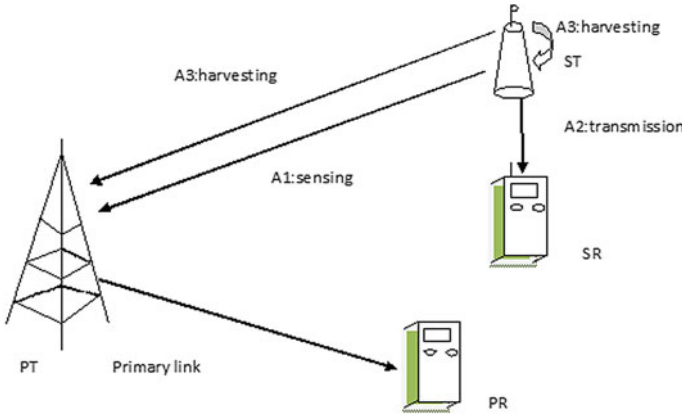
Fig. 1 System model [23]

A3 harvests energy either from primary transmitter, when it is busy or secondary transmitter, when it transmits the data. Antenna A3 also recycles the energy from ST itself. Hence, this harvested energy is again used for secondary transmitter sensing and transmitting.

3.2 Analysis of False Alarm Probability and Detection Probability

To determine the false alarm probability and the probability of detection, it is necessary to know the status of ST. Hence, we use the detection threshold ε_0, ε_1 to decide ST status, whether it is silent or active, respectively.

Secondary Transmitter (ST) is silent: If ST is silent or no data is being transmitted, there are two possible cases. It is possible that PT is active or PT is not active. Due to noise, ST may sense that PT is active, which is known as false alarm. We can formulate the PT detection with the help of two hypotheses H_{00} and H_{01}, where H_{00} presents waste of spectrum due to false alarm. The received signal at ST can be written as

$$y = \begin{cases} hp\,sp + w, & H_{01}, \\ w, & H_{00}, \end{cases} \tag{1}$$

Here, the Rayleigh channel gain with variance σ_{hp}^2 from PT to ST is denoted by hp, the PT's signal with variance σ_p^2 is denoted by sp. ω is the Gaussian noise with zero mean and variance σ_ω^2. The probability of detection $p_d^0(\varepsilon_0)$ and false alarm probability $p_f^0(\varepsilon_0)$ at ST [24] can be calculated as

$$p_d^0(\varepsilon_0) = Q\left(\left(\frac{\varepsilon_0}{(1+\gamma_s)\sigma_\omega^2} - 1\right)\sqrt{f_s\tau}\right), \tag{2}$$

$$p_f^0(\varepsilon_0) = Q\left(\left(\frac{\varepsilon_0}{\sigma_\omega^2} - 1\right)\sqrt{f_s\tau}\right), \tag{3}$$

Here, in this equation the SNR of sensing is denoted as $\gamma_s = \frac{\sigma_{hp}^2\sigma_p^2}{\sigma_\omega^2}$. f_s is the sampling frequency and complementary cumulative distribution function of standard Gaussian denoted as Q(.), which is $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty exp\left(-\frac{t^2}{2}\right)dt$.

ST is active: When ST is active or data is transmitted, we use the hypothesis H_{10} and H_{11} to present the PT state, either idle or busy. If the PT is active and ST is unable to detect the presence of primary signal due to miss detection, state of collision occurs. This is denoted by H_{11}. In such case, the received signal can be presented as

$$y = \begin{cases} hp\,sp + hi\,st + w, & H_{11}, \\ hi\,st + w, & H_{10}, \end{cases} \tag{4}$$

where self-interference channel gain from antenna A2 to antenna A1 is hi, st is the signal of the secondary transmitter. The variance of st is σ_s^2. The term $hi\,st$ can be modeled as Rayleigh distribution with zero mean and variance $\chi^2\sigma_s^2$. Self-interference suppression (SIS) factor χ^2 $(0 < \chi^2 < 1)$ can be defined as the ratio of power received at self-interference signal to secondary transmitters transmit power. The probability of detection $p_d^1(\varepsilon_1)$ and false alarm probability $p_f^1(\varepsilon_1)$ in this case can be written as [4]:

$$p_d^1(\varepsilon_1) = Q\left(\left(\frac{\varepsilon_1}{(1+\gamma_s+\gamma_i)\sigma_\omega^2} - 1\right)\sqrt{f_s\tau}\right) \tag{5}$$

$$p_f^1(\varepsilon_1) = Q\left(\left(\frac{\varepsilon_1}{(1+\gamma_i)\sigma_\omega^2} - 1\right)\sqrt{f_s\tau}\right) \tag{6}$$

where $\gamma_i = \frac{\chi^2\sigma_s^2}{\sigma_\omega^2}$ denotes the interference to noise ratio.

From the above equations, the detection probability and false alarm probability can be found as

$$P_d(\varepsilon_0\,\varepsilon_1) = \frac{P_d^1(\varepsilon_1)}{1 - P_d^0(\varepsilon_0) + P_d^1(\varepsilon_1)}, \tag{7}$$

$$P_f(\varepsilon_0\,\varepsilon_1) = \frac{P_f^1(\varepsilon_1)}{1 - P_f^0(\varepsilon_0) + P_f^1(\varepsilon_1)}, \tag{8}$$

To optimize the energy efficiency and energy harvesting, we need to understand the energy consumption and harvesting process in FD-CRN scenario. Here, we assume

that the secondary transmitter always has data to transmit. Hence, four cases arise as discussed below.

H_{00} denotes that the primary transmitter is not transmitting the data and secondary transmitter not detecting the status of PT. Hence, ST gives a false alarm and does not transmit the data. Therefore, the secondary transmitter is unable to harvest energy even it consumes for spectrum sensing.

H_{01} denotes that the primary transmitter is active and the secondary transmitter is silent. Hence, in this case, secondary transmitter harvests energy from primary transmitter and also consumes energy for spectrum sensing.

H_{10} denotes that the secondary transmitter transmits data along with harvesting energy from itself. Hence, in this case, the consumed energy is utilized to perform sensing and transmitting of data.

H_{11} denotes the primary transmitter and secondary transmitter both active at a same time due to miss detection. Hence, secondary transmitter consumes energy for sensing and transmission of data. Energy harvesting is also continuing from radio frequency signals of primary transmitter and secondary transmitter.

Here, we are assuming sensing duration, $\tau \ll T_0, T_1$, where T_0 represents the time duration when primary transmitter is not transmitting any data or is idle. T_1 represents the time duration when primary transmitter is transmitting the data or is busy. Here, the time duration is too short and can be neglected. Spectrum wastes time due to false alarm and miss detection can also be neglected. Therefore, we can find the consumed energy E_c and harvested energy E_h.

$$E_c = P_s(T_0 + T_1) + \sigma_s^2\big[(T_0 - \tau)\big(1 - P_f^0(\varepsilon_0)\big) + \tau\big], \tag{9}$$

$$E_h = P_{rs}\rho(P_{rs})\big[(T_0 - \tau)\big(1 - P_f^0(\varepsilon_0)\big) + \tau\big] + T_1 P_{rp}\rho\big(P_{rp}\big), \tag{10}$$

where Ps represents the sensing power, $Prs = \sigma_s^2\sigma_{hs_2}^2$ denotes the power of secondary transmitter's signal at the energy harvester, channel gain variance from the secondary transmitter's antenna to harvesting antenna, is $\sigma_{hs_2}^2$, $Prp = \sigma_p^2\sigma_{hp}^2$ denotes the power of PT's signal at the energy harvester, and efficiency of energy harvester is denoted by $\rho(\cdot)$.

The energy efficiency η_{EE} (bits/Joule/Hz) and harvesting efficiency η_{HE} can be defined as

$$\eta_{EE}(\varepsilon_0, \varepsilon_1) = \frac{R}{E_c/T} = \frac{T\,log_2(1 + \gamma_t)\big(1 - p_f(\varepsilon_0, \varepsilon_1)\big)}{P_s(T_0 + T_1) + \sigma_s^2\alpha} \tag{11}$$

$$\eta_{HE}(\varepsilon_0, \varepsilon_1) = \frac{E_h}{E_c} = \frac{Prs\rho(P_{rs})_2\alpha + T_1 P_{rp}\rho\big(P_{rp}\big)}{P_s(T_0 + T_1) + \sigma_s^2\alpha} \tag{12}$$

where $\gamma_t = \frac{\sigma_{hs_1}^2\sigma_s^2}{\sigma_\omega^2}$ denotes the SNR in transmission, $\alpha = \alpha = (T_0 - \tau)\big(1 - p_f^0(\varepsilon_0)\big) + \tau$.

Therefore, from Eqs. (11) and (12), we can calculate the energy efficiency and energy harvesting in full-duplex CRNs.

4 Throughput in Full-Duplex Cognitive Radio Network

4.1 Non-cooperative Spectrum Sensing System Model

System model for non-cooperative spectrum sensing is shown in Fig. 2 [25]. Here, SU receiver operates in the range of PU and this operation is performed in time-slotted manner with duration of T. SU uses one transmit antenna Tx for transmission and one receive antenna Rx for the sensing during the time T. It has the capability of self-interference suppression (SIS) which allows sensing and transmitting the data at the same time and frequency.

In this model, we assume that SU uses energy detection method to perform sensing of the signal. If the Ls is the total number of samples taken and Y is the received signal then the hypothesis testing can be written as

$$Y = \frac{1}{L_s} \sum_{l=1}^{L_s} |r(l)|^2, \tag{13}$$

where number of samples $Ls = fsT$ with fs sampling rate. If $r(l)$ is the lth sample of the received signal then it is presented as
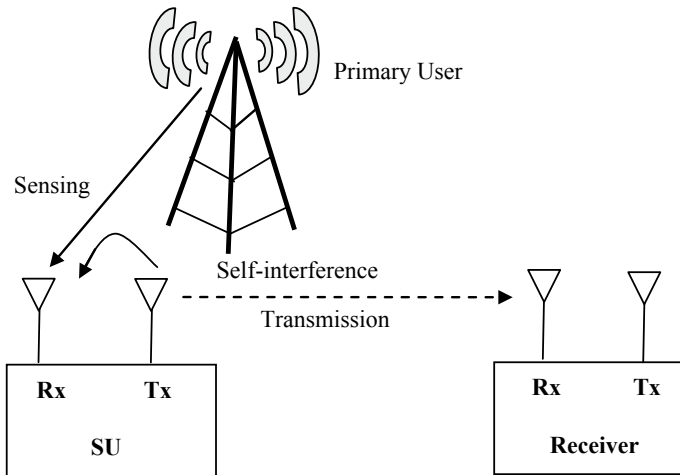
Fig. 2 Full-duplex CRN with non-cooperative system model [25]

$$r(l) = \begin{cases} \omega(l), & H_{00} \\ s(l) + \omega(l), & H_{10} \\ \chi \, p(l) + \omega(l), & H_{01} \\ s(l) + \chi \, p(l) + \omega(l), & H_{11} \end{cases}, \tag{14}$$

where $\omega(l)$ is the AWGN with variance σ_ω^2, $s(l)$ is the received primary user (PU) signal with SNR γ, χ is the signal to interference suppression (SIS) factor at SU receiver with $p(l)$ self-interference signal having zero mean σ_{SU}^2 variance. The interference to noise ratio will be as $\gamma_{SU} = \left(\chi^2 \sigma_{SU}^2 / \sigma_\omega^2\right)$. H_{00}, H_{10}, H_{01}, H_{11} are already discussed in Sect. 3.2 in detail.

Analysis of false alarm probability and detection probability

The false alarm probability and detection probability for FD-CRN can be calculated as given in equation in [24], based on signal to noise ratio (SNR) of PU and transmission power of secondary user.

Case 1: When the SU is not transmitting the data (idle state), the probability of false alarm p_f^0 and probability of detection p_d^0 will be as follows:

$$p_f^0 = Q\left(\left(\frac{\varepsilon_0}{\sigma_\omega^2} - 1\right)\sqrt{L_s}\right), \tag{15}$$

$$p_d^0 = Q\left(\left(\frac{\varepsilon_0}{(1 + \gamma_{PU})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right). \tag{16}$$

Case 2: When the SU is transmitting the data (active state) the probability of false alarm p_f^1 and probability of detection p_d^1 will be as follows:

$$p_f^1 = Q\left(\left(\frac{\varepsilon_1}{(1 + \gamma_{SU})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right), \tag{17}$$

$$p_d^1 = Q\left(\left(\frac{\varepsilon_0}{(1 + \gamma_{PU} + \gamma_{SU})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right). \tag{18}$$

Case 3: For the system, the probability of false alarm p_f and probability of detection p_d will be as follows:

$$p_f = \frac{p_f^1}{1 - p_f^0 + p_f^1}, \tag{19}$$

$$p_d = \frac{p_d^1}{1 - p_d^0 + p_d^1}, \tag{20}$$

where complementary distribution function of standard Gaussian variable is defined as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty exp\left(-\frac{t^2}{2}\right) dt$, p_d and p_f is the probability of detection and probability of false alarm.

4.2 Cooperative Spectrum Sensing System Model

System model for cooperative spectrum sensing is shown in Fig. 3 [25]. Figure 3 shows that the SU1 is assisted for spectrum sensing by (N − 1) SUs (SU$_2$, SU$_3$, …, SU$_N$). Each SU$_N$ with $n = 1, …, N$ locally sense the spectrum by its receiver antenna within the single time slot of duration T and send their sensing result to Fusion Centre (FC). FC receives the decision of all the SUn and makes a cooperative decision on presence of PU on the basis of OR fusion rule. In OR fusion rule, the FC decides the presence of PU if PU detected by any one of SU.

For n = 1, the spectrum sensing is equivalent to non-CSS. On the other hand, if n < 1, the SU1 transmitted data is assumed to interfere with SU$_n$ where $n = 2, 3, …, N$. The first sample of received SU$_n$ is given as $r_n(l)$
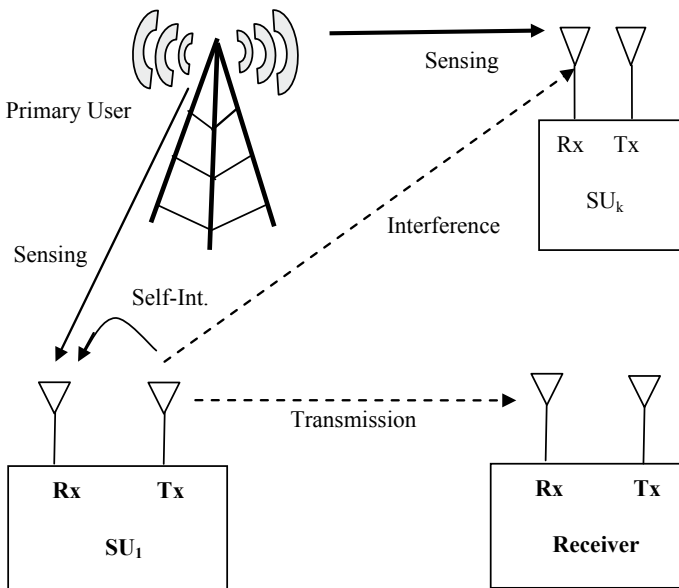
Fig. 3 Full-duplex CRN with cooperative system model [25]

$$r_n(l) = \begin{cases} \omega_n(l), & H_{00} \\ s_n(l) + \omega_n(l), & H_{10} \\ h_{1,n} \, p(l) + \omega_n(l), & H_{01} \\ s_n(l) + h_{1,n} \, p(l) + \omega_n(l), & H_{11} \end{cases} \qquad (21)$$

where $\omega_n(l)$ is AWGN channel with variance σ_ω^2, the received PU signal is $s_n(l)$ with signal to noise ratio $\gamma_{PU,n}$, and the Rayleigh channel gain from SU_1 to SU_n is $h_{1n}(l)$ with zero mean $\sigma_{1,n}^2$, variance. The transmitted signal at SU_1 is denoted by $p(l)$.

Analysis of probability of detection and probability of false alarm

The probability of detection and probability of false alarm can be calculated from [26], as the two detection thresholds, $\varepsilon_{0,n}$ and $\varepsilon_{1,n}$, are used, which determines the sensing results using energy detector scheme to decide the state of PU.

Case 1: When the SU is not transmitting the data (idle state), the probability of false alarm p_f^0 and probability of detection p_d^0 will be as follows:

$$p_{f,n}^0 = Q\left(\left(\frac{\varepsilon_{0,n}}{\sigma_\omega^2} - 1\right)\sqrt{L_s}\right), \qquad (22)$$

$$p_{d,n}^0 = Q\left(\left(\frac{\varepsilon_{0,n}}{(1 + \gamma_{PU,n})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right). \qquad (23)$$

Case 2: When the SU in transmitting the data (active state) the probability of false alarm p_f^1 and probability of detection p_d^1 will be as follows:

$$p_{f,n}^1 = Q\left(\left(\frac{\varepsilon_{1,n}}{(1 + \gamma_{SU1,n})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right), \qquad (24)$$

$$p_{d,n}^1 = Q\left(\left(\frac{\varepsilon_{0,n}}{(1 + \gamma_{PU,n} + \gamma_{SU1,n})\sigma_\omega^2} - 1\right)\sqrt{L_s}\right), \qquad (25)$$

where notations are as described in previous section. Hence, on the basis of the above analysis, we can calculate the probability of detection and probability of false alarm in both the cases of non-cooperative spectrum sensing and cooperative spectrum sensing in FD-CR networks.

5 Research Challenges and Issues

5.1 Spectrum-Related Issues and Research Directions

(1) Spectrum Sensing: Energy detection-based spectrum sensing is the most widely used spectrum sensing technique for FD-CRN. On the other hand, matched filter and cyclostationary-based spectrum sensing need more attention for various SIS and security aspects. Many spectrum sensing techniques have been proposed for FD-CRNs where frequency, time, and space dimensions are taken into account. Code and angle dimensions aspects are also needed to explore for spectrum sensing in FD-CRN [27].

(2) Spectrum Sharing: The key principle of spectrum sharing in CRNs is the dynamic spectrum sharing. Due to simultaneous spectrum sensing and data transmission, the efficiency of spectrum sensing greatly improves in FD-CRN in comparison to HD-CRN [28, 29]. In most of the existing work, the relay network is considered for spectrum sharing in FD-CRNs. However, overlay networks, multi-hop networks, and fading environments are also needed to be an explorer for spectrum sharing in FD-CRN.

(3) Spectrum Mobility and handoff: The spectrum mobility where SUs can use licensed spectrum band when PU is not active and SUs vacate the licensed band when PUs are active in licensed band, is widely explored for HD-CRN. However, the challenges arising from spectrum mobility for FD-CRNs need more attention for open research direction. For instance, PU node mobility and SU node mobility may be important issues for research in the direction of CR spectrum mobility.

Spectrum mobility is the major challenge in cognitive radio technology which is associated with spectrum handoff. Hence, spectrum handoff need more research to maintain the link performance and quality of service [19].

5.2 PU Activity

PUs follow different activity patterns like high, low, intermittent, and long [30, 31]. CRNs can operate in various PU activity patterns. SIS and spectrum sensing approaches become more effective when we consider different PU activity patterns. However, most of the current FD-CRN work focuses on ON/OFF model of PU activity [32–34]. In [35], FD spectrum sensing considering unknown PU activity has been focused. Data collected from real spectrum measurement for PU activity can be basis for further studies to enhance the underlying performance of FD-CRN.

5.3   Energy Harvesting and Green Communications

For energy-efficient wireless communication, harvesting approaches are very useful. Energy harvesting approaches, as simultaneous wireless information and power transfer (SWIPT) are widely used in FD-CRN [36]. FD-CRNs requires more energy for sensing and transmitting antenna in compared to HD systems. Hence, energy harvesting is one of the promising solutions for energy efficiency in FD-CRNs [37, 38].

Conservation of energy is also an interesting research area in wireless communications [39, 40]. Green communication is one of the innovative ideas to research while exploiting the licensed band in FD-CRN.

5.4   Self-interference Suppression (SIS)

Channel state information (CSI) can also be used for minimizing the self-interference in FD wireless communication as well as resource allocation [41]. Self-interference in FD-CRN occurs because of leakage due to the transmit chain impairments [42]. Therefore, it is required to suppress the self-interference happened due to different transmit chain leakages in FD-CRNs.

6   Conclusion

Ful-duplex CRN can improve the throughput by simultaneous sensing and transmission of data. Due to advancement in self-interference suppression (SIS) technology, FD-CRN made the more promising solution for licensed band exploitation. In this article, we have done extensive comparison of half-duplex (HD) and full-duplex (FD) communication and also covered energy efficiency analysis, energy harvesting analysis, and throughput calculations. After analyzing these parameters, we concluded that the probability of detection increased and probability of false alarm decreased significantly in FD cognitive radio networks. Analysis presented the improved energy efficiency, energy harvesting, and throughput in FD-CRNs in comparison to HD-CRN. This article gives direction to future researchers to limit the leakage due to the transmit chain impairments. Green communication is also more innovative idea to research FD-CRN. Different spectrum sensing techniques such as matched filter, cyclostationary-based sensing, code, and angle dimension need more attention.

References

1. Cisco, C.V.N.I.: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019 White Paper (2015). http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

2. Chandrasekhar, V., Andrews, J., Gatherer, A.: Femtocell networks: a survey. IEEE Commun. Mag. **46**(9) (2008)

3. Feng, M., Mao, S., Jiang, T.: Duplex mode selection and channel allocation for full-duplex cognitive femtocell networks. In: 2015 IEEE on Wireless Communications and Networking Conference (WCNC), pp. 1900–1905. IEEE (2015)

4. Rappaport, T.S., Sun, S., Mayzus, R., Zhao, H., Azar, Y., Wang, K., Wong, G.N., Schulz, J.K., Samimi, M., Gutierrez Jr., F.: Millimeter wave mobile communications for 5G cellular: It will work! IEEE Access **1**(1), 335–349 (2013)

5. Bharadia, D., Katti, S.: Full duplex MIMO radios. Self-1, A2 (2014): A3

6. Sultan, Radwa A., Song, Lingyang, Seddik, Karim G., Han, Zhu: Full-Duplex meets multiuser MIMO: comparisons and analysis. IEEE Trans. Vehicular Technol. **66**(1), 455–467 (2017)

7. Shin, K.G., Kim, H., Min, A.W., Kumar, A.: Cognitive radios for dynamic spectrum access: from concept to reality. IEEE Wirel. Commun. **17**(6) (2010)

8. Bhattarai, S., Jerry Park, J.-M., Gao, B., Bian, K., Lehr, W.: An overview of dynamic spectrum sharing: ongoing initiatives, challenges, and a roadmap for future research. IEEE Trans. Cogn. Commun. Netw. **2**(2), 110–128 (2016)

9. Rashid, Bushra, Rehmani, Mubashir Husain, Ahmad, Ayaz: Broadcasting strategies for cognitive radio networks: taxonomy, issues, and open challenges. Comput. Electr. Eng. **52**, 349–361 (2016)

10. Kiessling, M., Speidel, J.: Mutual information of MIMO channels in correlated Rayleigh fading environments-a general solution. In: 2004 IEEE International Conference on 2004 Jun 20, vol. 2, pp. 814–818. IEEE (2004)

11. Jain, M., Choi, J.I., Kim, T., Bharadia, D., Seth, S., Srinivasan, K., Levis, P., Katti, S., Sinha, P.: Practical, real-time, full duplex wireless. In: Proceedings of the 17th Annual International Conference on Mobile Computing and Networking 2011 Sep 19, pp. 301–312. ACM

12. Kim, Dongkyu, Lee, Haesoon, Hong, Daesik: A survey of in-band full-duplex transmission: from the perspective of PHY and MAC layers. IEEE Commun. Surv. Tutor. **17**(4), 2017–2046 (2015)

13. Liao, Y., Wang, T., Bian, K., Song, L. and Han, Z: Decentralized dynamic spectrum access in full-duplex cognitive radio networks. In: 2015 IEEE International Conference on Communications (ICC), pp. 7552–7557. IEEE (2015)

14. Liu, Gang, Yu, F.Richard, Ji, Hong, Leung, Victor C.M., Li, Xi: In-band full-duplex relaying: a survey, research issues and challenges. Resource **147**, 172 (2015)

15. Prasad, G., Lampe,, L.: Full-duplex spectrum sensing in broadband power line communications. In: 2017 IEEE International Symposium on Power Line Communications and its Applications (ISPLC), pp. 1–6. IEEE (2017)

16. Akyildiz, I.F., Lee, W.-Y., Vuran, M.C., Mohanty, S.: NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. Comput. Netw. **50**(13), 2127–2159 (2006)

17. Zhao, Q., Sadler, B.M.: Dynamic Spectrum Access: Signal Processing, Networking, and Regulatory Policy (2006). arXiv: cs/0609149

18. Azarfar, Arash, Frigon, Jean-Francois, Sanso, Brunilde: Improving the reliability of wireless networks using cognitive radios. IEEE Commun. Surv. Tutor. **14**(2), 338–354 (2012)

19. Arif, W., Hoque, S., Sen, D., Baishya, S.: A comprehensive analysis of spectrum handoff under different distribution models for cognitive radio networks. Wirel. Pers. Commun. **85**(4), 2519–2548 (2015)

20. Zhang, Y.: Spectrum handoff in cognitive radio networks: opportunistic and negotiated situations. In: ICC'09. IEEE International Conference on 2009 Communications, 2009 Jun 14, pp. 1–6. IEEE

21. Doan, X.T., Nguyen, N.P., Yin, C., Da Costa, D.B., Duong, T.Q.: Cognitive full-duplex relay networks under the peak interference power constraint of multiple primary users. EURASIP J. Wirel. Commun. Netw. 2017(1), 8 (2017)
22. Yin, C., Doan, T.X., Nguyen, N.P., Mai, T., Nguyen, L.D.: Outage probability of full-duplex cognitive relay networks with partial relay selection. In: International Conference on Recent Advances in Signal Processing, Telecommunications & Computing (SigTelCom), pp. 115–118. IEEE (2017)
23. Fan, M., Zhang, H., Hu, H., Wang, Q.: Detection threshold optimization for RF-powered full-duplex cognitive radio networks. In: 2016 IEEE International Conference on Communication Systems (ICCS), pp. 1–6. IEEE (2016)
24. Liao, Y., Wang, T., Song, L., Han, Z.: Listen-and-talk: Full-duplex cognitive radio networks. In: GLOBECOM, pp. 3068–3073 (2014)
25. Tuan, P.V., Koo, I.: Throughput maximisation by optimising detection thresholds in full-duplex cognitive radio networks. IET Commun. 10(11), 1355–1364 (2016)
26. Liao, Y., Wang, T., Song, L., Jiao, B.: Cooperative spectrum sensing for full-duplex cognitive radio networks. In: 2014 IEEE International Conference on Communication Systems (ICCS), 2014 Nov 19, pp. 56–60. IEEE
27. Akhtar, F., Rehmani, M.H., Reisslein, M.: White space: definitional perspectives and their role in exploiting spectrum opportunities. Telecommun. Policy 40(4), 319–331 (2016)
28. Khafagy, M.G., Alouini, M.S., Aissa, S.: Full-duplex opportunistic relay selection in future spectrum-sharing networks. In: 2015 IEEE International Conference on Communication Workshop (ICCW), pp. 1196–1200. IEEE (2015)
29. Chae, S.H., Jeon, S.-W., Lim, S.H.: Fundamental limits of spectrum sharing full-duplex multicell networks. IEEE J. Sel. Areas Commun. 34(11), 3048–3061 (2016)
30. Saleem, Yasir, Rehmani, Mubashir Husain: Primary radio user activity models for cognitive radio networks: a survey. J. Netw. Comput. Appl. 43, 1–16 (2014)
31. Canberk, B., Akyildiz, I.F., Oktug, S.: Primary user activity modeling using first-difference filter clustering and correlation in cognitive radio networks. IEEE/ACM Trans. Netw. (TON) 19(1), 170–183 (2011)
32. Tan, L.T., Le, L.B.: Distributed MAC protocol design for full-duplex cognitive radio networks. In: 2015 IEEE Global Communications Conference (GLOBECOM), pp. 1–6. IEEE (2015)
33. ElAzzouni, S., Ercetin, O., El-Keyi, A., ElBatt, T., Nafie, M.: Full-duplex cooperative cognitive radio networks. In: 2015 13th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOpt), pp. 475–482. IEEE (2015)
34. Askari, E. Aïssa, S.: Full-duplex cognitive radio with packet fragmentation. In: 2014 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1502–1507. IEEE (2014)
35. Hammouda, M., Zheng, R., Davidson, T.N.: Full-duplex spectrum sensing and access in cognitive radio networks with unknown primary user activities. In: 2016 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2016)
36. Liu, H., Kim, K.J., Kwak, K.S., Vincent Poor, H.: Power splitting-based SWIPT with decode-and-forward full-duplex relaying. IEEE Trans. Wireless Commun. 15(11), 7561–7577 (2016)
37. Xing, H., Kang, X., Wong, K.-K., Nallanathan, A.: Optimizing DF cognitive radio networks with full-duplex-enabled energy access points. IEEE Trans. Wireless Commun. 16(7), 4683–4697 (2017)
38. Zhang, R., Chen, H., Yeoh, P.L., Li, Y., Vucetic, B.: Full-duplex cooperative cognitive radio networks with wireless energy harvesting. In: 2017 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE (2017)
39. Huang, Xueqing, Han, Tao, Ansari, Nirwan: On green-energy-powered cognitive radio networks. IEEE Commun. Surveys Tutor. 17(2), 827–842 (2015)
40. Mahapatra, Rajarshi, Nijsure, Yogesh, Kaddoum, Georges, Hassan, Naveed Ul, Yuen, Chau: Energy efficiency tradeoff mechanism towards wireless green communication: a survey. IEEE Commun. Surveys Tutor. 18(1), 686–705 (2016)
41. Sabharwal, A., Schniter, P., Guo, D., Bliss, D.W., Rangarajan, S., Wichman, R.: In-band full-duplex wireless: Challenges and opportunities. IEEE J. Sel. Areas Commun. 32(9), 1637–1652 (2014)

42. Liao, Y., Song, L., Han, Z., Li, Y.: Full duplex cognitive radio: a new design paradigm for enhancing spectrum usage. IEEE Commun. Mag. **53**(5), 138–145 (2015)

Interaction of Electromagnetic Fields (100 KHz–300 GHz) Exposure with Respect to Human Body Model and Methods for SAR Measurement

Rashid Jamal, Rajiv Kumar Singh and Ekta Singh

Abstract   In the present paper, near- and far-field interaction mechanism of EM field and measurement of SAR over human body model for both near- and far-field from the electromagnetic source are described. For the modeling of equivalent human body part phantom model using Maxwell equations, parameters, like complex permittivity, polarization, internal dose, skin effect, etc., are considered. These parameters make SAR measurement easier. Further, the techniques and instruments that are used to measure the SAR for both far-field and near-field region are described. Thermal and electric probes used to measure the SAR have their remarkable boundaries while dealing with the measurement of low and high power levels. It has been identified that electric probes are used for both low- and high-power transceiver, whereas thermal probe is restricted only to high-power transceiver. Further, idea about the analysis of electromagnetic fields induced in biological tissues by thermographic camera is included.

Keywords Specific absorption rate · RF transparent temperature sensor · Electric field probe · Thermographic camera

1 Introduction

Electromagnetic field exposure causes some significant changes in physical parameter such as temperature and conductivity of biological tissues. Under the exposure condition of electromagnetic field, tissues store energy because tissues are a kind of equivalent lossy material. The stored energy inside tissue makes changes in their own physical parameters such as conductivity and temperature. Biological tissues store and dissipate the EM energy due to presence of complex permittivity [1]. Res-

R. Jamal · R. K. Singh (✉) · E. Singh
Department of Electronics Engg, Institute of Engineering and Technology, Lucknow, India
e-mail: rajivinbhu@gmail.com

Dr. APJ Abdul Kalam Technical University, Lucknow 226021, India

onance conditions are also associated with complex permittivity in tissues. Energy storing or dissipation characteristic of tissues cannot correlate directly to mass, therefore in 1981 National Council on Radiation Protection and Measurements [NCRP] has introduced a new dosimetry parameter called Specific Absorption Rate (SAR). Since 1981, SAR has been accepted widely as the quantification unit by scientific community to research over biological effects and medical applications of EM fields [2]. SAR calculation regarding far-field zone is comparatively easier than near-field because in near-field, SAR is the strong function of exposure source geometry, size, frequency, and exposure intensity. SAR measurement in near-field is very complex because significant gradient in spatial distribution of energy is possible in biological object. Probes are used to measure the SAR in field approach, but several numerical simulation methods are also available for SAR calculation such as FDTD.

2 Definition of SAR

The SAR is formally defined as the time derivative of the incremental energy (W) absorbed by (dissipated in) an incremental mass (m) contained in a volume (V) of a given density (ρ) [2, 3].

$$SAR = \frac{d}{dt}\left(\frac{dW}{dm}\right) = \frac{d}{dt}\left(\frac{dW}{\rho dV}\right) \tag{1}$$

SAR is expressed in units of watts per kilogram (W/kg). SAR definition is applied to electromagnetic fields, magnetic fields, or only to electric fields. In maximum cases, electric field dominates to magnetic field; hence, we only include the electric field in SAR calculation. SAR can be used to relate the electric field at a point for steady-state sinusoidal, as

$$SAR = \frac{\sigma}{\rho}E_{int}^2 \ \frac{W}{kg} \tag{2}$$

where

ρ is the mass density of the body ($\frac{kg}{m^3}$),
σ is the conductivity ($\frac{S}{m}$),
$Eint$ is the rms electric field strength in V/m at a point in the body.

(The subscript "int" used to emphasize the fact that the field inside the body is not the same as the external field strength surrounding an exposed object.)

SAR can be also related to define the temperature gradient at a point given, as

$$SAR = C\frac{\Delta T}{\Delta t} \tag{3}$$

where

ΔT　is the change in temperature (°C),
Δt　is the duration of exposure (seconds),
C　is the specific heat capacity (J/kg °C).

Here, it is assumed that measurements are made under "ideal" non-thermodynamic circumstances, i.e., no heat loss by thermal diffusion, heat radiation, or thermoregulation (blood flow, sweating, etc.).

3　Related Quantity to SAR Measurement

3.1　Dielectric Property

3.1.1　Complex Permittivity

Measurement of SAR is ethically not permissible over human body under RF exposure. So, we create a phantom model of human body. Permittivity and permeability of phantom must have same value as of human body. The permeability of human body is equivalent to air permeability so, we do not bother about permeability. Main issue is permittivity; human tissues are somewhere more aqueous and somewhere less aqueous in body. Therefore, we can not introduce a single permittivity constant, that is why we introduce the complex relative permittivity represented by the equation

$$\varepsilon = \varepsilon_0(\varepsilon' - j\varepsilon'') \tag{4}$$

where ε is the complex permittivity, ε_0 is air permittivity that is equal to 8.854×10^{-12} farads per meter, ε' is the real part of the complex permittivity, ε'' is the imaginary part of the complex permittivity [4].

Here, we introduce a term named loss tangent, used to measure how "lossy" or energy absorbing a material is. Loss tangent can be represented as

$$\tan \delta = \frac{\varepsilon''}{\varepsilon'} \tag{5}$$

Materials are divided into four categories on the basis of their own loss tangent [5].

(i)　Free space
(ii)　Lossless material ($\sigma \ll \omega\varepsilon$)
(iii)　Lossy material
(iv)　Conductor ($\sigma \gg \omega\varepsilon$).

If the angle of loss tangent is small, the material is said to be lossy and stores more energy in RF exposure field. Angle of loss tangent of the conductor is large
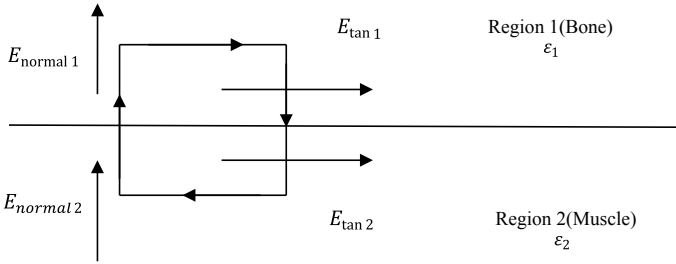
Fig. 1 Electric field boundary condition for two different dielectric layers of human body [1]

that means storing capacity of conductor is very small. Our interest lies in only to the lossy objects but not about conducting objects because our body equivalent is a lossy dielectric.

3.1.2 Boundary Condition

Appropriate boundary conditions have to be considered while dealing with the measurement of fields in a human body because of multilayered structure of human body. EM energy transfer from one layer to another (bone to muscle) is considered as energy transfer from one dielectric to another dielectric of different dielectric constants. The energy transfer mechanism must satisfy the Maxwell equations [1]. Boundary conditions for perfect dielectric are expressed as (Fig. 1).

$$E_{\tan 2} = E_{\tan 1} \tag{6}$$

$$\varepsilon_1 E_{normal1} = \varepsilon_2 E_{normal2} \tag{7}$$

Similarly, we can define the boundary condition for magnetic field [5]

$$\mu_1 H_{normal1} = \mu_2 H_{normal2} \tag{8}$$

$$H_{tan1} = H_{tan2} \tag{9}$$

3.1.3 Skin Effect

At low frequency, the wavelength of incident wave is greater than body dimension; therefore, the interdependency of different part of body exists. At low frequency, one part of the body strongly affects the other part of the body and field concentrations can be altered by combinations of size, shape, and orientation of the body in the incident

field. At intermediate frequency, field inside body is considered as propagating wave. Presence of multilayer structure inside body causes partial reflection and transmission of wave at boundaries. Also, at these intermediate frequencies, E and H cannot exist separately and are strongly coupled together. At very high frequency, the wavelength of incident wave is comparatively smaller than size of object (body part). Short wavelength EM waves can be thought as rays [1, 2]. High-frequency incident wave only produces skin effect to lossy object. At high frequency, the depth of penetration decreases rapidly and produces only skin effect. Skin effect depends upon frequency and conductivity both. Depth of penetration can be represented by

$$\delta = \frac{1}{\sqrt{(\pi f \mu \sigma)}} \tag{10}$$

where δ is depth of penetration, σ is conductivity, μ is permeability, and f is frequency of incident wave [6].

3.2 Internal Dose

Dose (specific absorption) is defined as "total amount of energy stored by biological object in EM field exposure per unit mass" [9]. $E_{incident}$ is aligned with the long axis of the object, strong electric fields are coupled into the object and thus stores more amount of energy as shown in Fig. (2a). Where $E_{incident}$ is normal to the long axis of the object, weaker electric fields are coupled into the object and thus stores less amount of energy as shown in Fig. (2b).

Lossy object (human muscle) behaves like a conductor at very low frequency and circulates an eddy current in magnetic field coupling. Eddy currents circulate in closed path around the incident magnetic field vector. Eddy current is used to represent the energy transfer mechanism from magnetic field vector to the object.

In some sense, the magnitude of the energy transferred is proportional to the cross-sectional area intercepting H_{inc}. The magnitude of energy transfer is greater in the case of Fig. (3b) than that of Fig. (3a) because in Fig. (3a) incident magnetic field acquire less area [10].

Fig. 2 Dielectric object placed in electric field [1]
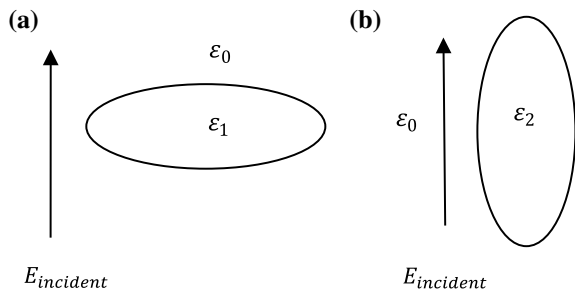
Fig. 3 Dielectric object
placed in magnetic field [1]

(a)

μ_1

$H_{incident}$

(b)

μ_2

$H_{incident}$
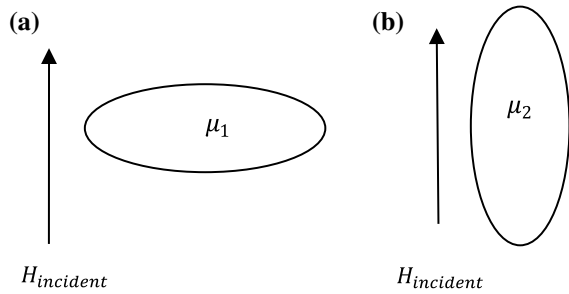
3.3 Polarization

For E-polarization, electric field is parallel to long axis of object; therefore, E and H coupling are strong. For H plane polarization, the magnetic field is parallel to long axis of the object; therefore, both E and H coupling are weak. For K-polarization where propagating vector K is parallel to the long axis of the object; therefore, E coupling is weak and H coupling is strong. Thus, for frequencies below resonance, the SAR is greatest for E-polarization, least for H-polarization, and intermediate for K-polarization [1].

3.4 Biological Object Geometry and Size

Under normal exposure condition, maximum value of local SAR is usually at or near the surface of biological object. But statement is not equally considerable for curved surfaces and resonant object. In curved surfaces or resonant objects high SAR (hot spot) exist at various locations. A human body has various geometrical configures of body parts. Human body is such a complex biological system regarding multilayered configurations of tissues. Each layer has different dielectric properties and causes different boundary conditions for EM exposure. In exposure conditions, field may propagate inside biological object in form of EM waves. Because of the presence of different dielectric layers with different dielectric properties, EM waves configure as in reflection, transmission, and standing waves. Different dielectric properties causes impedance mismatching; that's why, wave transmission is not possible completely. Fat thickness, tissue curvature, and dimensions of the body, limbs, and head relative to the wavelength, all affect the energy distribution [11, 12].

3.5   Field Frequency

Frequency-dependent quantities related to human body model are dielectric properties, the strength, and spatial distribution of internal fields. A general consideration about variation of average SAR with frequency for human-sized sphere is

(i)   *rage SAR* $\propto f^2$, at low frequency
(ii)   *average SAR* $\propto f$, at intermediate frequency
(iii) *(average SAR)$_{max}$*, at the resonance frequency

At resonance, the length of the long axis of the exposed body is approximately four-tenths of the field wavelength [13].

3.6   Source Configuration

In far-field, SAR measurement is typically independent of source configuration, because after $\frac{2D^2}{\lambda}$ there is no coupling of energy between source and object (Fig. 4). In near-field configuration, energy coupling depends upon source shape and size. For example, the SAR distributions of waveguide hyperthermia applicators are the strong function of source configuration [14].

3.7   Time Intensity Factors

Exposure intensity and duration both are the important parameters that are used to determine the total energy absorbed by tissue. Modulation of signal makes change in frequency, therefore time-averaged SAR also varies with modulation scheme. Long exposure duration or large intensity causes large amount of energy storage in biological object [2].
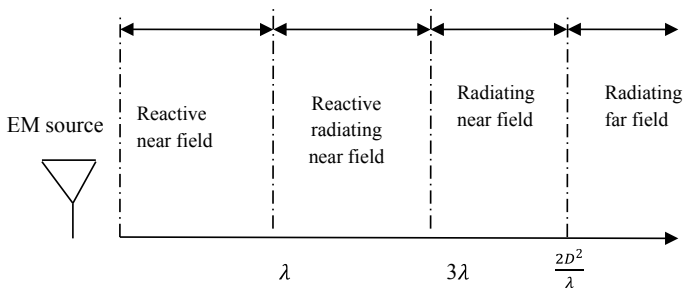
Fig. 4   Region around electromagnetic source [8]

3.8 Exposure Environment

The amount of energy stored in tissue also depends upon the exposure environment. Energy storage may be altered by nearby placed another biological object or any conducting reflector surface, etc.

3.9 Local SAR

Local SAR is used to define the magnitude of SAR in a small portion of an exposed biological object.

3.10 Whole Body Averaged SAR

It is a single SAR value that represents the magnitude of the spatially averaged SAR throughout an exposed biological object.

4 Methods for SAR Calculation

(SAR calculation is meaningful over 100 kHz to 6 GHz frequency range only)

Internal fields induced inside any object in EM fields by irradiation can be calculated by solving Maxwell's equation. In real time practice, calculation regarding internal field is very difficult, and could be done only for few special cases like as spheres or infinity long cylinders. Because of mathematical complexities appearing in SAR calculation; there are no general techniques that calculate SAR over whole frequency range. There are several techniques that calculate SAR for various models as functions of frequency. Each of these techniques has their own limitations. Combination of these techniques provides sufficient information of SAR as a function of frequency over a wide range of frequencies for a number of useful models. General considerable instrument used to measure SAR over specified frequency range is listed in Table 1.

Table 1 Distribution of SAR measuring instrument over frequency range [1]

S. No	Range of frequency	Instrument used
1	100 kHz to 6 GHz	RF transparent temperature sensor
2	300 MHz to 3 GHz	Electric field probe (using phantom)
3	Above 6 GHz	Thermographic camera

4.1 Quantities and Parameter to Be Measured

There is no instrument available, that measures the SAR directly. Measurement of SAR can be done by measuring either one or more component of electric field strength (E) r magnetic field strength (H) or both and then infer with to find an equivalent power density from the far-field. Theoretically, we first measure the power density (S) and then correlate with SAR. General equation of power density in far-field is given by

$$s = \frac{|E|^2}{120\pi} \quad or \quad \frac{|H|^2}{120\pi} \text{ W/m} \tag{11}$$

Although power density is difficult to determine for unknown polarization, but it is easier if polarization is known. Widely accepted quantity that is used to indicate the hazardous exposure is power density. SAR measurement in far-field is easier than near-field because of the presence of plane waves. Plane wave parameter is simply related to power density without loss of generality in far-field but it is not possible to apply same to near-field. In near-field configuration, SAR measurement is complex, because of the presence of scattering, reflection, or multiple sources. Therefore, we must have to measure the E or H field component individually and then correlate to SAR. In many cases, measurement will be required in reactive or radiating near-field. In reactive near zone, standing wave exists where time-averaged power density is not a valid unit to identify hazard [1, 6, 8].

4.2 General Considerations About Phantom Modeling

For health risk assessment, evaluation of SAR or induced current density in human body (body in high-frequency electromagnetic field) is necessary. Evaluation of electric field strength or temperature elevation in the actual human body is difficult using noninvasive methods. Therefore, a typical surrogate of human body is prepared designated as "phantom". Using phantom for RF, we simulate the electric properties equivalent to those of the human body [3]. Phantom can be modeled using various kinds of materials. Required material permittivity and conductivity are approximate to human body complex permittivity and conductivity. Liquid-based phantom model is widely used for RF range, because it provides some tremendous advantage comparatively, like (i) easy to prepare these materials (ii) electrical properties adjust easily (iii) easy to scan electric field sensors. Liquid phantom has disadvantage regarding electrical stability. It provides poor electrical stability due to presence of water evaporation. Although dry phantom provide fine stability, but its preparation requires complex and skilled procedures and high cost. Required phantom types depend over types of tissues. Highly water content tissue like muscle is modeled using liquid phantom. Low water content tissues such as fat and bone are usually modeled by dry

material. Measurement regarding effect of electromagnetic field over human body is a tough task because of heterogeneous structure of body tissues. To make measurement easy, generally, homogenous tissues are used for physical phantom like a standard head phantom for compliance tests of cellular phones or full-size models of human body. Difficulties for preparation and maintenance of the phantom are (i) adjustment of both imaginary and real part of complex permittivity of the phantom is not easy, (ii) temperature changes may lead to water evaporation that affects electrical properties of the phantom, and (iii) the uncertainty of the electrical properties measured by commercially available systems is sometimes considerable.

4.3 Problem Associated with RF Exposure Measurement

Following are the problems associated with RF exposure measurement:

i.   Source parameters such as modulation scheme, frequency, radiation pattern, and polarization;
ii. Interference pattern;
iii. Radiation leakage;
iv. Reactive near-field; and
v.   Passive, parasitic, or secondary re-radiators.

4.4 Instrumentation

Instruments for measuring SAR are implantable E-field probes and implantable temperature probes. Probe is an electronic device that has capability to measure the radio frequency field with negligible field perturbation. A probe contains sensor, high impedance connection, and conducting circuitry essentially. Sensor is used for detecting the radio frequency field component. High impedance connections are used to extract the rectified signal.

4.4.1 Implantable E-Field Probes

Implantable E-field probes are used to measure the local electric field strength in biological object at a specific point. If field polarization is unknown, device must be isotropic in nature and it's all dipoles are arranged orthogonally. Isotropic nature of device makes measurement independent regardless of orientation of field. Anisotropic devices are used where orientation of field is known. Implantable E-field probes with 1.5–2.5 mm long dipoles have been produced in small quantities by several commercial companies. Reviews of the theory of this type of probe are discussed by Bassen and Smith and Schmid [15] (Fig. 5).

Fig. 5 Elements of a typical implantable E-field probe (single-axis) [8]
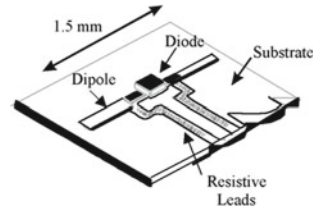
In order to make SAR measurement accurate using implantable E-field probe, calibration of probes is required. Two fundamental requirements that make calibration technique accurate are (i) model should be large compared to the length of the probe sensors (dipoles) (ii) SAR should be uniform throughout its volume. Calibrations are performed in spheres and in waveguides filled with lossy dielectric liquids. Uncertainty in calibration is typically 1–2 dB in any lossy, high dielectric constant object [15–18].

4.4.2 Implantable Temperature Probes

For successful measurement of SAR using implantable temperature probes, minimum requirements are EM field should not perturb by temperature sensor and their associated leads and the value of SAR should be large enough to produce a measurable temperature gradient during a period of less than 30 s. In order to achieve nonperturbing field, leads material should be highly resistive or optical fiber should be used instead of metallic lead. For single point SAR measurement, usually, microwave transparent temperature probes are used. Easily available temperature probe in market is Vitek-101 high-resistance-lead thermistor probe (carbon-loaded Teflon high-resistance leads) and Luxtron Model 3000 multi-channel fiber optic thermometer (temperature-sensitive phosphorus for sensor, and the leads are plastic optical fibers) [19, 20].

4.5 Measurement Techniques

Various tissue geometries and gross anatomies of human can be fabricated by phantom model of same dielectric property. Phantom model is fabricated such that it includes stratified layer of muscles and fat of various dimensions, and geometry like circular and irregular cylindrical structure for muscles, bones, and fat, and spherical for brain tissues. Liquids are poured to fill rigid, outer shells of simulated subcutaneous fat in the shape of a full-size human [7]. Liquid phantom model makes local SAR mapping through miniature E-field probe easy because movement of probe through continuous path is easy due to the use of mechanical scanner.

4.5.1 SAR Measurement with Implantable Temperature Probe

SAR measurement using temperature probe is possible because biological object temperature increases linearly during irradiation of high-intensity EM waves. To detect the SAR correctly using temperature probe, significant amount of positive temperature gradient is required for a duration of 30 s. Irradiation of tissue equivalent model is performed at high exposure intensity for short duration, therefore, temperature probe method is not suitable with handheld transceiver (mobile phone) [21–23]. For accurate measurement of SAR using temperature probe, irradiation duration must be in the range of 5–30 s and temperature rise in object should be such that there is no significant thermal conductive loss. Temperature rise during irradiation in tissue is kept below 10 °C from initial value of the temperature to prevent increase in tissue conductivity (approximately 2% per Celsius). Due to thermal runway conductivity of tissue alter because of temperature gradient. Some dielectric parameters also change the absorption rate of tissue. Therefore, the exposure intensity must be high for short duration to minimize heat diffusion. Implantable temperature probe techniques are generally used for SAR measurement in tissue and vitro cultural system. In biological system thermoregulatory effect of blood flow also reduces SAR accuracy.

To make SAR measurement accurate, initial rate of temperature rise determination is essential. Experimental setup used for thermal SAR measurement is shown in Fig. 6.

4.5.2 SAR Measurement with Miniature Implantable Electric Field Probe

Miniature isotropic E-field probe has much higher sensitivity than thermal probe. These types of probes are especially suitable for E-field measurement within high water simulated or biological tissues such as brain and muscle. It can measure the SAR accurately in the domain of E-field up to 10 mW/kg. Biological object exposed in the near- or far-field has possibility to larger spatial gradient and standing waves
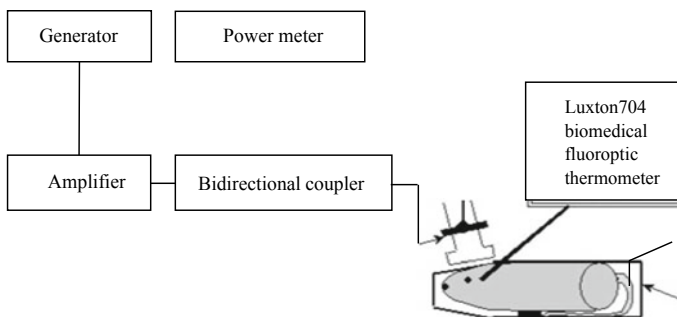
Fig. 6 Experimental setup used for thermal SAR measurement [27]
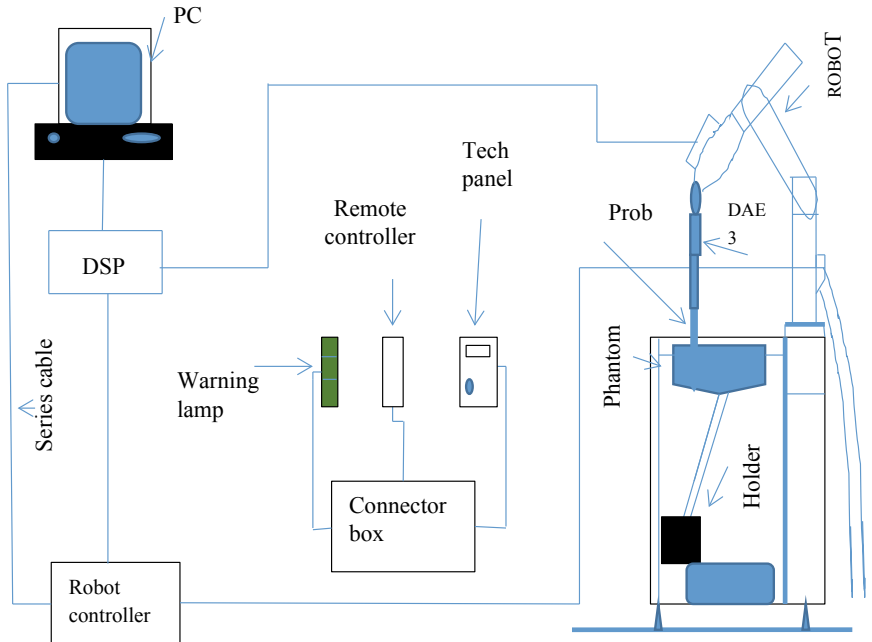
Fig. 7 SAR measurement system using DASY-6 [29]

existing within the object. Therefore, SAR measurement within simulated or actual biological object must be made at each individual point and at sufficient number of data points to make SAR measurement accurate. At each individual site, responses of three mutually orthogonal dipoles are added and location of probe must be recorded to prevent dislocation of site during repeatable measurement. The accurate position of sensing probe is achieved by using automated 3-D positioner. Sensing probes are moving along a continuous path , to scan the whole volume. E-field probe has a response time of the order of a few milliseconds. High sensitivity of electric field probes make them to measure the SAR of low power transceiver, e.g., cellular and personal communication equipment [24] (Fig. 7).

DASY6 is SAR assessment system used in SAR measurements. DASY6 is the latest generation of DASYX family and provides high precision.

4.5.3 SAR Measurement with Thermographic Camera

Exposure source frequency above 6 GHz does not impact up to full depth of biological object. Above 6 GHz frequency, depth of penetration reduces drastically and causes skin effects only. EM wave in millimeter range is attenuated in objects because biological tissues behave like reflecting surface at high frequency. Above 6 GHz
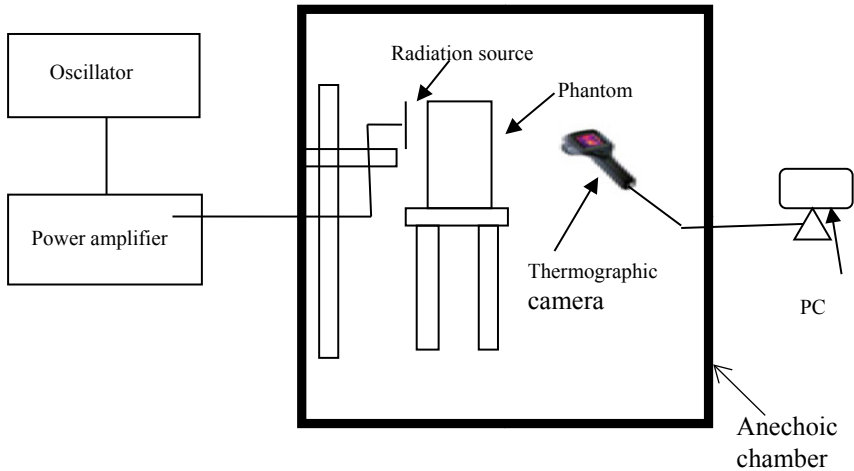
Fig. 8 SAR measurement system [28]

frequency, exposed field energy is stored at surface, therefore, thermographic camera is used to measure the SAR above 6 GHz [25] and [26] (Fig. 8).

Measurement Procedure

Thermographic method for SAR measurement is pertinent to both far and near-field zone. In this method, thermographic camera is used to record the spatial distribution of temperature inside phantom model. Under radiating field the phantom model of heads absorbs exposure energy. Absorbed electromagnetic energy is converted into thermal energy and produces spatial distribution of temperature through the phantom. Equation (3) is used to calculate the temperature distribution inside phantom. According to Eq. (3) SAR distribution is proportional to the temperature rise.

Figure (9) shows the procedure that is used to measure the SAR using thermographic camera. Measurement system includes thermographic camera, a phantom, an antenna, oscillator, radio anechoic chamber, computer, and a power amplifier. Solid phantom is preferred in the thermographic method rather than liquid phantom. Measurement procedures are as follows;

(1) Equivalent phantom model with uniform temperature is placed in a radio anechoic chamber and exposed to electromagnetic waves (above 3 GHz) by a nearby source for 2 min or so. The exposure duration is determined to yield in a temperature rise of at least 1 K.
(2) Phantom is split into two parts.
(3) After the exposure period, the phantom is opened quickly in front of a thermographic camera.
(4) Thermographic image is immediately captured to find the temperature rise on phantom surface.
(5) After this, using Eq. (2), find the SAR directly.
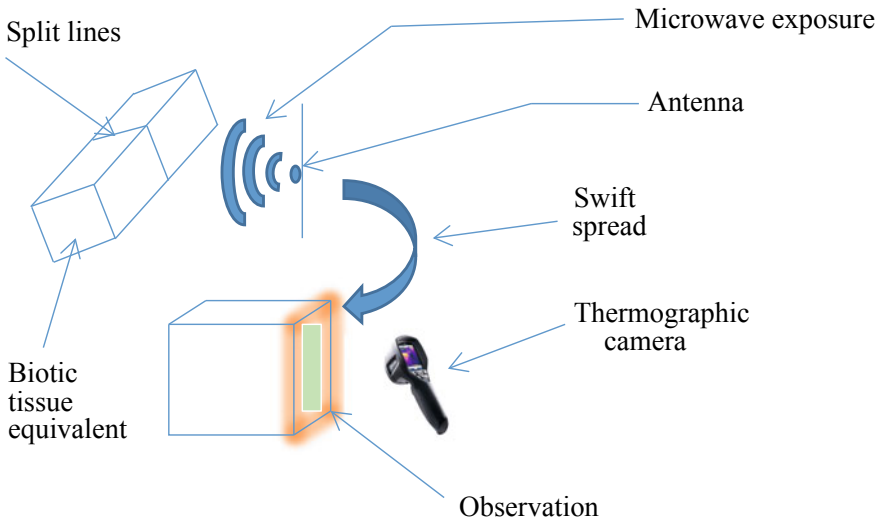
Fig. 9 SAR measurement procedure [28]

5 Conclusion

SAR is the most important parameter that defines electromagnetic energy stored within biological object. Large spatial gradient in stored energy in near-field zone makes the SAR calculation complex. Boundary condition and Maxwell equation are also essential for SAR calculation. For the measurement of EM exposure, electric probe is more versatile than thermal probe. Thermographic method is also useful above 6 GHz.

References

1. IEEE International Committee.: Practice for Measurements and Computations of Radio Frequency Electromagnetic Fields With Respect to Human Exposure to Such Fields, 100 kHz–300 GHz, IEEE Std. C95.3™-2002 (R2008)
2. Chou, C.Kc, Bassen, H., Osepchuk, J., Balzano, Q., Petersen, R., Meltz, M., Cleveland, R., Lin, J.C., Heynick, L.: Radio frequency electromagnetic exposure: Tutorial review on experimental dosimetry. Bioelectromagnetics **17**(3), 195–208 (1995)
3. NCRP.: Radiofrequency Electromagnetic Fields—Properties, Quantities and Units, Biophysical Interaction, and Measurements NCRP Report 67, National Council on Radiation Protection and Measurements, Washington, DC (1981)
4. Hayt Jr, W.H., Buck Title, J.A.: Engineering Electromagnetism, 6th edn., McGraw-Hill (2012)

5. Sadiku, M.N.O.: Principle of Electromagnetism, 4th edn. Oxford
6. WHO.: Environmental Health Criteria 137 Electromagnetic Field (300 Hz to 300 GHz) EHC 137 (2009) (1992)
7. TEC India.: Measurement of Electromagnetic Fields from Base Station Antenna, TEC/TP/EMF/001/02 (2012)
8. ICNIRP.: For limiting exposure to time varying electric, magnetic and electromagnetic fields (UP TO 300 GHZ), Health Phys. **74**(4):494–522 (1998)
9. Ahmad, S.N.: Physics and Engineering of Radiation Detection, 1st edn. Elsevier, Netherland (2007)
10. Guy, A.W., Lin J.C.: Quantities, Units, and Measurement Techniques for RF and Microwave Fields in Biological Tissues, URSI Fall Meeting, Williamsburg, VA (1972)
11. Johnson, C.C., Guy, A.W.: Nonionizing electromagnetic wave effects in biological materials and systems. Proc. IEEE **60**(6), 692–718 (1972)
12. Durney, C.H., Massoudi, H.: Radiofrequency Radiation Dosimetry Handbook, 4th edn. (Report TR-85-73). Brooks Air Force Base. USAF School of Aerospace Medicine, TX (1986)
13. Durney, C.H., Johnson, C.C.: Radiofrequency Radiation Dosimetry Handbook, 2nd edn. (Report SARTR- 78–22). Brooks (1978)
14. Chou, C.-K.: Evaluation of microwave hyperthermia applicators. Bioelectromagnetics **13**(6), 581–597 (1992)
15. Bassen, H., Smith, G.: Electric field probes–A review. IEEE Trans. Antennas Propag. **31**(5), 710–718 (1983)
16. Bassen, H., Herchenroeder, P., Cheung, A., Neuder, S.: Evaluation of an implantable electric-field probe within finite simulated tissues. Radio Sci. **12**(6S), 15–25 (1977)
17. Hill, D.A.: Waveguide technique for the calibration of miniature implantable electric-field probes for use in microwave-bioeffects studies. IEEE Trans. Microw. Theory Tech. **30**(1), 92–99 (1982)
18. Stuchly, M.A., Kraszewski, A., Stuchly, S.S.: Implantable electric-field probes-some performance characteristics. IEEE Trans. Biomed. Eng. **7**, 526–531 (1984)
19. Bowman, R.R.: A probe for measuring temperature in radio-frequency-heated material (short papers). IEEE Trans. Microw. Theory Tech. **24**(1), 43–45 (1976)
20. Wickersheim, K.A., Sun, M.H.: Fiberoptic thermometry and its applications. J. Microw. Power Electromagn. Energy. **22**(2), 85–94 (1987)
21. Balzano, Q., Garay, O., Steel, F.R.: Heating of biological tissue in the induction field of VHF portable radio transmitters. IEEE Trans. Veh. Technol. **27**(2), 51–56 (1978)
22. Balzano, Q., Garay, O., Steel, F.R.: Energy deposition in simulated human operators of 800-MHz portable transmitters. IEEE Trans. Veh. Technol. **27**(4), 174–181 (1978)
23. Balzano, Q., Garay, O., Steel, F.R..: An attempt to evaluate the exposure of operators of portable radios at 30 MHz. In: Vehicular Technology Conference, 1979. 29th IEEE, Vol. 29, pp. 187–188. IEEE (1979)
24. Guy, A.W., Chou, C.-K.: Specific absorption rates of energy in man models exposed to cellular UHF mobile-antenna fields. IEEE Trans. Microw. Theory Tech. **34**(6), 671–680 (1986)
25. Guy, A.W.: Analyses of electromagnetic fields induced in biological tissues by thermographic studies on equivalent phantom models. IEEE Trans. Microw. Theory Tech. **16**(2), 205–214 (1971)
26. Johnson, C.C., Guy, A.W.: Nonionizing electromagnetic wave effects in biological materials and systems. Proc. IEEE **60**(6), 692–718 (1972)
27. Lopresto, V., Pinto, R., De Vita, A., Mancini, S., Galloni, P., Marino, C., Ravazzani, P., Lovisolo, G.A.: Exposure setup to study potential adverse effects at GSM 1800 and UMTS frequencies on the auditory systems of rats. Radiat. Prot. Dosimetry. **123**(4), 473–482 (2006)
28. Wang, J., Lim, E., Leach, M., Wang, Z., Man, K., Huang, Y.: Two methods of SAR measurement for wearable electronic devices. In: Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol. 2 (2016)
29. s p e a g.: https://speag.swiss/products. Last accessed 2018/9/30

Noise-Induced Training for Weak Signal Detection in Neyman–Pearson Framework

Sumit Kumar, Ayush Kumar and Rajib Kumar Jha

Abstract Here, we propose a noise-induced neural network-based detector. The suggested method performs well in the detection of the known weak DC signal in additive to non-Gaussian noise. The precalculated noise is added in a neural network which helps in boosting the performance of the weak signal detector. This precalculated noise boosts the training process for signal detection. While training, the backpropagation (BP) algorithm acquires less error and it converges faster with the addition of the external noise. This method performs better than the traditional neural network-based detector in terms of its performance characteristics, i.e., the probability of detection (P_D) at a fixed value probability of false alarm (P_{FA}). We also test our noise-induced proposed detector under several signal-to-noise ratio (SNR) environments. The different state-of-the-art techniques have been compared with our proposed method.

Keywords Binary hypothesis · Signal detection · Neyman–Pearson framework

1 Introduction

The detection of a weak signal from noisy data is an interesting area of statistical signal processing. In different practical applications such as radar, biomedical, and underwater object detection the absolute signal gets corrupted with non-Gaussian noise. The performance of the detector is mostly degraded by natural environmental

Indian Institute of Technology Patna, Bihta, Bihar–801103.

S. Kumar (✉) · A. Kumar · R. K. Jha
Indian Institute of Technology Patna, Patna, Bihar, India
e-mail: sumitphd13@gmail.com

A. Kumar
e-mail: ayushkumar94314@gmail.com

R. K. Jha
e-mail: jharajib@gmail.com

sources such as rain, thundering, and lighting.These kinds of impulsive noises affect the signal detrimentally which cannot be recognized by the detectors. However, most of the desired optimum detectors such as uniformly most powerful (UMP), UMP unbiased, or invariant [1] do not exist for detection of the signal present in non-Gaussian noise.

The neural network (NN) provides the accountability for the robust detection. Therefore, it has been broadly studied and proposed for application in the various areas of signal processing. NN-based detector is essentially nonparametric, making no assumption about the noise. This is the basic reason about the robustness and the capability of detection of the weak signal in the non-Gaussian noise environment.

Poor [2] proposed an optimum linear detector which works in Gaussian noise. However, it suffers when the inherent noise is non-Gaussian. For small-signal strength and large sample size, the locally optimum detectors are used by Kassam [3]. Whereas for band-limited Gaussian noise, Watterson [4] proposes a NN-based model which consists of multilayer perceptron (MLP). Similarly, Lippmann and Beckman [5] employ a NN-based detector to enhance the performance of the matched filter in the presence of impulsive noise. Similarly, Michalopoulou et al. [6] train an MLP for the weak signal detection present in non-Gaussian noise whereas Gandhi et al. [7] discuss the performance of the NN-based detector which uses backpropagation (BP) algorithm. This detector performs better than the matched filter (MF) detector in both Gaussian and non-Gaussian noise.

However, in the paper [8], the measure-transformed Gaussian quasi-likelihood ratio test (MT-GQLRT) has been proposed for weak signal detection. It leads to a significant increase in decision performance whereas, in paper [9], a new optimality criterion has been proposed. The figure of merit has been considered to design the detector. However, Naderpour et al. [10] introduce the generalized Wald test (GWT) for a wide class of composite problems. The test statistics leads to the optimally uniform most powerful unbiased detector. However, for low SNR signal, the performance is slightly inferior. Moreover, Ma et al. [11] propose a novel algorithm for the detection of the compressed signal with low signal-to-noise ratio (SNR) value. But this algorithm performs poorly when the non-Gaussian is present as inherent noise. So, understanding the statistical characteristic of noise decides the performance of the detector.

However, the paper [12] added some uniformly distributed external noise to speed up the convergence in backpropagation (BP) training algorithm. Similar kinds of literature have been explored with an example [13–15].

Furthermore, the various methods such as matched filter (MF) [16], Guo et al. [17] and Urkowitz [18] perform poorly at low SNR in non-Gaussian noise. The general block representation of the detector has been shown in Fig. 1.

Organization of Paper: Section 2 deals with the basics of detection, internal noise, and the test statistics of the different methods whereas Sect. 3 shows the training of the proposed neural network. In Sect. 4, we discuss the results obtained from the different state-of-the-art methods. Finally, we conclude the paper in Sect. 5.
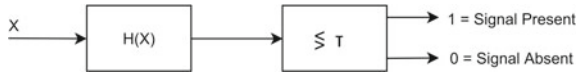
Fig. 1 Block diagram of a signal detector. Here, $H(X) = H_{NN}(X)$ broadly represents test statistics discussed in Eq. 8. T is the threshold chosen

2 Basic Mathematics

Consider a data vector $X = [X_1, X_2, \ldots, X_p]^T$ $(p \times 1)$ is input to the NN-based detector as shown in Fig. 1. With the observation model, we can formulate binary detection formula as follows:

$$X = \lambda s + N, \tag{1}$$

where $s = [s_1, s_2, \ldots, s_p]^T$ is the known signal vector, $N = [N_1, N_2, \ldots, N_p]^T$ is the noise vector, and λ represents the signal strength parameter. The case $\lambda > 0$ and $\lambda = 0$ refer to the presence and absence of a known signal s in the noisy sample X, respectively.

For performance analysis of proposed detector, the probability of detection (P_D) can be given as $P_D = Pr$ [signal present $|\lambda > 0$]. Similarly, the probability of false alarm (P_{FA}) can be stated as $P_{FA} = Pr$ [signal present $|\lambda = 0$]. For a signal detector defined in Fig. 1, the test statistics $H(X)$ is computed first and then it is compared with a threshold τ. If $H(X) > \tau$ then a signal presence is declared. Thus, P_D and P_{FA} are given as follows:

$$P_D = Pr[H(X) > \tau | \lambda > 0]. \tag{2}$$

$$P_{FA} = Pr[H(X) > \tau | \lambda = 0]. \tag{3}$$

For the model discussed in Eq. 1, test statistics $H(X) = T_{LR}(X)$ of the optimum likelihood ratio (LR) detector is given by

$$T_{LR}(X) = \prod_{i=1}^{i=p} \frac{f_{X_i}(x_i | \lambda > 0)}{f_{X_i}(x_i | \lambda = 0)}, \tag{4}$$

where $f_{X_i}(x_i)$ is the probability density function of data X_i. The use of a likelihood ratio test (LR) detector is very limited to a specific application because of unknown parameters. But generally, linear matched filter (MF) is used because of its computational simplicity. For MF, the test statistics is a linear combination of the p observations and can be written as follows:

$$T_{MF}(X) = \sum_{i=1}^{p} s_i X_i. \tag{5}$$

For the locally optimum (LO) detection, the test statistics $T_{LO}(X)$ which works better for non-Gaussian noise is given as follows:

$$T_{LO}(X) = \sum_{i=1}^{p} s_i \frac{f_N'(X_i)}{f_N(X_i)}, \tag{6}$$

where $f_N'(X_i) = \frac{df_N(X_i)}{dx}$. This LO detector behaves as a matched filter if the noise in Gaussian. But in case, the noise is non-Gaussian, the test statistics will be some nonlinear function of X.

3 Proposed Neural Network-Based Weak Signal Detection Algorithm

Here, we discuss the proposed noise enhanced neural network-based detector and conditions of improvability and non-improvability when SR noise is added. The inherent noise considered is non-Gaussian. We explore the optimum noise for the proposed detector.

The proposed detector has three layers. The three layers (input, hidden, and output) have p, q, and one node, respectively. The different weights have been shown in Fig. 2.

We employ the three-layered neural network as shown in Fig. 2 for detecting DC signal present in the different non-Gaussian noises. This neural network has $(p + 1)$ input nodes including one bias node, one hidden layer of $(q + 1)$ nodes including one bias node, and one output node. The bias nodes are unit-valued nodes and included
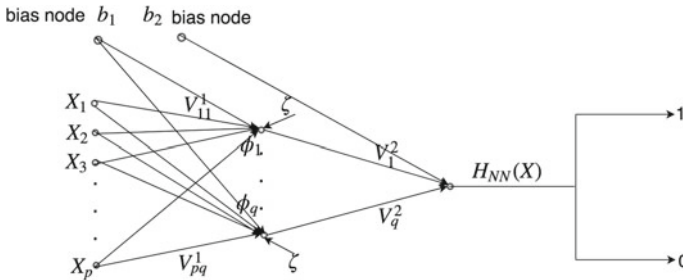
Fig. 2 Three-layers neural network. Input layer contains p nodes, the hidden layer contains q nodes whereas the output layer contains only one node. Here, b_1 and b_2 are the bias nodes to the hidden, and output layer, respectively. V_{ij}^1 and V_j^2 are the weight of the link from ith input node to jth node of the hidden layer and jth node of the hidden layer to output node, respectively, where $i = 1, \ldots, p$ and $j = 1, \ldots, q$. $\phi_1, \ldots, \phi_q$ are the weighted sum input applied to the activation function (See Eq. 7.) of the individual node of hidden layer. ζ is the external noise which is added when the backpropagation is applied for minimization of the error. The expression of $H_{NN}(X)$ is given in Eq. 8

as a part of the network for additional flexibility. Branch weights between ith input node and jth hidden node and are given by V_{ij}^1 whereas V_j^2 is between jth hidden node and output node where $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$. This produces a total of $(p + 2)q + 1$ real-valued network weights. In the hidden layer, the input $(\phi_1, \ldots, \phi_q)$ is passed through the sigmoid activation function, which is defined as follows.

$$a(t) = \frac{1}{1 + exp(-Rt)}, \quad -\infty < x < \infty \tag{7}$$

where t is defined as the sum of weight values to the jth hidden node. We consider $R = 10$ in our simulation. With inputs $X_1, \ldots, X_p$, the neural network-based detectors test statistic $H_{NN}(X)$ is given by

$$H_{NN}(X) = a \left[\sum_{j=1}^{q} V_j^2 a \left(\sum_{i=1}^{p} V_{ij}^1 X_i + V_j^{b_1} \right) + V^{b_2} \right], \tag{8}$$

where a is activation function defined in Eq. 7, $V_j^{b_1}$ is the weight of bias node b_1 to jth node of hidden layer and V^{b_2} is the bias weight to the output node. The statistics H_{NN} is compared with a real-valued threshold τ_{NN} for the decision of presence or absence of a signal in noisy data.

For training of the proposed noise enhanced neural network-based detector, we employ the backpropagation algorithm with the target of 0 and 1 for the absence and presence of signal, respectively. Both signal-plus-noise and only noise vectors are provided to the network as an input. During each training epoch, network weights are updated based on cost, which is given below

$$J = \frac{1}{2K} \sum_{k=1}^{p} (t_k - H_{NN}(X))^2, \tag{9}$$

where t_k is the target, p is the total number of inputs to the neural network and $H_{NN}(X)$ is the test statistics given in Eq. 8.

3.1 Noise Benefitted Neural Network Detector

Here, we discuss how the noise is added to the neural network-based detector and also exhibits the beneficial effect of the noise. Different strategies for the addition of external noise in the neural network have been discussed [12]. They mention that the backpropagation is basically a generalized expectation maximization (GEM) algorithm, which suggests that the noise assists in speeding up the expectation maximization (EM) algorithm. This is termed as noise expectation maximization (NEM) algorithm [14, 15]. The basic idea behind the NEM is that the noise increases the

average positivity condition, i.e., it reduces the error and increases the detection performance. Here in our proposed method, we add noise to the hidden layer when backpropagation is applied. Thereafter, we evaluate the weights of the proposed noise enhanced neural network detector.

Noise expectation maximization (NEM) algorithm supports the EM algorithm such that it increases the P_D at a constant value of P_{FA}. The NEM algorithm adds noise to the data at each EM iteration and it must satisfy the positivity condition. However, if noise satisfies the positivity condition, such noise is beneficial for us in order to increase the detection performance. Here, in this paper, we add the noise in the similar fashion as shown in Fig. 2.

The NEM idea suggests that when certain noise ζ is added, the probability of the signal gets increased [15]. Let us assume that the signal during the backpropagation is shown by y. Mathematically, we can write it as $p(y + \zeta|V) \geq p(y|V)$. This can be further written as $ln \frac{p(y+\zeta|V)}{p(y|V)} \geq 0$.

We consider that all random variables possess finite entropy. The average positivity condition can be defined as follows:

$$E_{y,\zeta|V}\left[ln \left(\frac{p(y + \zeta|V)}{p(y|V)} \right)\right] \geq 0. \tag{10}$$

We apply this condition to the backpropagation condition as shown in Fig. 2. Assume external noise ζ is added to every node of the hidden layer. Applying the likelihood ratio test and noise expectation maximization method produce the following inequality:

$$(V^T\zeta)^T ln(H_{NN}^t) \geq (U^T\zeta)^T ln(1 - H_{NN}), \tag{11}$$

where V is the synaptic weight, ln is logarithmic with base 'e' and $H_{NN}(X)$ is defined in Eq. 8, ζ is external noise, U is random variable with pdf $p(\frac{\zeta}{X})$.

In this paper, we focus mainly on uniformly and Irwin distributed noise as SR noise. Kosko et al. [12] also discuss the careful noise injection in the output layer. The detail of the proposed method has been discussed in Algorithm 1.

4   Simulation Results with Illustrative Example

In this section, we show and compare our proposed results with the state-of-the-art methods. The weak signal and different noises are discussed below.

$$s[i] = A, \quad i = 0, 1, 2, \ldots, (M - 1), \tag{12}$$

where M is the size of the signal. We consider the following two non-Gaussian noises which are inherently present with the signal.

Algorithm 1 Proposed Noise-induced Training Algorithm

1: **procedure** DETECTION(X)
2:    Generate the two set of samples following the Eq. 1. Let is name as X.
3:    The weights are initialized as V ← zeros$(1, (p + 2)q + 1)$;
4:    **Noise-induced training of the network is carried out as follows**:
5:    **while** $i : 1 \rightarrow q(< p)$ **do**
6:        Consider activation function as defined in Eq. 7.
7:        Evaluate $H_{NN}(X)$ as Eq. 8.
8:        SR noise is generated and name it as ζ.
9:        **if** $(V^T \zeta)^T ln(H_{NN}^t) \geq (I^T \zeta)^T ln(1 - H_{NN})$ **then**
10:            Add noise $y \leftarrow y + \zeta$, where y is the output of the hidden node and I is the identity matrix.
11:        **else**
12:            Add no noise.
13:        **end if**
14:        Evaluate the error cost function defined in Eq. 9.
15:        Apply backpropagation in order to minimize the cost function and update weight matrix.
16:    **end while**
17: **end procedure**

---

1. **Gaussian Mixed (GM)** The pdf of GM noise can be given as follows:

$$f_{int}(x) = \frac{c}{\sigma\sqrt{2\pi}} \left[\alpha \exp\left(-\frac{c^2 x^2}{2\sigma^2}\right) + \left(\frac{1-\alpha}{\beta}\right) \exp\left(-\frac{-c^2 x^2}{2\beta^2 \sigma^2}\right) \right] \qquad (13)$$

where $c = \sqrt{\alpha + (1-\alpha)\beta^2}$, $0 < \alpha < 1$, $\beta > 1$. For the GM noise, we consider the parameters $(\alpha, \beta, \sigma) = (0.3, 5, 1)$.

2. **Generalized Gaussian (GG)** The pdf of GG noise can be written as follows:

$$f_{int}(x) = \frac{c_1(\beta)}{\sigma} \exp\left(-c_2(\beta)\left|\frac{x}{\sigma}\right|^{\frac{2}{1+\beta}}\right), \qquad (14)$$

where,

$$\begin{cases} c_1(\beta) & = \dfrac{\Gamma^{\frac{1}{2}}\left(\frac{3(1+\beta)}{2}\right)}{(1+\beta)\left(\frac{1+\beta}{2}\right)}. \\[4mm] c_2(\beta) & = \left[\dfrac{\Gamma\left(\frac{3(1+\beta)}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right)}\right]^{\frac{1}{1-\beta}}. \end{cases} \qquad (15)$$

For the GG noise, we consider $(\beta, \sigma) = (1, 1)$ and, as SR noise, two noises have been considered which have to be added in the backpropagation.

1. **Uniformly Distributed Noise**

$$f_{ext}(x) = \frac{1}{2t}, \quad -t \leq x \leq t, \qquad (16)$$

where we consider $t = \frac{1}{\sqrt{3}}$.

2. **Irwin Noise** [19]

$$f_{ext}(x) = \frac{1}{(m-1)!} \sum_{k=0}^{x} (-1)^k \left({}^m C_k \right) (x-k)^{m-1}, \quad x \in R, \qquad (17)$$

where we consider $m = 5$.

In this paper, we have simulated the proposed algorithm in a similar fashion as suggested in the paper [7]. Similarly, variation in P_D for different SNR has also been discussed. The SNR is defined as follows:

$$SNR = \frac{[s]_{rms}}{\sigma^2}, \qquad (18)$$

where $[s]_{rms}$ represents the RMS value of the signal present in Eq. 1.

4.1 Weak DC Signal with Inherent Non-Gaussian Noise

Here, we show the performance and comparative study of the proposed method with different well-known methods for weak DC signal.

ROC Analysis In Fig. 3, we show the comparative performance in terms of receiver operating characteristic curve (ROC) for DC signal. The two Fig. 3a, b are the ROC curve with two different inherent noises Gaussian mixed and generalized Gaussian with uniform noise as stochastic resonance. However, Fig. 3c, d are the ROC curve when the Irwin noise is considered as stochastic noise with Gaussian mixed and generalized Gaussian noise as inherent noise.

When the Gaussian mixture (GM) $(\alpha, \beta, \sigma) = (0.9, 5, 1)$ noise is considered as inherent noise with the DC signal, the performance of the proposed detector with SR noise as uniform distribution and Irwin distribution is investigated, as shown in Fig. 3a, c, respectively. In both the cases, our proposed method works far better than ED [18], MF [16] and comparable to neural network-based detector [7] and Mandal et al. [17].

But in the case of generalized Gaussian (GG) $(\beta, \sigma) = (1, 1)$ noise, our proposed method with both the SR noise works better than all other methods discussed here. It can be seen in Fig. 3b, d.

In Table 1, we compare probability of detection, P_D at $P_{FA} = 0.1$, $M = 10$, and $A = 0.5$. The results show that our noise enhanced neural network proposed method is almost better in non-Gaussian noise. For example, in the case of Gaussian mixed noise, our P_D, $= 0.81$ and $= 0.85$ with uniformly distributed and Irwin noise, respectively. However, the different state-of-the-art methods such as Kay [16], Urkowitz [18], Guo et al. [17], Gandhi et al. [7], Halay et al. [8], and Naderpour et al. [10] produce 0.57, 0.25, 0.80, 0.85, 1, and 1, respectively.
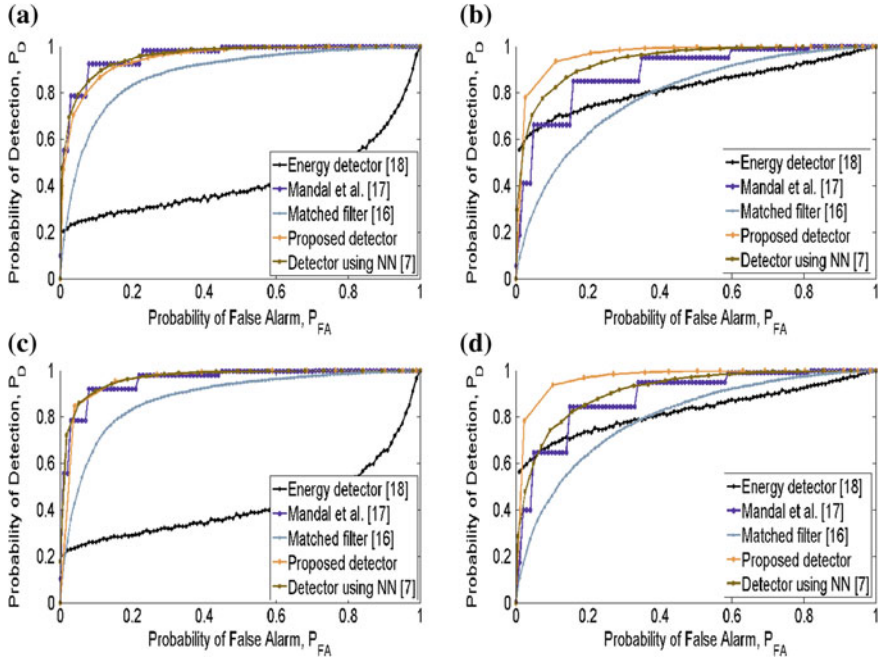
Fig. 3 ROC Curve for **DC a** (GM, Unif.) **b** (GG, Unif.) **c** (GM, Irwin) **d** (GG, Irwin). In the parenthesis, first noise is inherent which are discussed in Eqs. 13 and 14 whereas second is stochastic resonance noise given in Eqs. 16 and 17

Table 1 Comparision P_D for DC signal. The other parameters are as follows. M=10, s=0.5, $P_{FA} = 0.1$ and Monte Carlo run=10,000. Prop. (U) and Prop. (I) stand for the proposed method with SR noise as uniform and Irwin, respectively

Probability of detection, P_D								
Noise	[16]	[18]	[17]	[7]	**Prop. (U)**	**Prop. (I)**	[8]	[10]
GM.	0.57	0.25	0.80	0.85	**0.81**	**0.85**	1	1
GG.	0.40	0.65	0.65	0.79	**0.82**	**0.90**	1	1

P_D **Against SNR** We draw the curve for P_D at different SNR with constant $P_{FA} = 0.1$. For different SNR, we change the amplitude of DC (A) keeping the variance σ^2 of noise as constant. Figure 4 is the plot of P_D against SNR of noisy weak DC signal when the inherent noises are Gaussian mixed and generalized Gaussian with uniformly distributed noise and Irwin noise as SR noise.

Figure 4a, b are P_D against *SNR* for different noises Gaussian mixed and generalized Gaussian noise, respectively, with uniform noise as SR noise. Figure 4c, d are the plots for the probability of detection, P_D under different aforementioned noises with Irwin noise as SR noise.
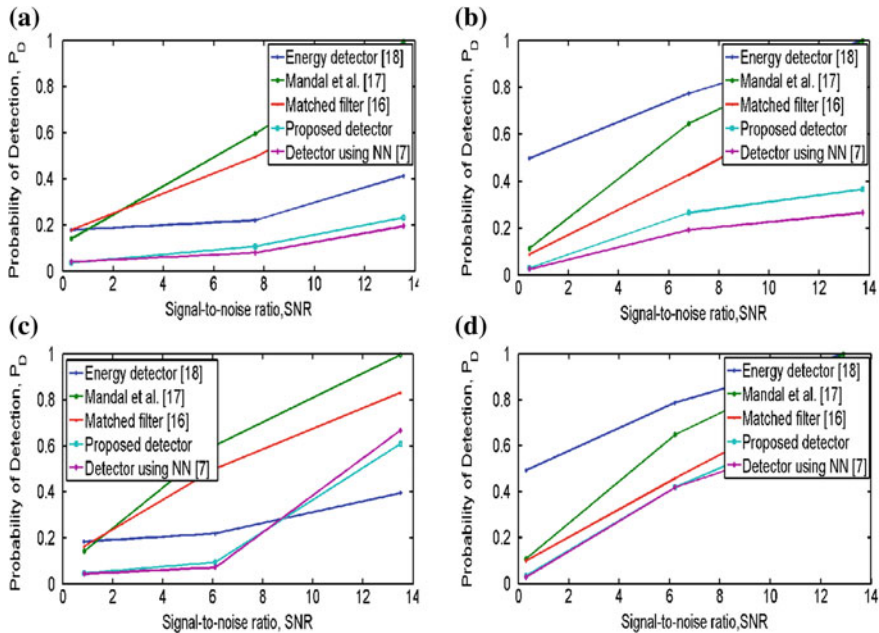
Fig. 4 P_D versus SNR for DC **a** (GM, Unif.) **b** (GG, Unif.) **c** (GM, Irwin) **d** (GG, Irwin). In the parenthesis, first noise is inherent which are discussed in Eqs. 13 and 14 whereas second is stochastic resonance noise given in Eqs. 16 and 17

Almost in every case, we observe that noise enhanced neural network based detector works better or comparable to the [7]. It can also be observed that as SNR increases, the value of P_D goes up. It justifies the theoretical approach for signal detection application. It also validates the benefits of the noise. As the SNR increases, the performance of the method based on neural network is comparatively not so better than other methods discussed here.

5   Conclusion

In this paper, the concept of stochastic resonance has been incorporated in neural network-based detector. This is termed as noise enhanced neural network-based detector. The positivity condition has also been discussed. Here, we consider two stochastic resonance noise, i.e., uniformly distributed noise and Irwin noise. The algorithm shows that there exist the weights between the two nodes when the noise is added during the backpropagation algorithm. The noise enhanced proposed detector yields superior performance in most of the cases.

References

1. Lehmann, E.L.: Testing Statistical Hypothesis. Wiley, New York (1986)
2. Poor, H.V.: An Introduction to Signal Detection and Estimation. Springer, Berlin (1988)
3. Kassam, S.A.: Signal Detection in Non-Gaussian Noise. Springer, Berlin (1988)
4. Watterson, J.W.: An optimum multilayer perceptron neural receiver for signal detection. IEEE Trans. Neural Netw. **1**, 298–300 (1990)
5. Lippmann, R.P., Beckman, P.: Adaptive neural net preprocessing for signal detection in non-Gaussian noise. Adv. Neural Inform. Process. Syst. **1**, 124–132 (1989)
6. Michalopoulou, Z., Nolta, L., Alexandrou, D.: Performance evaluation of multilayer perceptrons in signal detection and classification. IEEE Trans. Neural Netw. **6**, 381–386 (1995)
7. Gandhi, P.P., Ramamurti, V.: Neural networks for signal detection in non-Gaussian noise. IEEE Trans. Signal Process. **45**, 2846–2851 (1997)
8. Halay, N., Todros, K., Hero, A.O.: Binary hypothesis testing via measure transformed quasi-likelihood ratio test. IEEE Trans. Signal Process. **65**, 6381–6396 (2017)
9. Bell, M., Kochman, Y.: On composite binary hypothesis testing with training data. In: Fifty-Fifth Annual Allerton Conference, pp. 1026–1033 (2017)
10. Naderpour, M., Ghobadzadeh, A., Tadaion, A., Gazor, S.: Generalized wald test for binary composite hypothesis test. IEEE Signal Process. Lett. **22**, 2239–2243 (2015)
11. Ma, J., Xie, J., Gan, L.: Compressive detection of unknown-parameters signals without signal reconstruction. Signal Process. **142**, 114–118 (2018)
12. Audhkhasi, K., Osoba, O., Kosko, B.: Noise-enhanced convolutional neural networks. Neural Netw. **78**, 15–23 (2016)
13. Chen, H., Varshney, P.K., Kay, S.M.: Theory of stochastic resonance effects in signal detection: Part I-fixed detectors. IEEE Trans. Signal Process. **55**, 3172–3184 (2007)
14. Franzke, B., Kosko, B.: Using noise to speed up Morkov chain Monte Carlo estimation. Proced. Comput. Sci. **53**, 113–120 (2011)
15. Audhkhasi, K., Osoba, O., Kosko, B.: Noise benefits in backpropagation and deep bidirectional pre-training. In: IJCNN, pp. 1–8 (2013)
16. Kay, S.M.: Fundamentals of Statistical Signal Processing, Detection Theory, Vol II. Prentice Hall PTR, USA (2008)
17. Guo, G., Mandal, M., Jing, Y.: A robust detector of known signal in non-Gaussian noise using threshold system. J. Signal Process. **92**, 2676–2688 (2012)
18. Urkowitz, H.: Energy detection of unknown deterministic signals. Proc. IEEE **55**, 523–531 (1967)
19. Weissman, I.: Sum of squares of uniform random variables. Stat. Probab. Lett. **129**, 147–154 (2017)
20. Ikemoto, S., Libera, F.D., Hosoda, K.: Noise-modulated neural networks as an application of stochastic resonance. Neurocomputing **277**, 29–37 (2018)

Optimum APD Gain Evaluation of FSO System for Inter-building Laser Communication Application

Pritam Keshari Sahoo, Ajay Kumar Yadav, Y. K. Prajapati and Rajeev Tripathi

Abstract This paper evaluates the optimum gain performance of avalanche photodetector (APD) for various modulation techniques under dense fog condition of visibility within 50 m in Indian subcontinent. The comparative BER performance study has been outlined for the above conditions and weak atmospheric turbulence (Log-normal). Joint effects of shot, thermal, and background noise together with pointing error are taken into consideration for evaluation of error performance. In particular, it is observed that the performance of free-space optical (FSO) system at optimum APD gain is exceptionally well and good up to standard limit. Furthermore, the capacity evaluation in fog conditions is outlined for On-Off keying (OOK), binary phase-shift keying (BPSK), differential phase-shift keying (DPSK), and binary pulse position modulation (BPPM). From the diversity of performance study, this article finally suggests the suitable modulation technique for dense fog conditions and proposes the optimum gain of APD.

Keywords FSO · APD · Atmospheric turbulence · Pointing error · Outage probability

1 Introduction

Free-space optics (FSO) has been the fastest growing communication technology since last decade because of its unregulated spectrum, less cost of deployment, high-speed data transmission, and many more. FSO system is affected by atmospheric turbulences (AT), like fog, rain, snow fall, haze, etc., limiting the transmission reach [1]. The adverse effect of receiver noise, atmospheric loss, turbulence and pointing error (PE) on performance of FSO channel has been studied in literature [2]. Nevertheless, the effect of PE and aperture averaging (AA) is studied in [3], this

P. K. Sahoo (✉) · A. K. Yadav · Y. K. Prajapati · R. Tripathi
Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India
e-mail: sahoo.pritamkeshari@gmail.com

work presents the bit error rate (BER) performance comparison of various modulation techniques for inter-building laser communication in Indian subcontinent while visibility range is under 50 m due to dense fog.

Reference [4] examines the performance of BPPM with turbulence and pointing error. Similarly, the error performance study of OOK, DPSK, and BPSK with pointing error is reported in earlier literature [5–7]. However, the effect of AA in form of background noise due to large field of view (FOV) lens is analyzed. Recently in [8], authors have proposed a hybrid technique for a higher coverage distance of about 3 km. But, the study of pointing error impact on any communication system under log-normal turbulence channel and aperture averaging is yet left undisclosed. In this work, we have considered the turbulence channel model as log-normal and atmospheric attenuation as dense fog of visibility range under 50 m. In this paper, we analyze the BER performance comparison of all modulation techniques with PE.

2 System and Channel Model

We consider an intensity-modulated (IM)/direct detection (DD) FSO link employing various modulation techniques like OOK, BPSK, DPSK, and BPPM. The atmospheric scattering and absorption loss due to fog can be modeled by Beers–Lambert law,

$$h_l = e^{-\sigma L} \tag{1}$$

where σ is the wavelength and weather-dependent attenuation coefficient. L is the propagation distance. Through Kim's model [9]

$$\sigma = \frac{3.91}{V}\left(\frac{\lambda}{550}\right)^{-g} \tag{2}$$

where v is the visibility (in Km), λ is the laser wavelength (in nm), and g is the parameter related to the particle size distribution and v. Additive white Gaussian noise (AWGN) at the receiver we considered includes shot noise, thermal noise, and background noise as in [1].

We assumed the statistical misalignment-induced fading model proposed in [3] as

$$f_{h_p}(h_p) = \frac{\gamma^2}{A_0^{\gamma^2}}h_p^{\gamma^2-1}, \quad 0 \le h_p \le A_0 \tag{3}$$

where A_0 is the fraction of the collected power and γ is misalignment parameter, i.e., the ratio between equivalent beam radius at the receiver and the PE displacement

standard deviation at the receiver. AA is one of the simplest forms of spatial diversity, where the receiver aperture size is larger than fading correlation length. AA is considered to combat the scintillation effect. And the gain of AA is

$$A = \frac{\sigma_I^2(D)}{\sigma_I^2(0)} \tag{4}$$

For plane wave propagation [1],

$$\sigma_I^2(D) = exp\left[\frac{0.49\sigma_I^2}{\left(1 + 0.65d^2 + 1.11\sigma_I^{12/5}\right)^{7/6}} + \frac{0.51\sigma_I^2\left(1 + 0.69\sigma_I^{12/5}\right)^{-5/6}}{\left(1 + 0.9d^2 + 0.62d^2\sigma_I^{12/5}\right)^{7/6}}\right] - 1 \tag{5}$$

where $\sigma_I^2(D)$ and $\sigma_I^2(0)$ are the scintillation index of a receiver lens of diameter D and point receiver, respectively. σ_I^2 is Rytov variance of plane wave propagation [1] and parameter d is defined as

$$d = \sqrt{\frac{kD^2}{4L}} \tag{6}$$

Considering log-normal turbulence channel model with AA, the probability density function (pdf) of received irradiance is given by [3]

$$f_{h_s}(h_s) = \frac{1}{h_s\sqrt{2\pi\sigma_I^2(D)}}exp\left[-\frac{\left[ln(h_s) + \frac{1}{2}\sigma_I^2(D)\right]^2}{2\sigma_I^2(D)}\right] \tag{7}$$

Taking into account the atmospheric loss, pointing error, and atmospheric turbulence, the combined distribution $h = h_l h_p h_s$ is given as [7]

$$f_h(h) = \frac{\gamma^2}{2(A_0 h_l)^{\gamma^2}}h^{\gamma^2-1}e^{(0.5\sigma_I^2\gamma^2(1+\gamma^2))}erfc\left(\frac{ln\left(\frac{h}{A_0 h_l}\right) + \sigma_I^2\left(0.5 + \gamma^2\right)}{\sqrt{2\sigma_I^2}}\right) \tag{8}$$

3 Performance Analysis

a. BER Analysis

Next, we evaluate the error performance of FSO system in terms of received irradiance and APD gain. The average BER of all considered modulations under the effect of

combined channel model of (8) can be obtained as

$$P_e = \int\limits_0^\infty P_{ec} f_h(h) dh \tag{9}$$

where P_{ec} is the conditional BER in presence of AWGN channel. We have computed the above Eq. (9) through numerical integration using MATLAB and parameter values given in Table 1 under Appendix.

b. **Average Channel Capacity Analysis**

Average channel capacity signifies the information carrying capability (in bits/sec) of the channel as a function of instantaneous received signal-to-noise ratio (SNR). Assuming known channel pdf at the receiver end, the link capacity can be found through Shannon capacity [10].

$$C = \int\limits_0^\infty B \, log_2 \big(1 + SNR_{avg} \big) p(I) dI \tag{10}$$

where, B indicates the transmission bandwidth, SNR_{avg} is the average signal to noise ratio (SNR) and p(I) is the pdf of log-normal turbulence distribution. Using the characteristics of $log_2(x) = \ln(x)/\ln 2$, the link capacity under weak turbulence is shown in Fig. 4.

4   Result and Discussion

We evaluate the error performance of FSO system under pointing error effect and using AA. On a wavelength of 1550 nm and at a coverage length of 200 m, the error performances of various modulations are shown below at a range of APD gain. Figure 1 shows the error performance of binary PPM is better than OOK, DPSK, and BPSK for the same range of APD gain. However, the error performance study against APD gain shows, for a transmitted irradiance of 20 dBm, the optimum gain of APD is 16, 14, 13, and 11, respectively for binary PPM, BPSK, DPSK, and OOK.

Figure 2 shows the error performance comparative study for the same range of APD gain but at transmitted irradiance of 30 dBm. This result shows at an increased transmitted irradiance, the optimum gain of APD is approximately 16, 14, 13, 10, respectively for binary PPM, BPSK, DPSK, and OOK. In Fig. 3, however, the optimum gain falls to five for all modulation techniques. That means with increase in transmitted irradiance, the required gain of APD is minimum. Hence, this comparative study concludes that different modulation techniques give their optimum error performance at different APD gain.

Fig. 1 BER as a function of APD gain at a transmitter irradiance of 20 dBm
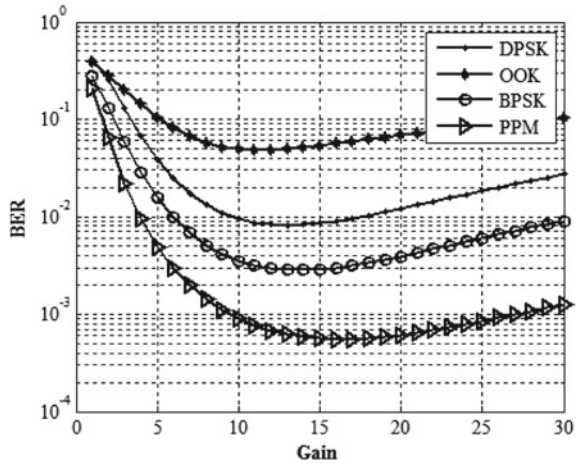
Fig. 2 BER as a function of APD gain at a transmitter irradiance of 30 dBm
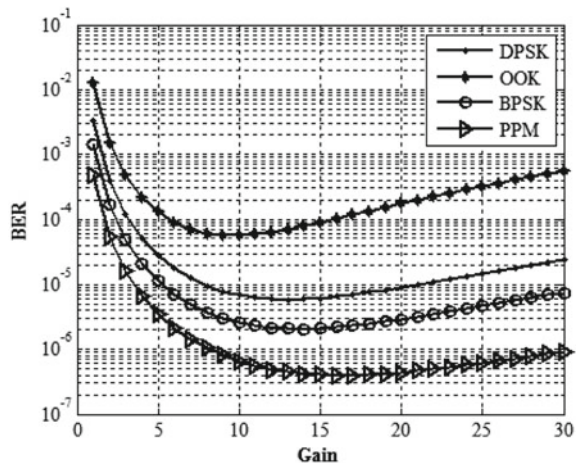
Authors in Fig. 4 have illustrated the transmission capacity for log-normal turbulence channel under pointing error effect. This shows with increase in transmitted irradiance the transmission capacity of the channel can be increased linearly. In comparison to DPSK, BPSK, and OOK modulation techniques, binary PPM gives highest capacity under pointing error effect.

5 Conclusion

The error performance and capacity study under pointing error effect for log-normal turbulence channel has been outlined in this paper. The research outcomes reveal

Fig. 3 BER as a function of APD gain at a transmitter irradiance of 40 dBm
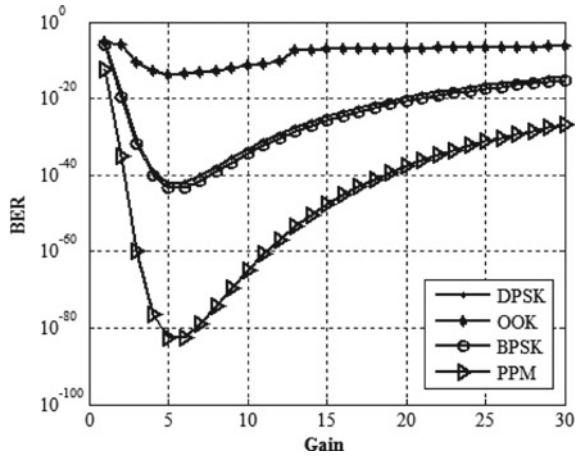
Fig. 4 Capacity against average received irradiance at an APD gain of 10
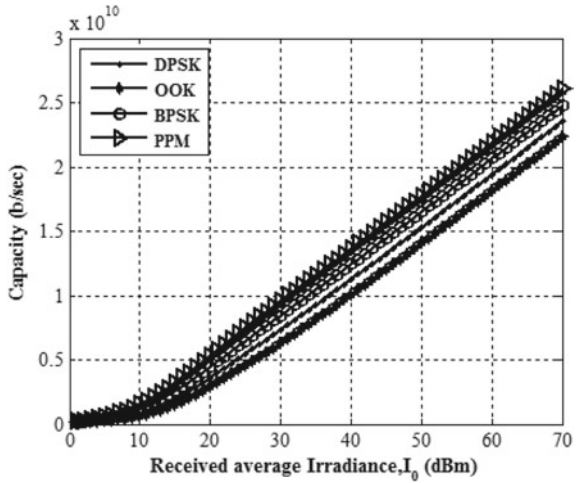
that through binary PPM, we can achieve superior error performance at 40 dBm optical irradiance for an APD gain of five, even at dense fog of visibility 50 m. This study has been carried out at a transmission distance of 200 m exclusively for inter-building LASER communication. Furthermore, a capacity of approximately 15 Gbps can be achieved through binary PPM under the same atmospheric turbulence of transmission irradiance 40 dBm.

Appendix

See Table 1.

Table 1 Parameter table

Parameters	Value
Bit rate (R_b)	5×10^9 bps
Wavelength (λ)	1550×10^{-9} m
Aperture diameter (D)	0.15 m
Propagation distance (L)	200 m
Temperature (T)	$300\,°$K
Load resistance (R_L)	1000 Ω
APD Gain (G)	10
Responsivity (R)	1
Amplifier noise figure (F_n)	2
Ionization factor (k_A)	0.7
BW of OBPF (Δλ)	0.8×10^{-9} m
Focal length (f)	50×10^{-3} m
Visibility (v)	50 m
Misalignment parameter (γ)	1.7725
Fraction of collected power (A_0)	0.8532

References

1. Ghassemlooy, Z., Popoola, W., Rajbhandari, S.: Optical Wireless Communications: System and Channel Modeling with MATLAB. Taylor & Francis (2012)
2. Sandalidis, H.G., Tsiftsis, T.A., Karagiannidis, G.K., Uysal, M.: BER performance of FSO links over strong atmospheric turbulence channel with pointing errors. IEEE Commun. Lett. **12**, 44–46 (2008). https://doi.org/10.1109/lcomm.2008.071408
3. Lee, I., Ghassemlooy, Z., Ng, W., Khalighi, M., Liaw, S.: Effects of aperture averaging and beam width on a partially coherent Gaussian beam over free-space optical links with turbulence and pointing errors. Appl. Opt. **55**, 1–9 (2016). https://doi.org/10.1364/ao.55.000001
4. Gappmair, W., Hranilovic, S., Leitgeb, E.: Performance of PPM on terrestrial FSO links with turbulence and pointing errors. IEEE Commun. Lett. **14**(5), 468–470 (2010). https://doi.org/10.1109/lcomm.2010.05.100202
5. Islam, A.K.M.N., Majumder, S.P.: Performance analysis of a free space optical link in the presence of pointing errors with space diversity. In: 2014 International Conference on Electrical Engineering and Information & Communication Technology, Dhaka, pp. 1–4 (2014). https://doi.org/10.1109/iceeict.2014.6919176
6. Ismail, T., Leitgeb, E.: Performance analysis of SIM-DPSK FSO system over lognormal fading with pointing errors. In: 2016 18th International Conference on Transparent Optical Networks (ICTON), Trento, pp. 1–4 (2016). https://doi.org/10.1109/icton.2016.7550350
7. Al-Nahhal, M., Ismail, T., Selmy, H., Elmesalawy, M.M.: BPSK based SIM-FSO communication system with SIMO over log-normal atmospheric turbulence with pointing errors. In: 2017 19th International Conference on Transparent Optical Networks (ICTON), Girona, pp. 1–4 (2017). https://doi.org/10.1109/icton.2017.8024953
8. Sahoo, P.K., Prajapati, Y.K., Tripathi, R.: PPM and GMSK based hybrid modulation technique for OWC cellular backhaul channel. IET Commun. (2018) (in press). https://doi.org/10.1049/iet-com.2018.5365

9. Kim, I.I., McArthur, B., Korevaar, E.: Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications. Proc. SPIE **4214**, 26–37 (2001). https://doi.org/10.1117/12.417512
10. Gopal, P., Jain, V.K., Kar, S.: Performance improvement of FSO satellite downlink using aperture averaging and receiver spatial diversity. IET Optoelectron. **10**(4), 119–127 (2016). https://doi.org/10.1049/iet-opt.2015.0102

EM Analysis of RF Interaction Structures for Gyrotron Devices

Rajiv Kumar Singh

Abstract In this work, an attempt is made to study the eigenvalue and dispersion characteristics of RF interaction structures to be used in gyrotron devices. Typically, a tapered structure, a vane-loaded structure, and a disc-loaded structure have been electromagnetically analyzed. The dependence of structure parameters has been observed using MATLAB simulations. For a tapered structure, tapering the RF structure results in flattening of the dispersion curve. For a vane-loaded structure, although, there is no considerable flattening in the dispersion curve, the cutoff frequency is highly dispersed, providing mode separation. In the case of disc-loaded structure, the structure parameter, viz., disc repetition has the highest impact on broadbanding the dispersion curve.

Keywords Gyrotron · Tapered structure · Disc-loaded structure · Vane-loaded structure

1 Introduction

High-power high-frequency gyrotron devices are undergoing a lot of advances due to their demand in the field of high-information density communications, nuclear reactors, medical electronics, industrial electronics, electronic warfare, radio astronomy, and many more areas. Due to their widespread applications, the gyrotron devices need to fulfill requirements like high power and high bandwidth. In order to meet the requirements, several modifications have been proposed in the recent past. RF structures used in the gyrotron devices play a very important role in the attainment of high power along with high-bandwidth requirements of the industry [1–5].

Several different types of RF interaction structures have been studied, analyzed, and experimentally tested in various parts of the globe. One such structure is a tapered structure that may be of either up-taper type or down-taper type. In such a tapered

R. K. Singh (✉)
Department of Electronics Engineering, Institute of Engineering & Technology, Lucknow, India
e-mail: rajivinbhu@gmail.com

Dr. APJ Abdul Kalam Technical University, Lucknow 226 021, India

structure, the magnetic field is also tapered so as to maintain synchronism between the RF wave and DC-applied magnetic field. It has been reported that a tapered structure, although provides large bandwidth, but at the expense of device gain [6, 7].

A central dielectric rod-loaded gyro-TWT amplifier has been proposed as a new interaction structure [8]. Slow time scale formalism based large signal analysis of such structure has been tried out with electron beam parameters: beam voltage 100 kV, beam current 5 A, beam pitch factor 0.59, and velocity spread 2%. Simulation result yielded 80 kW power from 9 to 11 GHz, 20% bandwidth, 30 dB saturated gain, and 15% efficiency.

The next conceptual step in the development of large-bandwidth gyro-TWT is based on the use of an all-metal structure in which a circular waveguide with helically grooved walls has been used. The bandwidth capability of such a gyro-TWT with a helically corrugated waveguide was experimentally verified by Bratman et al. [9]. A helix placed near the waveguide wall has also shown a capability to widen the coalescence between the beam mode line and waveguide mode curve [10–14].

The gyro-TWTs using the cylindrical waveguide with several types of corrugation have been adopted to improve its performance. A circular cylindrical waveguide loaded with an axially periodic annular disc, used in conventional TWTs as well as in linear accelerators, has been explored for its wideband potential in fast wave regime [10, 15–17]. In disc loading, the dispersion characteristics can be easily controlled by properly selecting the disc-loaded structure parameter. The small-signal analysis of a disc-loaded cylindrical waveguide given by Choe and Uhm considered only the fundamental component of field amplitude and ignored the effect of higher order harmonics [10]. Rekiouak et al. considered the higher order harmonics and gave the dispersion relation in the form of a determinant [18].

The vane-loaded cylindrical waveguide has been investigated and it was found that though it does not result in shaping the waveguide mode dispersion curve required for the wideband operation of the gyro-TWT; however, the cutoff frequency can be controlled by the optimization of the structure parameters [19]. Hence, the azimuthal periodic loading (vane loading) certainly helps in mode rarefaction, important for high beam-harmonic gyro-devices that can operate with lower beam energies and lesser magnetic fields. Coaxial waveguide resonator loaded with azimuthally separated vanes in the outer radius of the waveguide also facilitates mode selectivity [20–23]. Such type of structure enhances the RF fields in the slots, i.e., a strong coupling of higher harmonic fringing fields which ultimately enhances the growth rate. But, there is very little information available in the literature about the EM analysis of RF interaction structure at a single place. So, there is a lot of scope for the EM analysis of RF interaction structures to be used as an interaction structure in high-power microwave (HPM) gyrotron devices.

In the present paper, tapered structure, vane-loaded structure, and disc-loaded structure have been chosen for the electromagnetic analysis (Sect. 2). Field quantities in different regions of the structures along with appropriate boundary conditions resulted in the dispersion relation (Sect. 2). Result and detailed discussion is covered in Sect. 3.

2 EM Analysis of RF Structures

2.1 Tapered Structure

For a gyro-TWT operating in TE mode with a monoenergetic tenuous gyrating electron beam, the dispersion relation is expressed as [2]

$$\left(k_o^2 - \beta^2 - k_c^2\right)\left(\omega - \beta v_t - \frac{s\omega_c}{\gamma}\right)^2 = -\frac{\mu_o|c|^2 N_o(v_t/c)^2\left(\omega^2 - \beta^2 c^2\right)J_{x-m}^2\{k_c r_H\}J_S'^2\{k_c r_L\}}{\gamma m_{eo}\pi r_W^2[1 - \left(\frac{m}{k_c r_W)^2}\right]J_m'^2\{k_c r_W\}}$$

(1)

where ω is the wave frequency, β is the axial phase propagation constant, v_z and v_t are the axial and the transverse velocities of the electron beam, s is the beam cyclotron-harmonic mode number, m is the angular cyclotron-harmonic frequency of an electron of charge e corresponding to its rest mass m_{eo}, $k_o(= \omega/c)$ is the free-space propagation constant, $k_c(= \omega_{cut}/c = (k_o^2 - \beta_{cold}^2)^{1/2})$ is the cutoff wave number of the cold interaction structure, β_{cold} being the axial phase propagation constant of the guide, N_o is the number of electrons per unit axial length, r_W is the waveguide radius, μ_o is the permeability, and c the speed of light in free space. γ and γ_z are the relativistic mass factors given by

$$\gamma = (1 - \left(v_t^2 + v_z^2\right)/c^2)^{-1/2}$$

(2)

and

$$\gamma_z = \left(1 - v_z^2\right)/c^2)^{-1/2}$$

(3)

J with an integral subscript represents the ordinary Bessel function of the first kind, the subscript representing its order. The prime (′) with the Bessel function represents its derivative with respect to the argument.

Following Pierce's method for the conventional helix TWT, forward-wave solutions of the dispersion relation (1) for β around β_{cold} can be given as:

$$\beta = \beta_{cold} + j\beta_{cold}C\delta$$

(4)

such that $C\delta \ll 1$, where C and δ are each arbitrarily defined unitless quantities.

Also, Pierce's gain parameter C may be defined as

$$C = \left(\frac{K I_o}{4V_o}\right)^{1/3}$$

(5)

in which I_0 and V_0 being the beam current and voltage, respectively, and K, the interaction impedance of the structure. Substituting (4) into (1) in the following

form:

$$\delta(\delta + jb)^2 = j \tag{6}$$

where b is the velocity synchronization parameter defined as

$$b = \frac{\beta_e - \beta_{cold}}{\beta_{cold} C} \tag{7}$$

with β_e defined as a beam-propagation constant

$$\beta_e = \frac{\omega - s\left(\frac{\omega_c}{\gamma}\right)}{v_z} \tag{8}$$

The grazing-point magnetic flux density, corresponding to the grazing-point intersection is related to the cutoff frequency as

$$B_g = \frac{m_{eo}\gamma}{|e|s\gamma_s}\omega_{cut} \tag{9}$$

But, cyclotron frequency ω_c is related to the applied DC axial magnetic field Bo as

$$\omega_c = \frac{|e|B_o}{m_{eo}} \tag{10}$$

Combining (9) and (10) to express the ratio B_o/B_g as

$$\frac{B_o}{B_g} = \frac{s\gamma_z\omega_c}{\gamma\omega_{cut}} \tag{11}$$

2.2 Taper Design

In the problem, the radius of the waveguide r_W is tapered in the axial (z) direction flared-up in the forward direction of wave propagation (Fig. 1)—such that the start radius of the waveguide, vis-à-vis the frequency of operation, corresponds to the excitation of the input end of the device, and continues to propagate down the waveguide in compliance with the following characteristic equation:

$$\frac{\omega_{cut}\{z\}}{c}r_w\{z\} = constant \tag{12}$$

Fig. 1 Diagram showing
structure of tapered
waveguide with taper angle
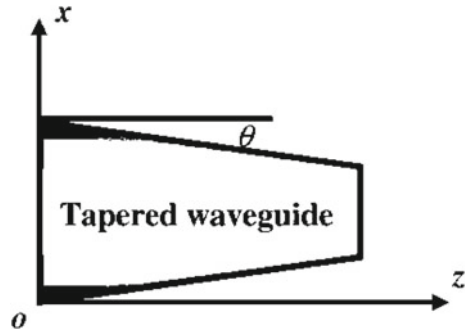
where $\omega_{\text{cut}\{z\}}$ is the cutoff angular frequency of the waveguide. It is obvious that, since $r_{W\{z\}}$ is tapered, one has to simultaneously taper the background magnetic flux density $B\{z\}$ and commensurate with the taper of the waveguide cross section in order to maintain the condition of cyclotron resonance throughout the length of the waveguide. Let us choose to taper $B_o\{z\}$ along the taper

$$\frac{B_o}{B_g} = constant \tag{13}$$

The left side (13) may be further expressed as

$$\frac{B_o}{B_g} = \frac{s\gamma_z\omega_c}{\gamma\omega_{cut}} \tag{14}$$

With the help of knowledge of equivalence of beam relativistic kinetic energy and the potential energy, the relativistic mass factor γ may be expressed in terms of the beam voltage V_o, as

$$\gamma = 1 + \frac{|e|V_o}{m_{eo}c^2} = constant \tag{15}$$

Therefore combining (10) and (12)–(15), we get the following for the taper:

$$\gamma_o\{z\}B_o\{z\}r_W\{z\} = constant \tag{16}$$

It is also implied that in view of the tapering of the background magnetic field, there will be an axial variation of the device parameters: Larmor radius (r_L), hollow-beam radius (r_H), axial electron velocity (v_z), transverse electron velocity (v_t), and beam velocity pitch factor α_o (=v_t/v_z).

Taking $r_L\{z\}$, $r_H\{z\}$, and $v_t\{z\}$ dependent on the axial distance—as per the taper of the magnetic flux density—as follows:

$$r_L\{z\}B_o^{1/2}\{z\} = constant \tag{17}$$

$$r_H\{z\}B_o^{1/2}\{z\} = constant \tag{18}$$

$$v\{z\}B_o^{1/2}\{z\} = constant \tag{19}$$

The relation (17) is valid under the adiabatic beam-flow approximation (15); the relation (18) follows from the consideration of the conservation of the magnetic flux; and the relation (19) from the conservation of electron magnetic moment. At this stage, it becomes convenient to define with the help of (13) and (17)–(19), the following quantities at any interaction length z relative to their corresponding values at the starting of the RF interaction structure (z = 0):

$$\left(\frac{B_o}{B_g}\right)\{z\} = \left(\frac{B_0}{B_g}\right)\{0\} \tag{20}$$

$$\left(\frac{r_L}{r_W}\right)\{z\} = \left(\frac{r_L}{r_W}\right)\{0\}\left(\frac{r_L\{z\}}{r_W\{0\}}\right)^{-1}\left(\frac{B_o\{z\}}{B_o\{0\}}\right)^{-1/2} \tag{21}$$

$$\left(\frac{r_H}{r_W}\right)\{z\} = \left(\frac{r_H}{r_W}\right)\{0\}\left(\frac{r_W\{z\}}{r_W\{0\}}\right)^{-1}\left(\frac{B_o\{z\}}{B_o\{0\}}\right)^{-1/2} \tag{22}$$

$$\left(\frac{v_t}{c}\right)\{z\} = \left(\frac{(\gamma^2 - 1)\alpha_o^2\{0\}}{(1 + \alpha_o^2\{0\})\gamma^2}\right)^{1/2}\left(\frac{B_o\{z\}}{B_o\{0\}}\right)^{-1/2} \tag{23}$$

$$\left(\frac{v_t}{c}\right)\{z\} = \left(\frac{\gamma^2 - 1}{\gamma^2}\right) - \left(\frac{(\gamma^2 - 1)\alpha_o^2\{0\}}{(1 + \alpha_o^2\{0\})\gamma^2}\right)^{1/2}\left(\frac{B_o\{z\}}{B_o\{0\}}\right)^{-1/2} \tag{24}$$

$$\alpha_o\{z\} = \frac{v_t\{z\}/c}{v_z\{z\}/c} \tag{25}$$

The right-hand side of (25) may be read with the help of (23) and (24). Thus in (25), we get the beam and magnetic flux density parameters at an axial distance from the start, in terms of their corresponding start quantities {at z = 0}, the taper parameter for the waveguide radius, and the taper parameter for the magnetic flux density.

With the help of (16) and (19), we may form a quadratic equation in the taper parameters, which can be solved to get

$$\frac{B_o\{z\}}{B_o\{0\}} = \frac{\gamma_z^2\{0\}\left(\frac{v_t\{0\}}{c}\right)^2 + \left(\gamma_z^2\{0\}\left(\frac{v_t\{0\}}{c}\right)^2\right)^2 + 4\left(\frac{r_W\{z\}}{r_w\{0\}}^2(\gamma_z\{0\}/\gamma)^2\right)^{1/2}}{2\left(\frac{r_W\{z\}}{r_w\{0\}}\right)^2} \tag{26}$$

$$\gamma_z\{0\} = (1 - (\frac{\gamma^2 - 1}{(1 + \alpha_o^2\{0\})\gamma^2})^2)^{-1/2} \tag{27}$$

$$\left(\frac{v_t}{c}\right)\{0\} = \frac{(\gamma^2 - 1)\alpha_o^2\{0\}}{(1 + \alpha_o^2\{0\})\gamma^2} \tag{28}$$

2.3 Vane-Loaded Structure

Figure 2 shows a vane-loaded circular cylindrical RF structure, in which wedge-shaped vanes are placed at regular intervals azimuthally.

EM Field Expression in Vane-Loaded Structure.
In order to write the EM field expressions in a vane-loaded RF structure, first of all, the complete structure is divided into two parts, the first part (region 1) being the vane-free cylindrical free space extending from axis of the structure to the tip of the vanes, i.e., $(0 \le r \le r_V; 0 \le \theta \le 2\pi)$, and the second part (region 2) being the vane-free wedge-shaped free space extending from the start of vane tips to the structure inner wall, i.e., $(r_V \le r \le r_W; \phi/2 \le \theta \le 2\pi/N - \phi/2)$. The structure parameters r_V, r_W, and N are inner radius of the vane edge, inner radius of the waveguide wall, and number of vanes. Due to the angular periodicity of the vanes, angular harmonics will be generated and thus the RF dependence is considered as $\exp(jm(2\pi/\phi)\theta - \exp(jmN\theta)$, where $\phi(= 2\pi/N)$ is the angular periodicity of the vanes, in which m is an integer representing the angular harmonic, and N is the number of vanes.

For a TE mode excited structure, the wave equation of a vane-loaded structure is given as

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + k_t^2 - \frac{(mN)^2}{r^2}\right)H_z = 0 \tag{29}$$

Fig. 2 Cross-sectional view of vane-loaded cylindrical waveguide
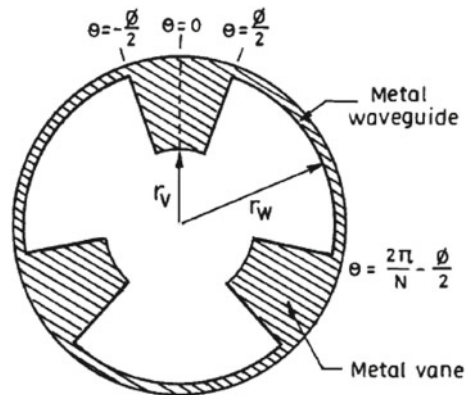
where, the transverse propagation constant is $k_t = \left(k_o^2 - \beta^2\right)^{1/2}$. For fast wave operation, the transverse propagation constant is a real quantity for $k_o > \beta$. For k_o corresponding to $\beta = 0 k_t$ is defined as the cutoff wave number $k_c (= \omega_{cut}/c)$ in which ω_{cut} is the cutoff frequency of the vane-loaded RF structure.

The solution of wave Eq. (29) of the vane-loaded RF interaction structure is given as

$$H_{z,p} = \sum_{m=-\infty}^{\infty} \left(A_{m,p} J_{mN}\{k_c r\} + B_{m,p} Y_{mN}\{k_c r\}\right) \tag{30}$$

where field constants are $A_{m.p}$ and $B_{m,p}$. $p = 1$ refers to the region 1 and $p = 2$ to region 2. J_{mN} and Y_{mN} are the $(mN)^{th}$ order ordinary Bessel functions of the first and second kinds, respectively.

Using Maxwell's equations and Eq. (29), the field quantities in region 1 can be found as

$$E_{z,1} = 0 \tag{31}$$

$$H_{z,1} = \sum_{m=-\infty}^{\infty} \left(A_{m,1} J_{mN}\{k_c r\} + B_{m,1} Y_{mN}\{k_c r\}\right) \tag{32}$$

$$E_{r,1} = \sum_{m=-\infty}^{\infty} -\left(mN\omega\mu_o/\left(k_c^2 r\right)\right)\left(A_{m,1} J_{mN}\{k_c r\} + B_{m,1} Y_{mN}\{k_c r\}\right) \tag{33}$$

$$H_{r,1} = \sum_{m=-\infty}^{\infty} -(j\beta/(k_c))\left(A_{m,1} J'_{mN}\{k_c r\} + B_{m,1} Y'_{mN}\{k_c r\}\right) \tag{34}$$

$$E_{\theta,1} = \sum_{m=-\infty}^{\infty} -(j\omega\mu_o/(k_c))\left(A_{m,1} J'_{mN}\{k_c r\} + B_{m,1} Y'_{mN}\{k_c r\}\right) \tag{35}$$

$$H_{\theta,1} = \sum_{m=-\infty}^{\infty} -\left(mN\beta/\left(k_c^2 r\right)\right)\left(A_{m,1} J_{mN}\{k_c r\} + B_{m,1} Y_{mN}\{k_c r\}\right) \tag{36}$$

The field quantities in region 2 can be obtained in a similar manner and written as

$$E_{z,2} = 0 \tag{37}$$

$$H_{z,2} = \sum_{m=-\infty}^{\infty} \left(A_{m,2} J_{mN}\{k_c r\} + B_{m,2} Y_{mN}\{k_c r\}\right) \tag{38}$$

$$E_{r,2} = \sum_{m=-\infty}^{\infty} -\left(mN\omega\mu_o/\left(k_c^2 r\right)\right)\left(A_{m,2} J_{mN}\{k_c r\} + B_{m,2} Y_{mN}\{k_c r\}\right) \tag{39}$$

$$H_{r,2} = \sum_{m=-\infty}^{\infty} -(j\beta/(k_c)) \left(A_{m,2} J'_{mN}\{k_c r\} + B_{m,2} Y'_{mN}\{k_c r\} \right) \tag{40}$$

$$E_{\theta,2} = \sum_{m=-\infty}^{\infty} -(j\omega\mu_o/(k_c)) \left(A_{m,2} J'_{mN}\{k_c r\} + B_{m,2} Y'_{mN}\{k_c r\} \right) \tag{41}$$

$$H_{\theta,2} = \sum_{m=-\infty}^{\infty} -(mN\beta/(k_c^2 r)) \left(A_{m,2} J_{mN}\{k_c r\} + B_{m,2} Y_{mN}\{k_c r\} \right) \tag{42}$$

where the differentiation of Bessel functions with respect to their argument is represented as primes.

Boundary Conditions in Vane-Loaded Structure.
Of the four field constants $A_{m,1}$, $B_{m,1}$, $A_{m,2}$, and $B_{m,2}$, the constant $B_{m,1}$ is considered as null so as to prevent the field quantities from increasing to infinity. The boundary conditions in a vane-loaded RF interaction structure are given as

$$E_{\theta,1} = E_{\theta,2}\big|_{r=r_V} \ (\phi/2 \leq \theta \leq 2\pi/N - \phi/2) \tag{43}$$

$$E_{\theta,2} = 0\big|_{r=r_W} \ (\phi/2 \leq \theta \leq 2\pi/N - \phi/2) \tag{44}$$

The other two boundary conditions of the vane-loaded RF interaction structure are expressed as

$$H_{\theta,1} = H_{\theta,2}\big|_{r=r_V} \ (\phi/2 \leq \theta \leq 2\pi/N - \phi/2) \tag{45}$$

$$H_{r,1} = 0\big|_{r=r_V} \ (-\phi/2 \leq \theta \leq \phi/2) \tag{46}$$

Dispersion Relation of Vane-Loaded Structure.
Considering the nonzero constants $A_{m,2}$ and $B_{m,2}$ and representing them in terms of $A_{m,1}$ and with the help of the boundary conditions (43)–(46), into which the field expressions (31)–(42) are substituted, we may easily show that

$$\alpha_q A_{q,1} + \sum_{\substack{m=-\infty \\ m \neq q}}^{\infty} \delta_{m,q} A_{m,1} = 0 \tag{47}$$

where

$$\alpha_q = J'_{qN}\{k_C r_V\}\phi + \left((1+\eta_q) J_{qN}\{k_C r_V\} - \eta_q \frac{J'_{qN}\{k_C r_V\}}{Y'_{qN}\{k_C r_V\}} Y_{qN}\{k_C r_V\} \right)(2\pi/N - \phi) \tag{48}$$

and

$$\delta_{m,q} = \frac{2\sin((m-q)(N\phi/2))}{(m-q)N}$$
$$\times \left(J'_{mN}\{k_c r_V\} - \left((1+\eta_m) J_{mN}\{k_c r_V\} + \eta_m \frac{J'_{mN}\{k_c r_V\}}{Y'_{mN}\{k_c r_V\}} Y_{mN}\{k_c r_V\} \right) \right)$$
(49)

For: (a) q = 0, m = 1, m = −1; (b) q = 1, m = 0, m = −1; and (c) q = −1, m = 0, m = 1, referring to the three lowest-order modes, the following set of three equations are written

$$\left.\begin{array}{l} \alpha_0 A_{0,1} + \delta_{1,0} A_{1,1} + \delta_{-1,0} A_{-1,1} = 0 \\ \delta_{0,1} A_{0,1} + \alpha_1 A_{1,1} + \delta_{-1,1} A_{-1,1} = 0 \\ \delta_{0,-1} A_{0,1} + \delta_{1,-1} A_{1,1} + \alpha_{-1} A_{-1,1} = 0 \end{array}\right\}$$
(50)

The nontrivial solution of Eq. (50) will exist if the determinant formed by the coefficients of the constants occurring in the Eq. (50) vanish, i.e.,

$$\begin{vmatrix} \alpha_0 & \delta_{1,0} & \delta_{-1,0} \\ \delta_{0,1} & \alpha_1 & \delta_{-1,1} \\ \delta_{0,-1} & \delta_{1,-1} & \alpha_{-1} \end{vmatrix} = 0, \text{ which simplifies to}$$

$$\alpha_0\alpha_1\alpha_{-1} - \alpha_0\delta_{-1,1}\delta_{1,-1} + \delta_{1,0}\delta_{0,-1}\delta_{-1,1} + \delta_{1,0}\delta_{0,1}\alpha_{-1} + \delta_{-1,0}\delta_{0,1}\delta_{1,-1} - \delta_{-1,0}\delta_{0,-1}\alpha_1 = 0$$
(51)

After evaluating individual terms of Eq. (51), it has been found that the term $\alpha_0\alpha_1\alpha_{-1}$ dominates over all other terms, thus retaining only this term, Eq. (51) it can be written as

$$\alpha_0\alpha_1\alpha_{-1} = 0$$
(52)

All the three factors of (52) when equated to zero results in three dispersion relations for each of the three possible waves. Thus, considering

$$\alpha_0 = 0$$
(53)

For the special case of $\phi = 0$ and $r_v = r_W$, and considering $\alpha_0 = 0$, Eq. gives $J'_0\{k_c r_W\} = 0$ which is the dispersion relation of a smooth waveguide excited in the TE_{0n} mode. Equation (51) gets modified, for mode orders more than three, as

$$\prod_{m=0,\pm1,\pm2,\ldots} \alpha_m = 0$$
(54)

2.4 Disc-Loaded Structure

A disc-loaded RF interaction structure is shown in Fig. 3.

Field Expressions and Boundary Conditions.
In order to write the EM field expressions inside the disc-loaded RF structure, the structure has been physically divided into two distinct parts: the first region identified as region I extends from $0 \leq r < r_D$ and $0 < z < L$, while the second part identified as region II extends from $r_D \leq r < r_W$ and $0 < z < L - T$, where r_D, r_W, L, and T are disc hole radius, waveguide-wall radius, axial periodicity of discs, and disc thickness.

For TE mode of operation and taking azimuthally symmetric mode ($\partial/\partial\theta = 0$), the solution of the wave equation in the structure region I of disc-loaded RF structure is written as

$$H_z^I = \sum_{n=-\infty}^{+\infty} H_{z,n}^I = \sum_{n=-\infty}^{+\infty} [A_n^I J_0\{\gamma_n^I r\} + B_n^I Y_0\{\gamma_n^I r\}] \exp j\,(\omega t - \beta_n^I z) \qquad (55)$$

where radial propagation constant

$$\gamma_n^I = (k^2 - \beta_n^{I2})^{1/2} \qquad (56)$$

in which β_n^I is axial phase propagation constant in the structure region I; and

$$k(= \omega(\mu_0 \varepsilon_0)^{1/2}) \qquad (57)$$

Fig. 3 Schematic of a circular waveguide loaded with annular metal discs showing the structure regions (I and II), structure surfaces (i, ii, iii, and iv), and disc parameters
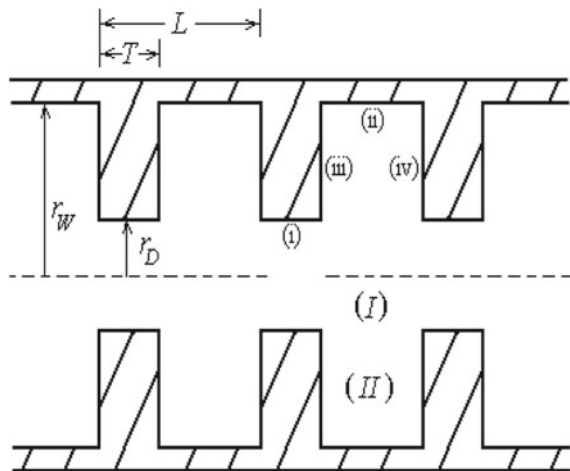
is the free-space propagation constant. Here, n represents the space-harmonic number $(n = 0, \pm 1, \pm 2, \pm 3 \ldots \pm \infty)$ resulting due to the axial periodicity of the structure. A_n^I and B_n^I are the field constants.

Further, it is noteworthy here that $Y_0\{\gamma_n^I r\} \rightarrow -\infty$ as $r \rightarrow 0$. Thus, the field constant B_n^I is taken as zero so as to prevent the field from blowing up to infinity. Thus,

$$H_z^I = \sum_{n=-\infty}^{+\infty} H_{z,n}^I = \sum_{n=-\infty}^{+\infty} A_n^I J_0\{\gamma_n^I r\} \exp j(\omega t - \beta_n^I z). \tag{58}$$

EM Field Quantities in the Disc-Free Region.
The Floquet's theorem for periodic structures says that, for a given mode of propagation and at a given steady-state frequency, the fields at one cross section of the structure differ from those one period away by a complex constant [1]. Thus, according to this theorem and taking the axial functional dependence of RF field quantities $f\{z\}$, considering space harmonics generated by the axial periodicity of the discs, one can write as

$$f\{z + L\} = f\{z\} \exp(-j\beta_0 L) \tag{59}$$

where $f\{z\} = \exp(-j\beta_n z)$, L is the axial periodicity of the structure, and β_n is the axial phase propagation constant of the nth space-harmonic mode supported by the structure, β_0 referring to the fundamental mode. $f\{z + L\}$ can be expressed with the help of $f\{z\}$, as

$$f\{z + L\} = \exp[-j\beta_n(z + L)]. \tag{60}$$

From (59) and (60), one may write

$$\exp[-j\beta_n(z + L)] = f\{z\} \exp(-j\beta_0 L). \tag{61}$$

Substituting $f\{z\}$ in (61), one obtains

$$\exp[-j\beta_n(z + L)] = \exp[-j(\beta_n z + \beta_0 L)]. \tag{62}$$

It then immediately follows from the phase part of (62) and will satisfy the relation $\beta_n(z + L) = \beta_n z + \beta_0 L \pm 2n\pi$, which algebraically simplifies to $\beta_n = \beta_0 + 2n\pi/L$, this may be written, interpreting n such that it can assume both positive and negative values, as

$$\beta_n = \beta_0 + \frac{2n\pi}{L}, \quad (n = 0, \pm 1, \pm 2, \pm 3 \ldots) \tag{63}$$

The relation (63), interpreted here from Floquet's theorem, has then to be used subsequently for stating the field components in region I (disc-free region) (Fig. 3).

Thus, the wave equation can be written as

$$H_z^I = \sum_{n=-\infty}^{+\infty} H_{z,n}^I = \sum_{n=-\infty}^{+\infty} A_n^I J_0\{\gamma_n^I r\} \exp j(\omega t - \beta_n^I z) \qquad (64)$$

The field quantities can be obtained with the help of Maxwell equations for azimuthally symmetric TE_{0n} modes as

$$E_\theta^I = \frac{j\omega\mu_0}{(\gamma_n^I)^2} \frac{\partial H_z^I}{\partial r} \qquad (65)$$

$$E_r^I = 0 \qquad (66)$$

$$H_\theta^I = 0 \qquad (67)$$

$$H_r^I = -\frac{j\beta_n^I}{(\gamma_n^I)^2} \frac{\partial H_z^I}{\partial r}. \qquad (68)$$

Substituting the expression for the axial component magnetic field intensity H_z^I from Eq. (64) into (65) and (68), then one can respectively obtain, for region I, the following expressions for the azimuthal component of electric field intensity and radial component of magnetic field intensity:

$$H_z^I = \sum_{n=-\infty}^{+\infty} H_{z,n}^I = \sum_{n=-\infty}^{+\infty} A_n^I J_0\{\gamma_n^I r\} \exp j(\omega t - \beta_n^I z) \qquad (69)$$

$$E_z^I = 0 \qquad (70)$$

$$E_\theta^I = \sum_{n=-\infty}^{+\infty} E_{\theta,n}^I = j\omega\mu_0 \sum_{n=-\infty}^{+\infty} \frac{1}{\gamma_n^I} A_n^I J_0'\{\gamma_n^I r\} \exp j(\omega t - \beta_n^I z) \qquad (71)$$

$$E_r^I = 0 \qquad (72)$$

$$H_r^I = \sum_{n=-\infty}^{+\infty} H_{r,n}^I = -j \sum_{n=-\infty}^{+\infty} \frac{\beta_n^I}{\gamma_n^I} A_n^I J_0'\{\gamma_n^I r\} \exp j(\omega t - \beta_n^I z) \qquad (73)$$

$$H_\theta^I = 0 \qquad (74)$$

The axial phase propagation constant β_n^I occurring in the field expressions (69), (71), and (73) may be written, in view of (63), as follows:

$$\beta_n^I = \beta_0^I + \frac{2\pi n}{L} \quad (n = 0, \pm 1, \pm 2, \pm 3 \ldots) \qquad (75)$$

EM Field Quantities in the Disc-Occupied Region.

The region *II* of the disc-loaded waveguide (Fig. 3) consists of a cell-type structure that supports stationary waves due to reflection at discs. The separation between two consecutive discs can be written in terms of an integral number of half guide wavelengths as

$$L - T = \frac{m\lambda_m^{II}}{2} \quad (m = 0, 1, 2, 3 \ldots) \tag{76}$$

where *m* is the modal number of the stationary wave and λ_m^{II} is guide wavelength of the stationary wave existing between successive discs. The relation between λ_m^{II} and β_m^{II} in region *II* is given as

$$\lambda_m^{II} = \frac{2\pi}{\beta_m^{II}}. \tag{77}$$

With the help of (76) and (77), β_m^{II} can be expressed as

$$\beta_m^{II} = \frac{m\pi}{L - T} \quad (m = 0, 1, 2, 3, \ldots.) \tag{78}$$

Due to reflections from the disc wall in region II, standing waves are formed. These standing waves are composed of forward-wave and the backward-wave component. So, these components are expressed with the help of (57) as

$$H_{z,m,f}^{II} = [A_{m,f}^{II} J_0\{\gamma_m^{II} r\} + B_{m,f}^{II} Y_0\{\gamma_m^{II} r\}] \exp j(\omega t - \beta_m^{II} z) \tag{79}$$

$$H_{z,m,b}^{II} = [A_{m,b}^{II} J_0\{\gamma_m^{II} r\} + B_{m,b}^{II} Y_0\{\gamma_m^{II} r\}] \exp j(\omega t + \beta_m^{II} z) \tag{80}$$

where $H_{z,m,f}^{II}$ and $H_{z,m,b}^{II}$ are the forward and backward components of magnetic fields, respectively. $\gamma_m^{II} (= (k^2 - \beta_m^{II2})^{1/2})$ is the radial propagation constant in region *II*. The boundary condition that the normal component of magnetic flux density and hence also that of the magnetic field intensity in region *II* vanishes, at the interface, interpreted at $z = 0$, between the metal disc and the free space, gives

$$(H_{z,m,f}^{II} + H_{z,m,b}^{II})\big|_{z=0} = 0, \tag{81}$$

the direction of the normal to the interface at $z = 0$ being essentially in the $+z$ direction. Putting (79) and (80), each taken at $z = 0$, in (81), one then obtains the following relation for the field constants referring to region *II*: $(A_{m,f}^{II} + A_{m,b}^{II}) J_0\{\gamma_m^{II} r\} + (B_{m,f}^{II} + B_{m,b}^{II}) Y_0\{\gamma_m^{II} r\} = 0$, which, in view of $J_0\{\gamma_m^{II} r\}$ and $Y_0\{\gamma_m^{II} r\}$, each taking on determinate values in region *II* $(r_D \leq r < r_W)$ (Fig. 3), immediately yields

$$A_{m,b}^{II} = -A_{m,f}^{II} \atop B_{m,b}^{II} = -B_{m,f}^{II} \Big\}. \tag{82}$$

Now, mth modal harmonic of the axial magnetic field can be expressed as, with the help of (79), (80), and (82)

$$\begin{aligned} H_{z,m}^{II} &= H_{z,m,f}^{II} + H_{z,m,b}^{II} \\ &= [A_{m,f}^{II} J_0\{\gamma_m^{II}r\} + B_{m,f}^{II} Y_0\{\gamma_m^{II}r\}] \exp(j\omega t)[\exp(-j\beta_m^{II}z) - \exp(j\beta_m^{II}z)], \end{aligned}$$

which, in view of the relation $\exp(j\beta_m^{II}z) - \exp(-j\beta_m^{II}z) = 2j\sin(\beta_m^{II}z)$, may be read as

$$H_{z,m}^{II} = -2j[A_{m,f}^{II} J_0\{\gamma_m^{II}r\} + B_{m,f}^{II} Y_0\{\gamma_m^{II}r\}] \exp(j\omega t)\sin(\beta_m^{II}z). \tag{83}$$

However, the field constant $B_{m,f}^{II}$ is not independent of $A_{m,f}^{II}$. The relation between these field constants can easily be found. For this purpose, remembering the azimuthal symmetry of the problem ($\partial/\partial\theta = 0$), one may first write mth stationary-wave modal harmonic azimuthal component of electric field intensity as

$$E_{\theta,m}^{II} = \frac{j\omega\mu_0}{\gamma_m^{II2}} \frac{\partial H_{z,m}^{II}}{\partial r}. \tag{84}$$

Also, recalling the concept of the null value of the electric field intensity at a tangent of waveguide wall $r = r_W$, i.e.,

$$E_{\theta,m}^{II} = 0, \quad (0 \le z < \infty; r = r_W). \tag{85}$$

One then obtains the following relation for $B_{m,f}^{II}$ in terms of $A_{m,f}^{II}$, by first substituting (83) in (84), and then putting the resulting expression, in (85):

$$B_{m,f}^{II} = -A_{m,f}^{II} \frac{J_0'\{\gamma_m^{II}r_W\}}{Y_0'\{\gamma_m^{II}r_W\}}. \tag{86}$$

Using (86), the Eq. (83) is rewritten as

$$H_{z,m}^{II} = \left(-2j \frac{A_{m,f}^{II}}{Y_0'\{\gamma_m^{II}r_W\}}\right) Z_0\{\gamma_m^{II}r\} \exp(j\omega t)\sin(\beta_m^{II}z), \tag{87}$$

where

$$Z_0\{\gamma_m^{II}r\} = J_0\{\gamma_m^{II}r\} Y_0'\{\gamma_m^{II}r_W\} - J_0'\{\gamma_m^{II}r_W\} Y_0\{\gamma_m^{II}r\}. \tag{88}$$

In order to use the simpler expression, the quantity in parenthesis in (87) is taken as

$$A_m^{II} = -2j \frac{A_{m,f}^{II}}{Y_0'\{\gamma_m^{II} r_W\}}.$$ (89)

Thus, putting (89) in (87), one may express (87) as

$$H_{z,m}^{II} = A_m^{II} Z_0\{\gamma_m^{II} r\} \exp(j\omega t) \sin(\beta_m^{II} z).$$ (90)

Therefore, the field expressions in region II may be rewritten with the help of (90) and (84) as

$$E_z^{II} = 0$$ (91)

$$H_z^{II} = \sum_{m=1}^{\infty} H_{z,m}^{II} = \sum_{m=1}^{\infty} A_m^{II} Z_0\{\gamma_m^{II} r\} \exp(j\omega t) \sin(\beta_m^{II} z)$$ (92)

$$E_\theta^{II} = \sum_{m=1}^{\infty} E_{\theta,m}^{II} = j\omega\mu_0 \sum_{m=1}^{\infty} \frac{1}{\gamma_m^{II}} A_m^{II} Z_0'\{\gamma_m^{II} r\} \exp(j\omega t) \sin(\beta_m^{II} z)$$ (93)

$$H_\theta^{II} = 0$$ (94)

$$E_r^{II} = 0$$ (95)

$$H_r^{II} = -\frac{j\beta_m^{II}}{\gamma_m^{II}} \sum_{m=1}^{\infty} A_m^{II} Z_o'\{\gamma_m^{II} r\} \exp(j\omega t) \sin(\beta_m^{II} z)$$ (96)

where Z_0' is the first derivative of Z_0.

Electromagnetic Boundary Conditions.
For characterizing the disc-loaded circular waveguide, one has to combine the relevant boundary conditions with the field expressions for the waveguide, excited in azimuthally symmetric TE modes ($E_z = 0$). The free-space interface of region I and region II, at $r = r_D$, establishes continuity of the tangential components of electric and magnetic field intensities. Similarly, the tangential component of electric field intensity at the metal inner circumferential edge of the discs, $r = r_D$, vanishes. Making use of these concepts, the boundary conditions at $r = r_D$ can be written as

$$E_\theta^I = \begin{cases} E_\theta^{II} & 0 < z < (L-T) \\ 0 & (L-T) \le z \le L \end{cases} \quad (r = r_D)$$ (97)

$$H_z^I = H_z^{II} \quad 0 < z < (L-T) \quad (r = r_D).$$ (98)

Similarly, the boundary condition at $r = r_W$ can be written with the help of knowledge of vanishing of the tangential component of electric field intensity at the metallic surface as

$$E_\theta^{II} = 0 \ (0 \le z < \infty; \ r = r_W). \tag{99}$$

Dispersion Relation.

First of all, the field expression (69) for H_z^I in region I and the field expression (92) for H_z^{II} in region II are substituted into the boundary condition (98) to obtain

$$\sum_{n=-\infty}^{\infty} A_n^I J_0\{\gamma_n^I r_D\} \exp(-j\beta_n^I z) = \sum_{m=1}^{\infty} A_m^{II} Z_0\{\gamma_m^{II} r_D\} \sin(\beta_m^{II} z), \tag{100}$$

where γ_n^I and γ_m^{II} are defined following (69) and (92), respectively. Z_0 may be interpreted, by taking $r = r_D$ in (88), as $Z_0\{\gamma_m^{II} r_D\} = J_0\{\gamma_m^{II} r_D\} Y_0'\{\gamma_m^{II} r_W\} - J_0'\{\gamma_m^{II} r_W\} Y_0\{\gamma_m^{II} r_D\}$.

Similarly, the field expression (71) for E_θ^I in region I and the field expression (93) for E_θ^{II} in region II are substituted into the boundary condition (97) to obtain

$$\sum_{n=-\infty}^{\infty} \frac{1}{\gamma_n^I} A_n^I J_0'\{\gamma_n^I r_D\} \exp(-j\beta_n^I z) = \sum_{m=1}^{\infty} \frac{1}{\gamma_m^{II}} A_m^{II} Z_0'\{\gamma_m^{II} r_D\} \sin(\beta_m^{II} z) \tag{101}$$

Multiplying (100) by $\sin(\beta_m^{II} z)$, integrating it from $z = 0$ to L, and using the relevant definite integrals, an expression for A_m^{II} in the form of a series relating A_n^I $(-\infty < n < \infty)$ may be obtained as

$$A_m^{II} = \sum_{n=-\infty}^{\infty} A_n^I P_{nm} \quad (m = 1, 2, 3 \ldots) \tag{102}$$

where

$$P_{nm} = \frac{2}{L} \frac{J_0\{\gamma_n^I r_D\}}{Z_0\{\gamma_m^{II} r_D\}} \frac{\beta_m^{II}[1 - (-1)^m \exp(-j\beta_0^I L)]}{[(\beta_m^{II})^2 - (\beta_n^I)^2]} \tag{103}$$

For getting A_n^I in terms of A_m^{II}, let us first multiply (101), in which β_n^I is taken as (75), by $\exp(j\beta_n^I z)$ and then integrate it from $z = 0$ to L, while using the relevant definite integrals to obtain

$$A_n^I = \sum_{m=1}^{\infty} A_m^{II} R_{nm}, \quad (n = 0, \pm 1, \pm 2 \ldots) \tag{104}$$

where

$$R_{nm} = \frac{1}{L} \frac{\gamma_n^I}{\gamma_m^{II}} \frac{Z_0'\{\gamma_m^{II} r_D\}}{J_0'\{\gamma_n^I r_D\}} \frac{\beta_m^{II}[(-1)^m \exp(j\beta_0^I L) - 1]}{[(\beta_n^I)^2 - (\beta_m^{II})^2]}. \tag{105}$$

Substitution of (104) into (102) yields

$$A_m^{II} = \sum_{n=-\infty}^{+\infty} \sum_{m=1}^{+\infty} A_m^{II} R_{nm} P_{nm}. \tag{106}$$

Putting in (106) A_1^{II} ($m = 1$) and taking rest all A_m^{II}'s equal to zero, it has been obtained as $1 - \sum_{n=-\infty}^{+\infty} R_{n1} P_{n1} = 0$. Now using (103) and (105), the dispersion relation of the structure is obtained as

$$\sum_{n=-\infty}^{+\infty} \gamma_n^I \frac{J_0\{\gamma_n^I r_D\}}{J_0'\{\gamma_n^I r_D\}} \frac{1}{[(\pi/L)^2 - (\beta_n^I)^2]^2}$$
$$= \frac{\gamma_1^{II} \exp(j\beta_0^I L)}{[1 + \exp(j\beta_0^I L)]^2} \frac{L^4}{2\pi^2} \frac{Z_0\{\gamma_1^{II} r_D\}}{Z_0'\{\gamma_1^{II} r_D\}}, (-\infty < n < \infty, m = 1) \tag{107}$$

3 Result and Discussion

3.1 Tapered Structure

Cross-sectional view of the designed tapered waveguide is shown in Fig. 4. In order to estimate the effect of taper on the response of the device, the interaction length has been divided into small sections of uniform length, say, of the order of a few microns, and approximated each section to be of uniform radius equal to its actual midpoint radius, so that dispersion relation could be applied individually to each of these directions. The calculations have been done for the entire length of the structure.

The linear taper for the waveguide radius, r_W, may be put in terms of the taper angle Ψ, as follows:

$$\frac{r_W\{z\}}{r_W\{0\}} = 1 + \frac{z \tan \Psi}{r_W\{\frac{l}{2}\} - (\frac{l}{2}) \tan \Psi} \quad (34) \tag{108}$$

where $\Psi\left(\tan^{-1}(r_W\{l\} - r_W\{0\})/l\right)$ is the taper angle that is defined by the initial start radius of the waveguide $r_W\{0\}$, the final radius of the waveguide $r_W\{l\}$, and l is the interaction length.
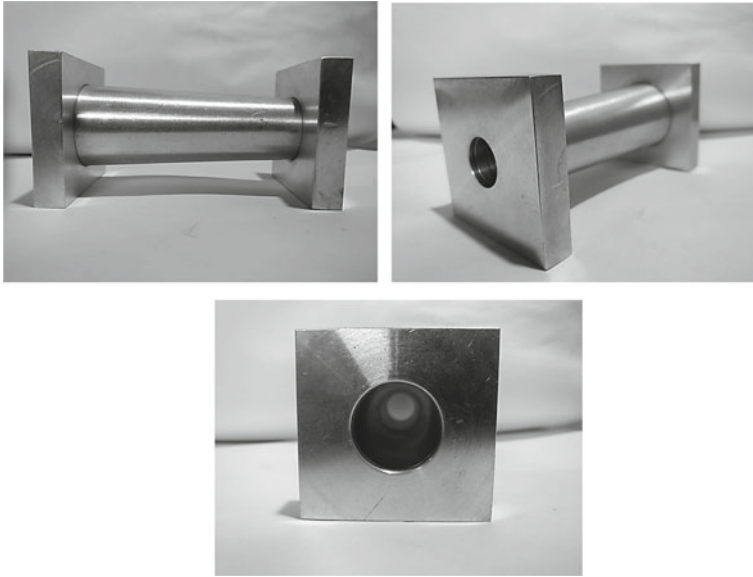
Fig. 4 Cross-sectional and side views of designed tapered waveguide of dimensions: Initial radius—3.73 mm, Final radius—7.37 mm, Length—10 cm, Material—Aluminum

For the non-tapered configuration, the waveguide radius and the background magnetic field can be chosen so that the device operates in narrow bands of frequencies over the scheme of concurrent tapering of the waveguide cross section and the applied DC magnetic field choosing an appropriate value of the taper parameter Ψ.

For a tapered configuration with the same mean waveguide radius, one must choose to change the taper angle Ψ in two ways: (i) by changing the initial waveguide radius $r_W\{0\}$ and the final waveguide radius $r_W\{l\}$, while keeping the interaction length to be constant, and (ii) by changing the interaction length but keeping the initial and final waveguide radii $\{r_W\{0\}, r_W\{l\}\}$, each to be constant. In both these ways, one can control the gain and bandwidth by the choice of the taper angle Ψ. Thus by choosing taper parameters, we can widen the device bandwidth. Dispersion diagram for the tapered waveguide operating in TEm1 mode is shown in Fig. 5. For TE_{0n} modes, i.e., for azimuthally symmetric modes, it is shown in Fig. 6. It is evident from inspection of Figs. 5 and 6 that a tapered structure has more potential to offer wide device bandwidths. Moreover, it is also worth noting the aspects of mode competition of the tapered device with those of the non-tapered. In general, the tapered device has to face more mode competition than the non-tapered. Finally, it is felt that the tapered-waveguide scheme which has wide device bandwidths has the drawback of yielding poor gain. It is felt that it is not difficult to use the present approach of the effects of tapering on high-gain gyro-TWT structures, like the vane-loaded (magnetron-like) structure, for wide device bandwidths at reasonably high gains.
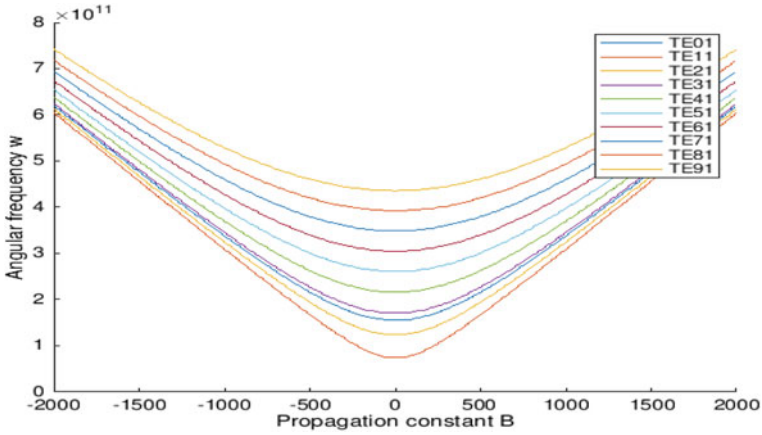
Fig. 5 Dispersion plot for the taper waveguide. The plot is for the TEm1 mode
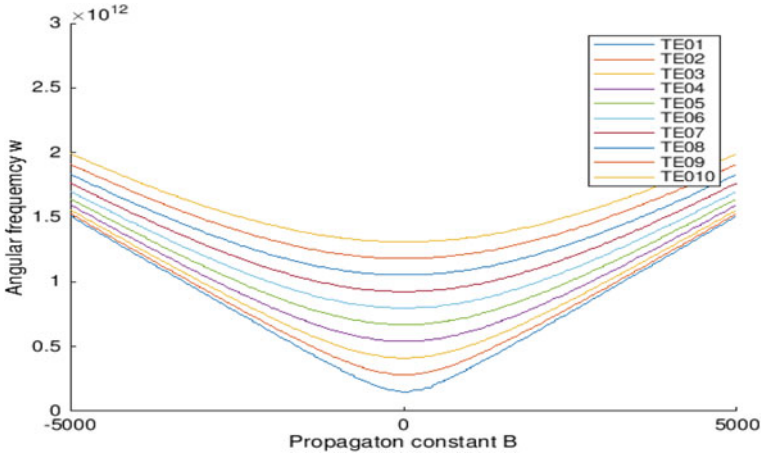
Fig. 6 Dispersion plot for taper waveguide. The plot is for the TE_0p mode, i.e., n = 0 and p is varied

3.2   Vane-Loaded Structure

A vane-loaded RF interaction structure has been designed. Cross-sectional and side views of the vane-loaded structure are shown in Fig. 7. The dispersion relation of the vane-loaded waveguide is solved using MATLAB simulation. First of all, the cutoff frequency of the structure is obtained by taking axial phase propagation constant equal to zero. Using this cutoff value as a guess point, the dispersion equation is solved for $\beta > 0$. Thus, the dispersion curve is obtained for the structure.
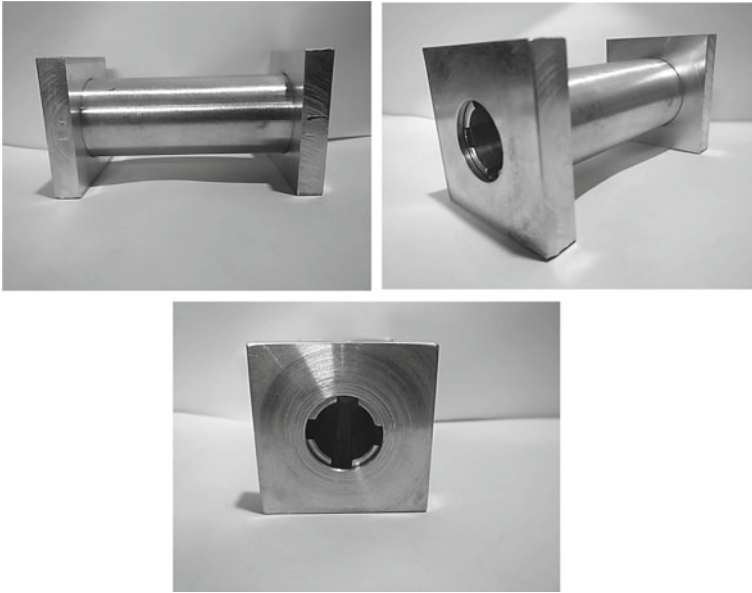
Fig. 7 Cross-sectional and side views of designed vane-loaded waveguide

Now, in order to the know the effect of structure parameter variation on the dispersion curve, any one structure parameter is varied and other remaining two structure parameters are kept constant. Thus, three dispersion curves are obtained for the three structure parameters, namely, the number of vanes N, the vane depth r_V/r_W, and vane angular width ϕ (Figs. 8,9, and 10).

As the value of N_v is increased or decreased (Fig. 8), or that of ϕ_v is increased or decreased (Fig. 10), the ω–β plot of the waveguide goes downward or upward respectively, in the ω scale, and also, correspondingly, the cutoff frequency ω_{cut} decreases or increases, respectively. But, with the increase or decrease of r_V/r_W (Fig. 9), the dispersion curve goes downward or upward, respectively. Here, it is noteworthy that the dispersion curve merely shifts upward or downward without any appreciable change in the slope of the dispersion curve. However, it is observed that the cutoff frequency is changed with the change in the value of structure parameters. Thus, it is concluded that the vane-loaded structure does not allow any appreciable change in the slope of the dispersion curve whereas it certainly changes the cutoff frequency. So, a vane-loaded structure can be used for mode separation.
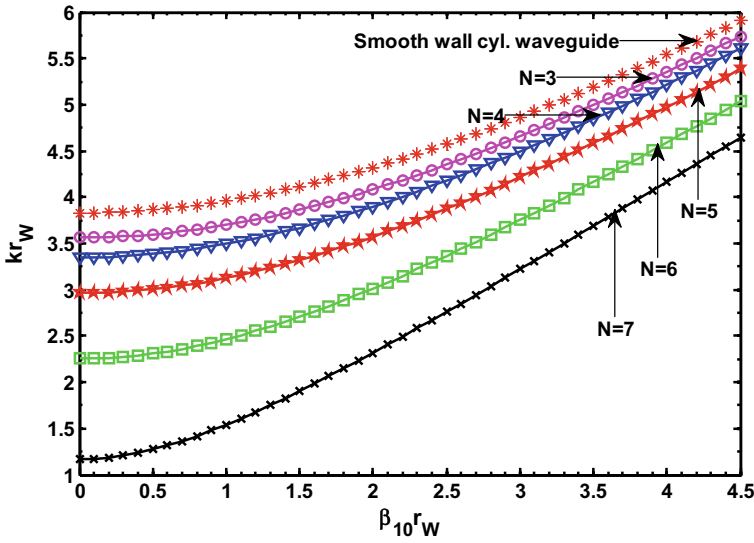
Fig. 8 Dispersion characteristics of a vane-loaded circular waveguide excited in the TE_{01} mode with structure parameters $r_v/r_w = 0.8$ and $\emptyset = 45°$; taking the number of vanes N as the parameter
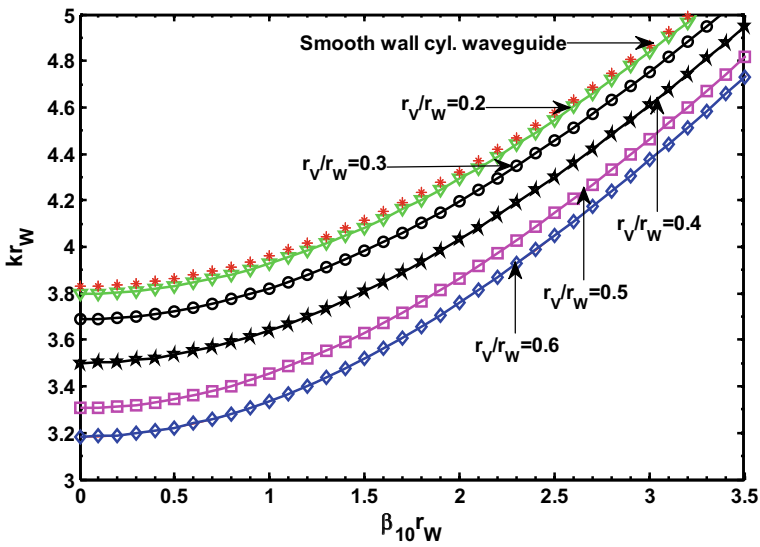
Fig. 9 Dispersion characteristics of a vane-loaded circular waveguide excited in the TE_{01} mode with structure parameters $N = 4$ and $\emptyset = 45°$; taking the relative vane depth r_V/r_W as the parameter
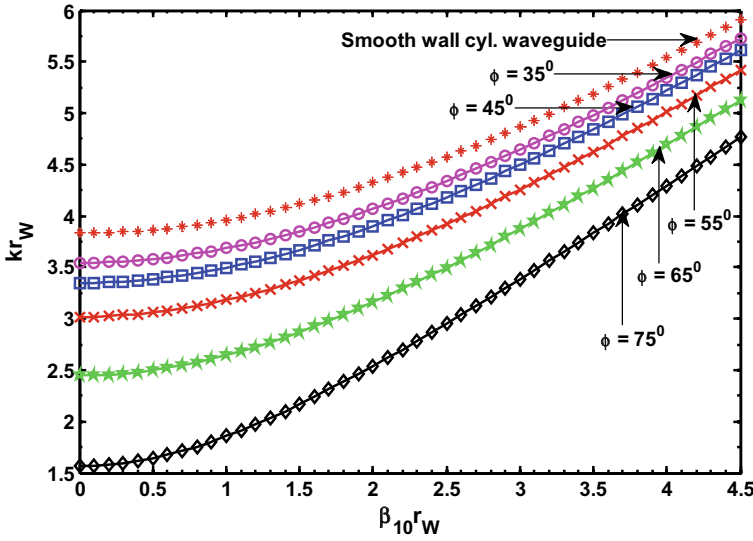
Fig. 10 Dispersion characteristics of a vane-loaded circular waveguide excited in the TE_{01} mode with structure parameters N = 4 and $r_v/r_w = 0.8$; taking the vane angle Ø as the parameter

3.3 Disc-Loaded Structure

A MATLAB code has been developed for the dispersion Eq. (107) of disc-loaded structure. Again, a similar algorithm has been used as that of the vane-loaded structure. Taking axial phase propagation constant equal to zero, the cutoff frequency of the disc-loaded structure is obtained for a set of structure parameter. Further, this cutoff value is used as an initial guess point for getting a set of data for the operating frequency. For a special case of $r_D/r_W = 1$, the dispersion Eq. (107) of disc-loaded RF interaction structure becomes a dispersion equation of a smooth wall circular cylindrical waveguide. Thus, the derived expression (107) is valid.

Structure parameters, namely, disc hole radius r_D/r_W, disc repetition L/r_W, and disc thickness T/r_W are varied one by one taking other two as a constant and the dispersion curve is obtained. In this way, the effect of parameters is studied (Figs. 11, 12, and 13). From the observation of Fig. 11, it is found that lower edge frequencies are almost the same while the upper edge frequencies are lowered. In this way, the dispersion curve becomes more flat with the increase in the value of L/r_W.

Figure 12 shows variation in the behavior of dispersion curve taking the disc hole radius r_D/r_W as a parameter. Here, it is observed that the increase in the value of r_D/r_W lowers the value of cutoff frequency but the upper edge frequency is almost the same. Thus, a little flatness in the dispersion curve is obtained. In other words, it can be said that disc hole radius does not appreciably widen the bandwidth of the structure.
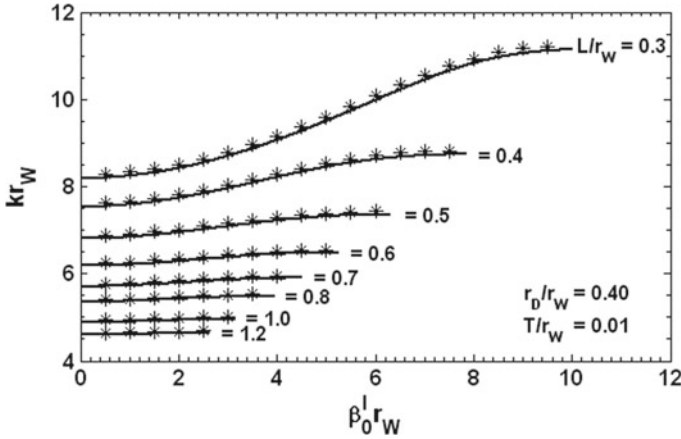
Fig. 11 Dispersion curve of a disc-loaded RF interaction structure taking the relative structure periodicity L/r_W as the parameter, and disc hole radius r_D/r_W and disc thickness T/r_W constant

Fig. 12 Dispersion curve of a disc-loaded RF interaction structure taking the relative structure disc hole radius r_D/r_W as the parameter, and periodicity L/r_W and disc thickness T/r_W constant
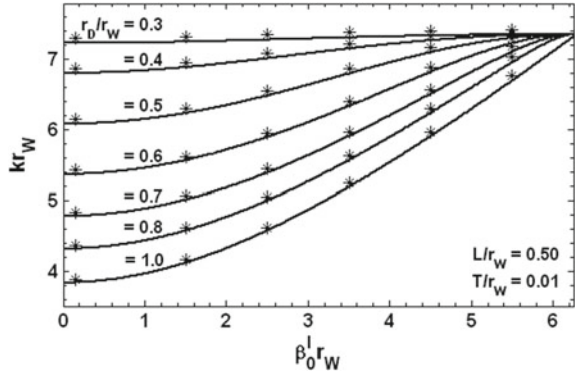
Fig. 13 Dispersion curve of a disc-loaded RF interaction structure taking the relative structure disc thickness T/r_W constant as the parameter, and disc hole radius r_D/r_W and periodicity L/r_W constant
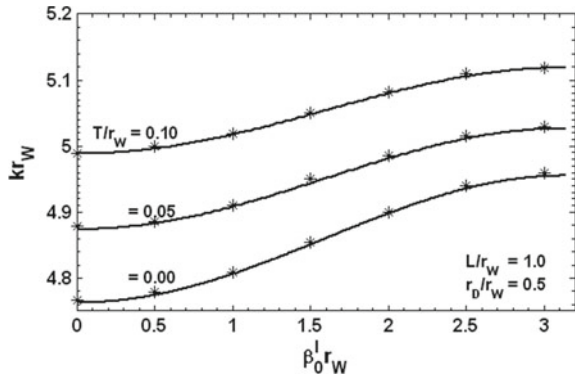
The effect of disc thickness on the dispersion curve is shown in Fig. 13. It is observed that disc thickness does not influence the dispersion characteristics up to an appreciable level. But, here it is important to mention that disc thickness plays a very crucial role in deciding the power handling capability of the device.

4 Conclusion

This study reveals that the dispersion characteristic analysis of any RF structure is very important for getting information about the available bandwidth. The structure parameters play a very important role in widening the device bandwidth. For a tapered structure, it has been observed that the tapering angle alters the bandwidth of the structure but at the cost of the device gain. Similarly, a vane-loaded RF interaction provides mode separation, thus, can be used in mode rarefaction. Likewise, it is observed that the dispersion characteristics of disc-loaded RF interaction structure are highly influenced by the disc periodicity. So, a judicious choice of structure parameters can be made to achieve the wideband performance of the gyrotron device.

References

1. Granatstein, V.L., Parker, R.K., Armstrong, C.M.: Vacuum electronics at the dawn of the twenty-first century. Proc. IEEE **187**, 702–716 (1999)
2. Basu, B.N.: Electromagnetic Theory and Applications in Beam-Wave Electronics. World Scientific, Singapore (1996)
3. Kartikeyan, M.V., Borie, E., Thumm, M.K.A.: Gyrotrons High-Power Microwave and Millimeter Wave Technology. Springer, Germany (2004)
4. Sakamoto, K.: Gyrotron and millimeter wave technology. IEEE Trans. Plasma Sc. **34**, 635–639 (2006)
5. Thumm, M.: History, Present and Future of Gyrotrons. International Vacuum Electronics Conference (IVEC – 2009), Rome, Italy (2009)
6. Park, G.S., Park, S.Y., Kyser, R.H., Armstrong, C.M., Ganguly, A.K.: Investigation of the stability of a tapered Gyro-TWT amplifier. In: International Electron Devices Meeting, IEDM'91, Technical Digest, pp. 779–781 (1991)
7. Park, G.S., Park, S.Y., Kyser, R.H., Armstrong, C.M., Ganguly, A.K., Parker, R.K.: Broadband operation of a Ka-Band tapered Gyro-traveling wave amplifier. IEEE Tran. Plasma Sci. **22**(5), 536–543 (1994)
8. Leou, K.C., McDermott, D.B., Luhmann, N.C.: Dielectric-Loaded wideband gyro-TWT. IEEE Trans. Plasma Sci. **20**(3), 188–196 (1992)
9. Bratman, V.L., Gross, A.W., Denisov, G.G., He, W., Phelps, A.D.R., Ronald, K., Samsonov, S.V., Whyte, C.G., Young, A.R.: High-gain wide-band gyrotron travelling wave amplifier with a helically corrugated waveguide. Phys. Rev. Lett. **84**, 2746–2749 (2000)
10. Choe, J.Y., Uhm, H.S.: Theory of gyrotron amplifiers in disc or helix-loaded waveguides. Int. J. Electronics **53**, 729–741 (1982)

11. Cooke, S.J., Denisov, G.G.: Linear theory of wide-band gyro-TWT amplifier using spiral waveguide. IEEE Trans. Plasma Sci. **26**, 519–530 (1998)
12. Denisov, G.G., Bratman, V.L., Cross, A.W., He, W., Phelps, A.D.R., Ronald, K., Samsonov, S.V., Whyte, C.G.: Gyrotron travelling wave amplifier with a helical interaction waveguide. Phys. Rev. Lett. **81**, 5680–5683 (1998)
13. Uhm, H.S., Choe, J.Y.: Theory of gyrotron amplifier in a tape helix loaded waveguide. J. Appl. Phys. **54**, 4889–4894 (1983)
14. Takukdar, I., Tripathi, V.K.: Excitation of whistler modes in a sheath-helix loaded waveguide. J. Appl. Phys. **65**, 1479–1483 (1989)
15. Yue, L., Wang, W., Gong, Y., Zhang, K.: Analysis of coaxial ridged disk-loaded slow-wave structures for relativistic travelling wave tubes. IEEE Trans. Plasma Sci. **32**, 1086–1092 (2004)
16. Zhang, Y., Mo, Y., Zhou, X.: "Rigorous analysis of the disk-loaded waveguide slow-wave structures. Int. J. Infrared Millimeter Waves **24**, 525–535 (2003)
17. Amari, S., Vahldieck, R., Bornemann, J.: Analysis of propagation in periodically loaded circular waveguides. IEE Proc. Microw. Antennas Propag. **146**, 50–54 (1999)
18. Rekiouak, A., Cheo, B.R., Wurthman, G., Bates, C.: A slow wave structure for gyro-TWA H_{11} operation. IEEE Trans. Microw. Theor. Tech. **42**, 1091–1094 (1994)
19. Barroso, J.J., Correa, R.A., de Castro, P.J.: Gyrotron coaxial cylindrical resonators with corrugated inner conductor: theory and experiment. IEEE Trans. Microw. Theor. Tech. **46**, 1221–1230 (1998)
20. Zaginaylov, G.I., Iaremenko, S.S.: Efficient method for analysis of coaxial gyrotron cavity with corrugated inner insert. In: Proceedings of 41th EUMW, pp. 183–186 (2011)
21. Singh, S., Kartikeyan, M.V.: Analysis of a triangular corrugated coaxial cavity for megawatt-class gyrotron. IEEE Trans. Electron Devices **62**(7), 2333–2338 (2015)
22. Singh, S., Kartikeyan, M.V.: Analysis of plasma-loaded noncorrugated and triangular corrugated coaxial cavity. IEEE Trans. Electron Devices **63**(10), 4060–4066 (2016)
23. Singh, S., Kartikeyan, M.V.: Full wave analysis of coaxial gyrotron cavity with triangular corrugations on the insert. IEEE Trans. Electron Devices **64**(4), 1756–1762 (2017)

LTE-Advanced Carrier Aggregation for Enhancement of Bandwidth

Akshay Goyal and Krishan Kumar

Abstract The advancement in technology and smart gadgets have led to the demand for more data and higher data rates. Despite efficient use of spectrum, it is not able to achieve the data rate of 1 Gbps as required by Long Term Evolution Advanced (LTE-A). To achieve such a high data rate, this requires a higher transmission bandwidth which cannot be achieved by a single carrier. So, a new technique is introduced called carrier aggregation to aggregate two or more component carriers to achieve higher bandwidth. The carriers that are to be aggregated can be from continuous spectrum and can also be from different bands which result in three types of carrier aggregation, i.e., intra-band contiguous, intra-band noncontiguous, and inter-band carrier aggregation. The most important feature to move from LTE to LTE-A is carrier aggregation and the main advantage of this is that it is backward compatible. As per 3GPP release it is given that up to five component carriers can be aggregated to achieve a bandwidth of 100 MHz. Thus, in this paper the carrier aggregation up to three-component carriers by taking different number of downlink resource block for each aggregated carrier and comparative study of increment in the bandwidth will be carried out and the simulation results the enhancement of bandwidth using carrier aggregation.

Keywords Carrier aggregation · Component carriers · Bandwidth · LTE · 3GPP

1 Introduction

In recent years, it has been seen that the demand for mobile data is increasing tremendously and it is observed that it will continue to grow. It can be seen that today more than 2.6 billion mobile subscriptions exist worldwide and this number is growing continuously. It is well known that the mobile traffic is not only limited to the voice

A. Goyal (✉) · K. Kumar
National Institute of Technology, Hamirpur, Hamirpur 177005, India
e-mail: me.akshaygoyal@gmail.com

K. Kumar
e-mail: Krishan_rathod@nith.ac.in
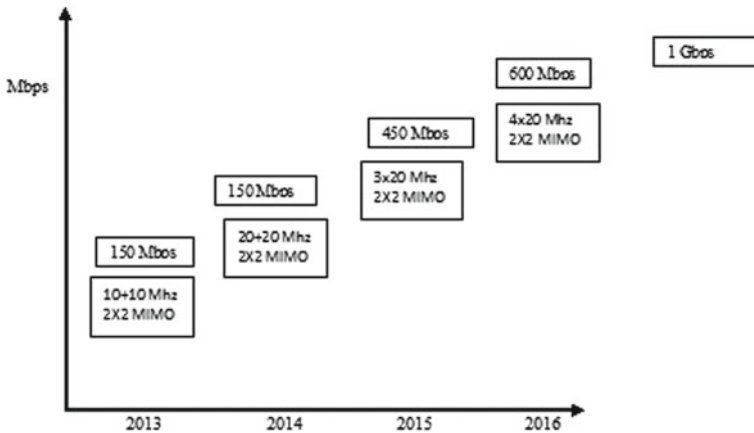
Fig. 1 Evolution of data rate with carrier aggregation [4]

calls but there are number of other purposes such as video calling, live streaming, and online gaming which urges for higher data rate. There is a challenge to address these rising data demands. So, this requires new technologies that can make use of segmented spectrum efficiently to work for this purpose. In today's world higher data rates and speed are required. So, one of the solutions to serve this purpose is to aggregate more than one carrier to achieve higher bandwidths known as Carrier Aggregation (CA) [1]. In LTE-advanced carrier aggregation refers combining multiple long-term evolution component carriers across the available spectrum [2]. There are number of bandwidths which can be aggregated like 1.4, 3, 5, 10, 15 and 20 MHz supported by LTE standards and will have 6, 15, 25, 50, 75, 100 Resource Blocks (RB), respectively [3]. CA also reduces interference as it says that only those base stations can use extra component carrier which does not cause much interference to the nearby cells based on downlink reference signal received power measurements which are known as autonomous component carrier selection. Figure 1 explains how carrier aggregation along with MIMO will help to attain higher data speeds in the future.

1.1 Related Work

Weihong Fu et al. in their paper discussed the CA in LTE systems and explained how it can help us to meet the demand for bandwidth. Along with that in their paper they proposed a resource scheduling algorithm to improve the performance of cell-edge users by providing a variable which limits the number of carriers each user can access and the weight factor depends upon the coverage weight factor and user category weight factor [5].

This paper is organized as follows. In Sect. 2 it provides system model and some useful features of LTE advanced. Section 3 provides performance analysis for the CA along with the steps for optimization. Section 4 provides design specification. Section 5 provides results and discussion. Section 6 finally concludes the paper.

2 System Model

2.1 LTE Advanced and Its Features

The LTE advanced mainly focusses on higher capacities and higher data rates. The main force which drives LTE toward LTE advanced is to provide higher data rates in a cost-efficient manner. To fulfill the IMT 4G requirements it is needed to move from LTE to LTE advanced [6]. The main features include increased data rates, higher spectral efficiency, capacity to accommodate more number of mobile subscribers, and performance improvement at the cell edges [7]. The main functionality added to LTE advanced is CA and enhanced use of MIMO systems.

2.2 Carrier Aggregation

Carrier aggregation is a mean through which 3GPP is aiming at higher data rates and higher bandwidths by aggregating multiple component carriers in both uplink and downlink directions [8]. Today 3GPP is allowing aggregation of carriers up to 5 components carriers, to achieve a bandwidth of 100 MHz. There can be number of scenarios based on how different bands and number of resource blocks within one or more band are chosen for aggregation by which carriers can be aggregated.

Three types of CA such as intra-band contiguous, intra-band noncontiguous, and inter-band carrier aggregation are explained and intra-band contiguous is used in this work. This provides us with the possibility of aggregating licensed band with the unlicensed band. This is the first time when the licensed band can be used along with unlicensed band to meet the user's requirements [9]. In this work only intra-band contiguous CA is analyzed (Fig. 2).

2.3 Intra-band Contiguous Carrier Aggregation

It is a continuous CA scheme in which multiple adjacent component carriers were used to achieve higher bandwidth and better data rates. In Fig. 3, intra-band contiguous carrier aggregation is represented. In this, two-component carriers are aggregated
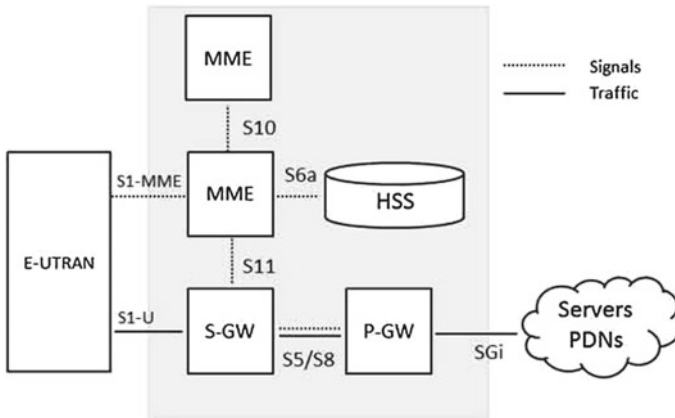
Fig. 2 LTE architecture
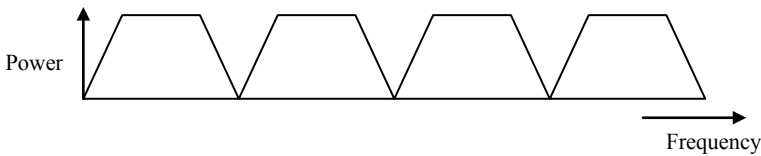
Fig. 3 Intra-band contiguous CA

from the same band which is continuous. For intra-band carrier aggregation supported bandwidths are 15 and 20 MHz in E-UTRA band 1 and 10, 15 and 20 MHz in Band 40 [10].

2.4 Intra-band Noncontiguous Carrier Aggregation

It a noncontinuous type of carrier aggregation scheme in which multiple component carriers are not continuous and may have different number of resource blocks and bandwidths can be aggregated to boost the performance of LTE system. In Fig. 4,
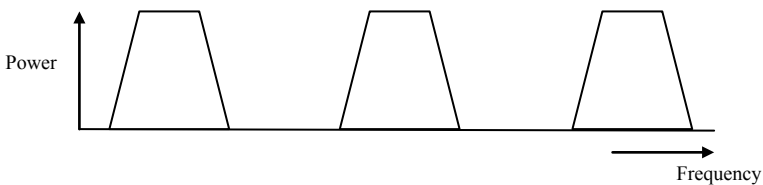
Fig. 4 Intra-band noncontiguous CA

intra-band noncontiguous carrier aggregation is represented. Noncontiguous aggregation is used in situations where it is not possible to assign continuous spectrum or we can say continuous spectrum is not available to be aggregated [11].

2.5 Inter-band Carrier Aggregation

In this type of carrier aggregation multiple component carriers which are to be aggregated are from different bands. In this user equipment implementation is very challenging as intermodulation and cross modulation may occur within the device when number of transmitters and receivers works together at the same time [12] (Fig. 5).

User equipment which has the capability of inter-band carrier aggregation has to decode the data over a large bandwidth and these types of aggregated carriers do not have same coverage [13].

3 Performance Analysis

Aggregated bandwidth is calculated with the help of given expression using Matlab [14]

$$\text{Aggregated bandwidth of the channel\_CA} = F_{\text{edge, high}} - F_{\text{edge, low}} \tag{1}$$

The lower bandwidth edge and the higher bandwidth edge are the reference points for the transmitters and receivers which is calculated as [14]

$$F_{\text{edge, low}} = FC_{\text{offset, low}} - F_{\text{offset, low}} \tag{2}$$

$$F_{\text{edge, high}} = FC_{\text{offset, high}} + F_{\text{offset, high}} \tag{3}$$

The lower and the higher frequency offsets entirely depend upon the component carrier configuration and the number of downlink resource blocks in the lowest and highest component carriers, which is calculated as [14]
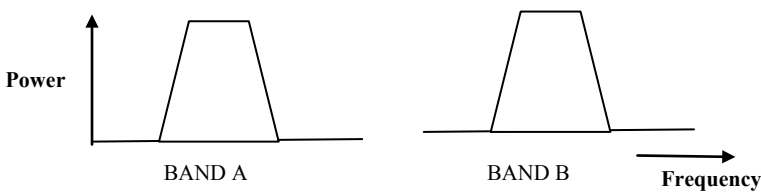
Fig. 5 Inter-band CA

$$F_{\text{offset, low}} = \left(0.18_{\text{NRB, Low}}\right)/2 + \text{NGBW} \tag{4}$$

$$F_{\text{offset, high}} = \left(0.18_{\text{NRB, High}}\right)/2 + \text{NGBW} \tag{5}$$

Here, the factor 0.18 is used for representing bandwidth of 1 RB in MHz, and NGBW stands for nominal guard bandwidth.

3.1 Steps to Optimize the Bandwidth

1. Choose number to component carriers to be aggregated.
2. Choose number of downlink resource blocks for the respective component carriers.
3. Configure corresponding eNode B for the given parameters.
4. Sample all the component carriers at a common sampling rate.
5. Calculate the low edge, high edge and center frequency and aggregated bandwidth.
6. At the receiver side choose the component carrier which is to be received.
7. Specify filter parameters as per the received carrier.
8. Check for CRC and display results.

4   Design Specification and Simulation for Carrier Aggregation

For each new component carrier of different or same name of resource blocks e-node B parameters needs to be configured according to the requirements of resource blocks which are described in Table 1.

First, two-component carriers are aggregated with the given parameters having 25 and 50 resource blocks, respectively, and the power spectrum aggregated waveform is shown in Fig. 6b. Component carrier 1 is of 5 MHz and eNode B 1 is configured with the following parameters represented in Table 2.

Now three-component carriers are aggregated with different number of resource blocks and bandwidth. Specification for three-carrier components aggregation with component carriers having 50, 75, and 100 resource blocks, respectively, are provided in Table 3.

Table 1 Number of resource blocks for a given bandwidth [3]

Bandwidth (MHz)	1.4	3	5	10	15	20
NRB	6	15	25	50	75	100
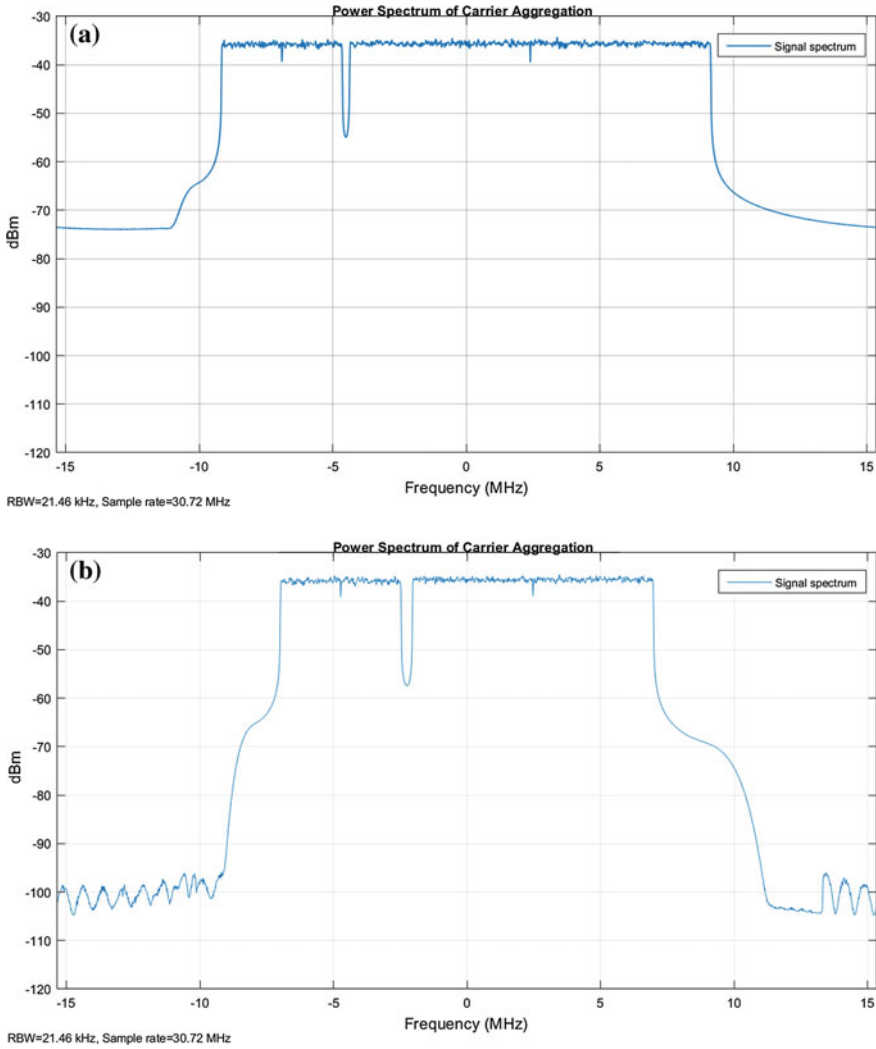
Fig. 6 Power spectrum of intra-band contiguous CA of 2 component carrier

5 Results and Discussion

The final comparison is carried out for bandwidth increment by carrier aggregation. Different types of carriers are aggregated with different number of downlink resource blocks. Table 1 shows the number of resource blocks used in different component carriers and corresponding bandwidth. First, two-component carriers are aggregated with the bandwidth 3 MHz each and 15-downlink resource blocks per aggregated component carrier, so the corresponding aggregated bandwidth 6 MHz is achieved

Table 2 Parameters for component carrier 1 and 2

Description	Metric value	Metric value
Number of downlink resource blocks	25	50
Cyclic prefix	Normal	Normal
NcellID	10	10
Duplexing mode	FDD	FDD
Sub carrier spacing (kHz)	15	15
Total number of subframes	10	10
CFI	1	2
Modulation	64 QAM	64 QAM

Table 3 Parameters for component carrier 1, 2 and 3

Description	Metric value	Metric value	Metric value
Number of downlink resource blocks	50	75	100
Cyclic prefix	Normal	Normal	Normal
NcellID	10	10	10
Duplexing mode	FDD	FDD	FDD
Sub carrier spacing (kHz)	15	15	15
Total number of subframes	10	10	10
CFI	1	2	3
Bandwidth (MHz)	10	15	20
Modulation	64 QAM	64 QAM	64 QAM

which can be seen from table. In second, the number of resource blocks of one of the carrier is changed and a corresponding change in the aggregated bandwidth is achieved. In final stage, three-components carriers were aggregated with a bandwidth of 20 MHz each and having 100 downlink resource blocks for each aggregated carriers and a resultant 59.6 MHz aggregated bandwidth. Figure 6a represents power spectrum for aggregated carriers having number of resource blocks 25 and 75 each and results in an aggregated bandwidth of 20 MHz. Similarly, Fig. 6b shows carrier aggregation of 5 and 10 MHz component carriers which results in an aggregated bandwidth of 15 MHz. Figure 7a, b represents aggregation of three-component carriers and results in an aggregated bandwidth of 30 MHz and 59.6 MHz, respectively. In Fig. 8, the aggregated bandwidth corresponding to the different NRBs and aggregated component carriers are represented.
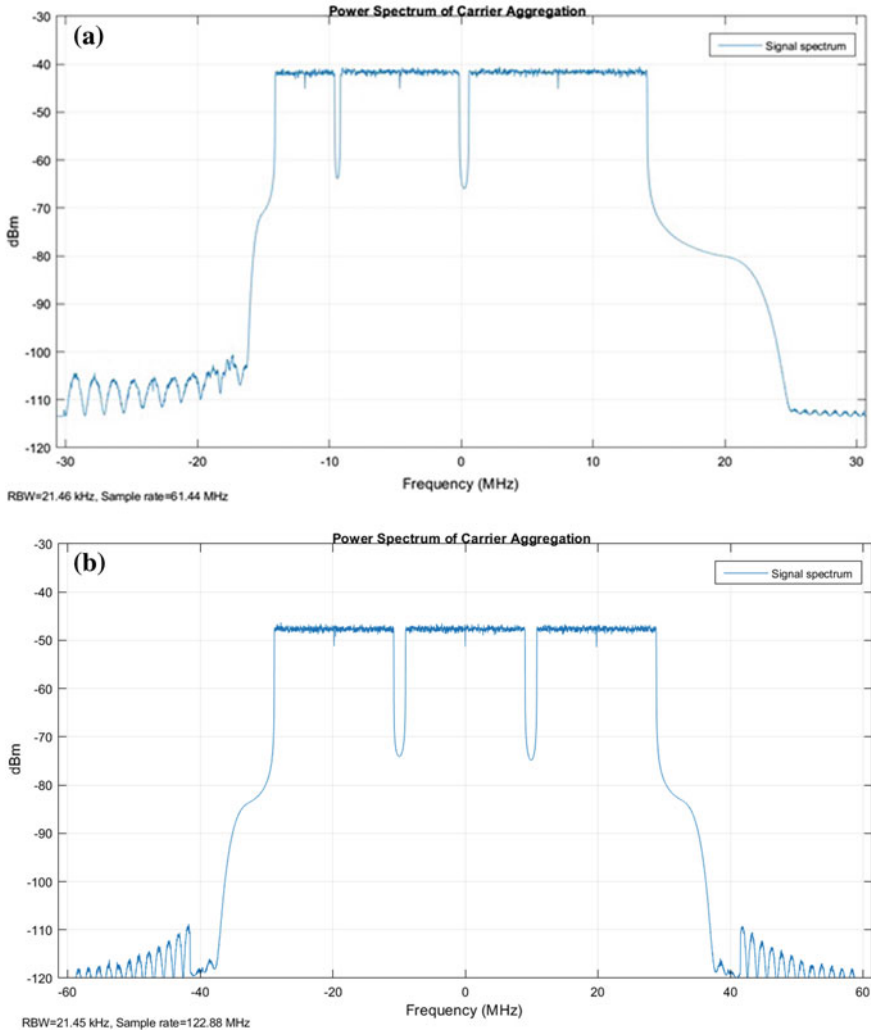
Fig. 7 Power spectrum of intra-band contiguous carrier aggregation of three-component carriers aggregation

6   Conclusion

In this work, three-component carriers are aggregated for LTE advanced for enhancement of bandwidth and this can be seen that component carriers of different bandwidths can be aggregated to get higher bandwidth. Table 4 supports that the number of RBs is increased in the component carriers which are to be aggregated and a corresponding increase in the aggregated bandwidth is achieved and the simulated results validate the theoretical study for enhanced bandwidth. As the number of
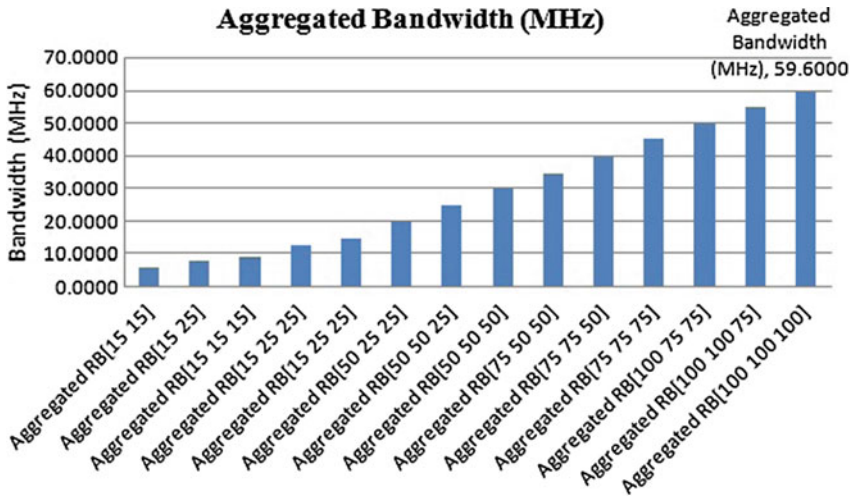
Fig. 8 Comparison of bandwidth for different number of resource blocks

Table 4 Aggregated component carriers and bandwidth

Aggregated carriers RB (No. of RB)	Aggregated bandwidth (MHz)
Aggregated RB [15 15]	6.0000
Aggregated RB [15 25]	8.0000
Aggregated RB [15 15 15]	9.0000
Aggregated RB [15 25 25]	12.8000
Aggregated RB [15 25 25]	14.6000
Aggregated RB [50 25 25]	19.7500
Aggregated RB [50 50 25]	24.8500
Aggregated RB [50 50 50]	29.8000
Aggregated RB [75 50 50]	34.6500
Aggregated RB [75 75 50]	39.7500
Aggregated RB [75 75 75]	45.0000
Aggregated RB [100 75 75]	49.8500
Aggregated RB [100 100 75]	54.6500
Aggregated RB [100 100 100]	59.6000

mobile subscribers is increasing continuously and the demands for higher data rates are raising this carrier aggregation can be a fruitful alternative to address all such type of problem and research issues.

References

1. 3rd Generation Partnership Project: LTE; Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (3GPP TR 36.913 version 11.0.0 Release 11), vol. 0, pp. 0–16 (2012)
2. Kumar, K., Prakash, A., Tripathi, R.: Spectrum handoff in cognitive radio networks: a classification and comprehensive survey. J. Netw. Comput. Appl. **61**, 161–188 (2016)
3. T. Specification: TS 136 101—V10.6.0—LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 10.6.0 Release 10), vol. 0 (2012)
4. N. Networks: LTE-Advanced Carrier
5. Fu, W., Kong, Q., Zhang, Y., Yan, X.: A Resource Scheduling Algorithm Based on Carrier Weight in LTE-Advanced System with Carrier Aggregation, no. 1, pp. 1–5
6. Ghosh, A., Ratasuk, R.: LTE-Advanced (2011)
7. Dahlman, E., Parkvall, S., Sköld, J.: Carrier Aggregation, 4g, LTE Evol. Road to 5G, pp. 309–330, Jan 2016
8. Fadlan, L.: Carrier Aggregation on Planning and Dimensioning LTE-Advanced in Bandung City, pp. 129–133 (2017)
9. Alkhansa, R., Artail, H., Gutierrez-Estevez, D.M.: LTE-WiFi carrier aggregation for future 5G systems: a feasibility study and research challenges. Procedia Comput. Sci. **34**, 133–140 (2014)
10. Shen, Z., Papasakellariou, A., Montojo, J., Gerstenberger, D., Xu, F.: Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications. IEEE Commun. Mag. **50**(2), 122–130 (2012)
11. Pedersen, K.I., Frederiksen, F., Rosa, C., Nguyen, H., Garcia, L.G.U., Wang, Y.: Carrier aggregation for LTE-advanced: functionality and performance aspects. IEEE Commun. Mag. **49**(6), 89–95 (2011)
12. Kumar, K., Prakash, A., Tripathi, R.: A spectrum handoff scheme for optimal network selection in cognitive radio vehicular networks: a game theoretic auction theory approach. Phys. Commun. (2017)
13. Park, C.S., Sundström, L., Wallén, A., Khayrallah, A.: Carrier aggregation for LTE-advanced: design challenges of terminals. IEEE Commun. Mag. **51**(12), 76–84 (2013)
14. T. Specification: TS 136 101—V13.3.0—LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 13.3.0 Release 13), vol. 0 (2016)

VLSI

Temperature-Dependent Analog, RF, and Linearity Analysis of Junctionless Quadruple Gate MOSFETs for Analog Applications

Prateek Kishor Verma [ID], **Akash Singh Rawat and Santosh Kumar Gupta** [ID]

Abstract In this paper, junctionless quadruple gate (JLQG) MOSFET is analyzed for its temperature-dependent characteristics. The junctionless MOSFET makes manufacturing simpler because it has no p-n junction as the doping of the channel is same as source/drain (S/D) region. Various performance parameters for short channel effects (SCEs), analog/RF, and linearity distortion of the JLQG MOSFET such as drain current (I_D), transconductance (gm_1), transconductance generation factor (*TGF*), output conductance (*gd*), early voltage (V_{EA}), intrinsic gain (*gm/gd*), cut-off frequency (f_T), gain frequency product (*GFP*), transconductance frequency product (*TFP*), gain trans-conductance frequency product (*GTFP*), second order derivative (gm_2), third order derivative (gm_3), second order voltage intercept point (*VIP2*), third order voltage intercept point (*VIP3*), third order input intercept point (*IIP3*), and third order intermodulation distortion (*IMD3*) with respect to temperature variations are presented and discussed. The study reveals the different zero-crossing points for temperature-dependent characteristics leading to guidelines for temperature-insensitive designs using JLQG MOSFETs.

Keywords Junctionless MOSFETs · Quadruple gate · Temperature · Linearity distortion · Analog · Radio frequency (RF)

P. K. Verma (✉) · A. S. Rawat · S. K. Gupta (✉)
Department of Electronics and Communication Engineering, Motilal Nehru National
Institute of Technology, Allahabad 211004, India
e-mail: prateek.kishor05@gmail.com

S. K. Gupta
e-mail: skg@mnnit.ac.in

A. S. Rawat
e-mail: akash09251991@gmail.com

1   Introduction

In recent years, semiconductor devices are used for a wide range of application at different temperatures. Nuclear energy and metallurgical sector require devices to be operated at high temperature whereas biomedical industry and space require low temperature [1, 2]. So there is a need of the devices which can perform efficiently at different temperature with optimum performance. Variation in temperature affects the electrical parameters of the device which affects the performance of the device [3]. So, we need to bias the device meant for different temperature applications at a point where it shows the least variation with respect to temperature. This bias point is known as zero-temperature coefficient point. In recent years, multi-gate MOS-FETs such as Double gate MOSFET, Fin-FETs, and the surrounding gate have been found superior to planar MOSFETs as the device channel length had become less than 45 nm [4–9]. It is difficult to fabricate the ultrasharp doping profile between source/drain region and channel region in nanometer range. To overcome this fabrication challenge Colinge et al. fabricated nanowire junctionless MOSFETs. JL-MOSFETs have better characteristics than conventional MOSFETs as they have S/D and body channel region with same type of high doping concentration. NMOS have high n-type and PMOS have high p-type doping with p-poly and n-poly gate [10, 11]. The investigation of analog/RF figure merits (FOMs) of JLDG MOSFET confirmed that junctionless MOSFET is better choice in comparison to junction-based MOSFET [11].

A JL transistor in which channel is controlled from all around leads to better performance compared to a JL transistor controlled by single, double, or triple gates. A junctionless quadruple gate (JLQG) MOSFET [12, 13] has been shown to fetch better RF and analog performance. In this proposed work, the temperature-dependent characteristics of JLQG MOSFETs by varying temperature from 200 to 400 K in terms of standard FOMs such as drain current (I_D), transconductance (gm_1), transconductance generation factor (TGF), output conductance (gd), early voltage (V_{EA}), intrinsic gain (gm/gd), cut-off frequency (f_T), gain frequency product (GFP), transconductance frequency product (TFP), and gain trans-conductance frequency product ($GTFP$), second order derivative (gm_2), third order derivative (gm_3), *VIP2*, *VIP3*, *IIP3,* and *IMD3*. We have also found the zero-temperature coefficient point for I_D, *gm1*, *gd*, V_{EA}, C_{gs}, f_T, and *GFP*. The ZTC_{Id}, ZTC_{gm1}, ZTC_{gd}, ZTC_{VEA}, ZTC_{Cgs}, $ZTC_{fT,}$ and ZTC_{GFP} are found to be 0.88 V, 0.72 V, 0.12 V, 0.84 V, 0.82 V, 0.7 V, and 0.7 V, respectively.

2   Device Structures

The schematic of the quadruple gate junctionless MOSFET is presented in Fig. 1. The device structure has uniform and similar doping concentration profile (N^+–N^+–N^+) for S/D and channel regions and the gate contact is made of P^+ polysilicon having a
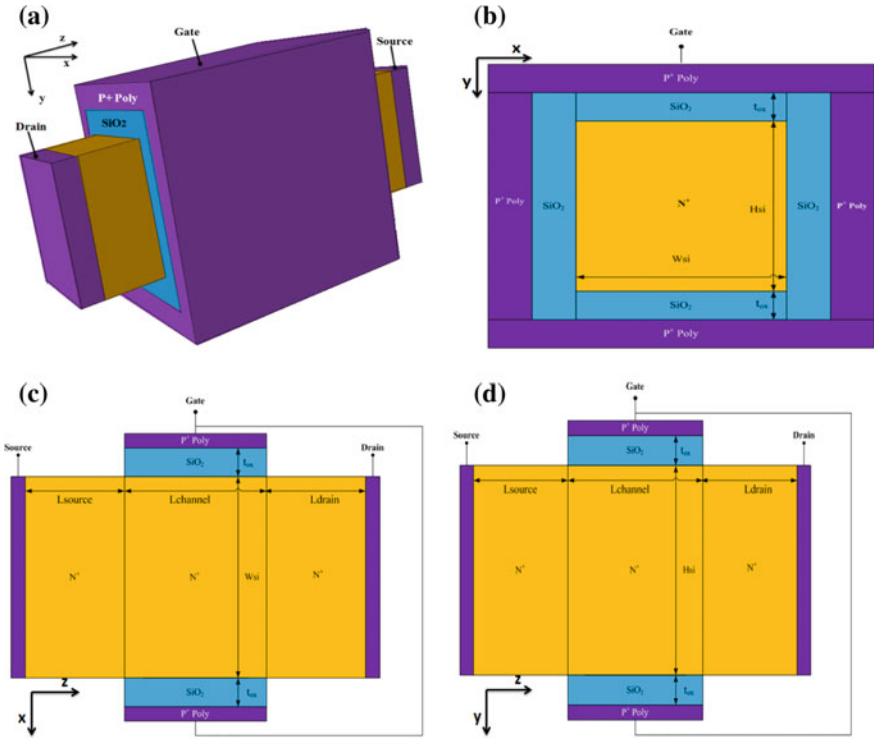
Fig. 1 Schematic diagram of n-type JLQG MOSFET, **a** 3D view, **b** cross sectional view along xy plane, **c** zx plane and **d** yz plane

work function of 5.2 eV. The physical device dimensions of the structure are taken with mentioned specifications; length of the channel (L_{ch}) is 20 nm; thickness of silicon (H_{si}) is 10 nm; width of silicon (W_{si}) is 10 nm; thickness of oxide (t_{ox}) is 2 nm; donor concentration (N_d) is 1×10^{19} cm^{-3} and S/D extensions are taken to be 5 nm each; then temperature is varied from 200 to 400 K with a step of 50 K by retaining other specifications same to investigate the variation in short channel effects, analog/RF performance, and linearity distortion performance parameters. The numerical analysis is carried out with the help of device simulation tool, i.e., Silvaco TCAD (ATLAS) [14], which is commonly used for characterization of the semiconductor devices in terms of electrical performances. Numerical calculations have been done with the help of different models such as Shockley-Read-Hall (SRH) recombination model, Lombardi model (CVT), concentration dependent mobility model (CONMOB), field-dependent mobility model (FLDMOB), and Boltzmann transport model. These models have been used to calculate the current and electrostatics of the device. There are no quantum effects that have been considered in this numerical analysis.

3 Temperature-Dependent Performance Analysis

The temperature-dependent performance analysis of JLQG is presented and analyzed in this section. First, the short channel effect parameters such as subthreshold swing (SS) and drain induced barrier lowering (DIBL) parameters are presented and discussed. In subsequent subsections various analog/RF performance parameters and linearity distortion analysis have been also discussed.

3.1 Temperature-Dependent Short Channel Effects (SCEs)

SCE parameters at different temperatures are shown in Table 1. SS is directly proportional to the temperature, i.e., it increases with temperature. At higher temperatures, SS value increases w.r.t. increasing V_{DS}. Its value increases by 190% for $V_{DS} = 0.1$ V and by 104% for $V_{DS} = 1.0$ V from 200 to 400 K. I_{on}/I_{off} ratio decreases as the temperature increases because of large increase in the OFF-current and decrease in the ON-current. The I_{off} increases approximately by a factor 10^8 from 200 to 400 K. I_{on} decreases by 50% for $V_{DS} = 0.1$ V and 24% for $V_{DS} = 1.0$ V from 200 to 400 K. This is due to the fact that minority current is high at high temperature since number of the high energy minority carrier increases at high temperature on the other hand the high energy majority carrier result in more collisions resulting in less drift velocity and hence less ON current.

DIBL increases by 27% from 200 to 400 K. DIBL increases with increase in temperature because of the fact that even when barrier lowering at a given drain voltage is same; the number of majority carriers with minimum energy to overcome that barrier increases. So, for lower short channel effects the device should be operated at lower temperatures.

3.2 Temperature-Dependent Analog Performance Analysis

In Fig. 2a drain current variations with increasing V_{GS} has been shown. It is observed that at the $V_{GS} = 0.88$ V, the current is same for all the temperatures called as "Zero Temperature coefficient (ZTC_{Id})" point. For V_{GS} less than ZTC_{Id} the drain current at higher temperature is higher but for V_{GS} more than ZTC_{Id} drain current at lower temperature becomes higher. The carriers at high temperature have higher energy so at high-temperature carriers have higher conduction at lower V_{GS} values but at higher V_{GS} value no. of collisions increases decreasing conduction.

The peak value of drain current decreases by 24% from 200 to 400 K as shown in Table 2. Transconductance variation with V_{GS} is shown in Fig. 2b. It has been observed that at $V_{gm1} = 0.72$ V, the transconductance is same for all temperatures called as "Zero Temperature coefficient (ZTC_{gm1})" point. The transconductance at

Table 1 SCE parameters at different temperatures

Temp. (K)	SS (mV/decade)		I_{off}		I_{on}		I_{on}/I_{off}		DIBL (mV/V)
	$V_{ds} = 0.1$ V	$V_{ds} = 1$ V	$V_{ds} = 0.1$ V	$V_{ds} = 1$ V	$V_{ds} = 0.1$ V	$V_{ds} = 1$ V	$V_{ds} = 0.1$ V	$V_{ds} = 1$ V	
200	31.708	45.618	1.46×10^{-22}	1.44×10^{-21}	8.68×10^{-6}	1.33×10^{-5}	5.93×10^{16}	9.22×10^{15}	43.634
250	56.516	57.205	3.65×10^{-19}	2.21×10^{-18}	6.59×10^{-6}	1.19×10^{-5}	1.81×10^{13}	5.40×10^{12}	46.421
300	68.395	69.661	1.57×10^{-16}	9.05×10^{-16}	5.41×10^{-6}	1.12×10^{-5}	3.45×10^{10}	1.23×10^{10}	52.896
350	79.901	80.89	6.13×10^{-15}	1.37×10^{-14}	4.74×10^{-6}	1.05×10^{-5}	7.74×10^{8}	7.64×10^{8}	53.204
400	91.895	93.085	5.32×10^{-14}	2.27×10^{-13}	4.34×10^{-6}	1.01×10^{-5}	8.15×10^{7}	4.47×10^{7}	55.548
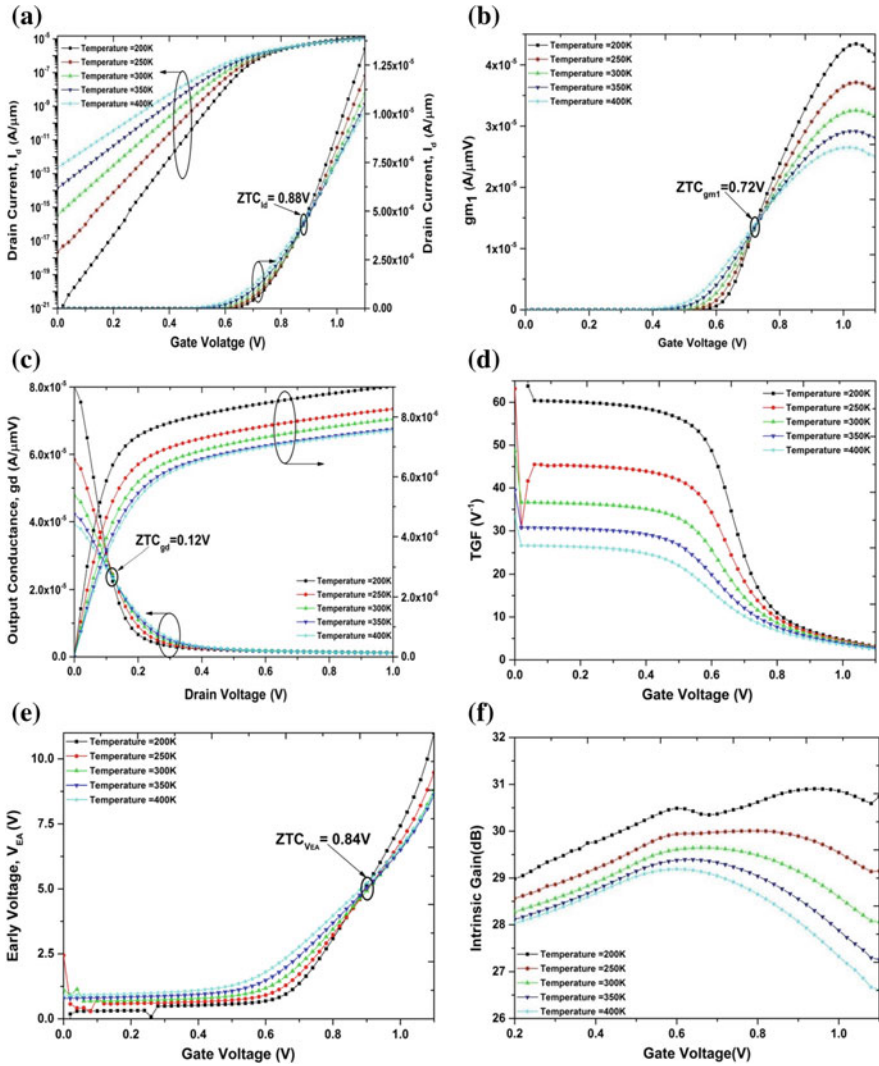
Fig. 2 **a** Drain current versus V_{GS} at $V_{DS} = 1$ V, **b** transconductance (g_{m1}) versus V_{GS} at $V_{DS} =$ 1 V, **c** drain current and output conductance versus V_{DS} at $V_{GS} = 1$ V, **d** *TGF* versus V_{GS} at V_{DS} $= 1$ V, **e** early Voltage versus V_{GS} at $V_{DS} = 1$ V and **f** intrinsic gain versus V_{GS} at $V_{DS} = 1$ V for different temperatures

Table 2 Peak values of analog parameters at $V_{DS} = 1.0$ V for different temperatures

Temp. (K)	200	250	300	350	400
Id (A/μm)	1.33×10^{-5}	1.19×10^{-5}	1.10×10^{-5}	1.05×10^{-5}	1.01×10^{-5}
gm1 (A/μmV)	4.34×10^{-5}	3.71×10^{-5}	3.26×10^{-5}	2.91×10^{-5}	2.65×10^{-5}
gd (A/μmV)	1.26×10^{-6}	1.28×10^{-6}	1.26×10^{-6}	1.23×10^{-6}	1.18×10^{-6}
TGF (V^{-1})	60.40991	45.47786	36.66163	30.76545	26.53385
V_{EA} (V)	11	9.47	8.81	8.6	8.57
Av (dB)	30.90318	30.00543	29.64708	29.39133	29.18896

higher temperature is higher for V_{GS} less than ZTC_{gm1} but for V_{GS} more than ZTC_{gm1} it becomes higher at lower temperatures. Maximum peak is observed at $V_{GS} = 1.04$ V at the temperature of 200 K. The Peak value of transconductance is decreased by 38.9% from 200 to 400 K as shown in Table 2. For better analog performance we need higher g_{m1} for which the device should be operated at lower temperature and higher V_{GS}.

In Fig. 2c variation in drain current and output conductance with increasing V_{DS} has been shown at $V_{GS} = 1$ V. Output conductance is same at $V_{DS} = 0.12$ V (zero-temperature coefficient point, ZTC_{gd}) for all the temperatures. For V_{DS} less than ZTC_{gd}; g_d is high for lower temperatures but for V_{DS} more than ZTC_{gd}, it is high for the higher temperatures. For better analog performance, lower values of g_d are required to achieve higher gains. So, for higher gains (i.e., lower g_d) the device should be biased at higher V_{DS} and lower temperatures. Figure 2d–f present the variations of TGF, early voltage (V_{EA}) and intrinsic gain (A_V) with increasing V_{GS}. TGF decreases as temperature increases which means there is less gain per unit power dissipation available at high temperatures. Early voltage at V_{GS} less than Zero-Temperature coefficient point, ZTC_{VEA} ($= V_{GS} = 0.84$) is high at higher temperatures (viz., 400 K) but after this early voltage is higher at the lower temperatures (viz., 200 K). Gain is high at lower temperatures (viz., 200 K) and lower at higher temperatures (viz., 400 K). Peak values of the TGF, V_{EA} and Gain (A_V) reduce by 56%, 22% and 5.54%m respectively, from 200 to 400 K.

3.3 Temperature-Dependent RF Performance Parameters

Figure 3a represents C_{gs} and C_{gd} variation with respect to V_{GS}. The extraction of C_{gs} and C_{gd} is done using AC small-signal analysis after getting DC solutions at $V_{DS} = 1$ V. The C_{gs} at gate voltage less than ZTC_{Cgs} ($V_{gs} = 0.82$ V) is high for the higher temperatures but after this point lower temperatures (viz., 200 K) provides high C_{gs}. The cut-off frequency (f_T) variation with respect to V_{GS} has been shown in Fig. 3b. f_T at higher temperatures is high for the V_{GS} less than ZTC_{ft} ($V_{GS} = 0.7$ V) after that it becomes more for lower temperatures. Peak value of f_T decreases by 28.9% from 200 to 400 K as shown in Table 3. In Fig. 3c, GFP at higher temperatures is more
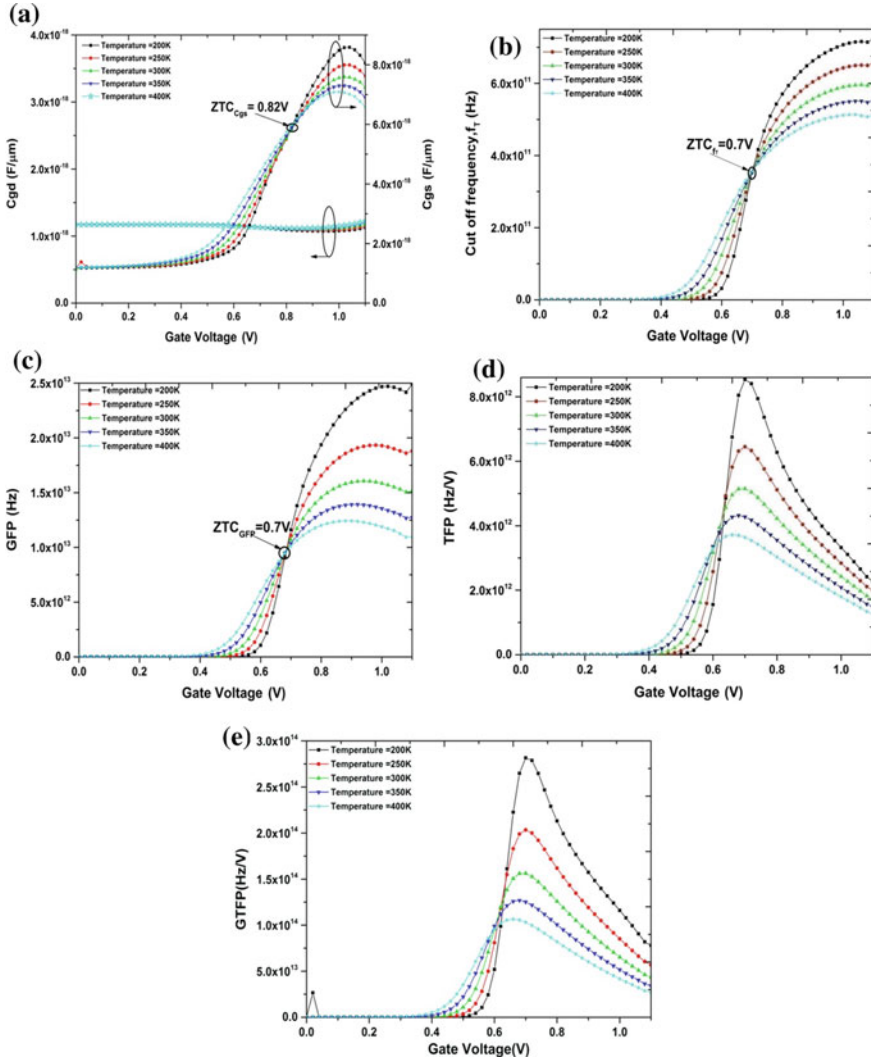
Fig. 3 **a** C_{gs} and C_{gd} versus V_{GS}, **b** cut-off frequency versus V_{GS}, **c** *GFP* versus V_{GS}, **d** *TFP* versus V_{GS}, **e** *GTFP* versus V_{GS}; for different temperatures at $V_{DS} = 1.0$ V

for low V_{GS} values up to ZTC_{GFP} ($V_{gs} = 0.7$ V) after that *GFP* is more for lower temperatures. Peak value of the *GFP* decreases by 50% from 200 to 400 K. *TFP* as shown by Fig. 2d is high for higher temperatures till ZTC_{TFP} ($V_{GS} = 0.6$ V) but after that at lower temperatures, *TFP* starts increasing and peak value is found at $V_{GS} = 0.7$ V. The peak values at other temperatures are also around $V_{GS} = 0.7$ V but the absolute value of *TFPs* reduce with temperature. Peak value of the *TFP* decreases by 56.4% from 200 to 400 K. *GTFP* shown in Fig. 3e is high at higher temperatures

Table 3 Peak values of RF parameters for different temperature at $V_{DS} = 1.0$ V

Temp. (K)	200	250	300	350	400
C_{gd} (F/µm)	1.17×10^{-18}	1.17×10^{-18}	1.17×10^{-18}	1.20×10^{-18}	1.22×10^{-18}
C_{gs} (F/µm)	8.59×10^{-18}	8.00×10^{-18}	7.59×10^{-18}	7.30×10^{-18}	7.09×10^{-18}
f_T (Hz)	7.23×10^{11}	6.56×10^{11}	6×10^{11}	5.52×10^{11}	5.14×10^{11}
GFP (Hz)	2.49×10^{13}	1.93×10^{13}	1.61×10^{13}	1.39×10^{13}	1.24×10^{13}
TFP (Hz/V)	8.54×10^{12}	6.45×10^{12}	5.15×10^{12}	4.32×10^{12}	3.72×10^{12}
GTFP (Hz/V)	2.82×10^{14}	2.03×10^{14}	1.56×10^{14}	1.27×10^{14}	1.06×10^{14}

till ZTC_{GTFP} ($V_{GS} = 0.6$ V) but after that at lower temperatures (viz., 200 K); *GTFP* starts increasing and peak value is found at $V_{GS} = 0.7$ V. The peak values of *GTFP* shift toward lower gate voltage as temperature increases. The peak value of *GTFP* decreases by 62.4% from 200 to 400 K.

3.4　Temperature-Dependent Linearity Distortion Analysis

Figure 4a and b represent the higher order transconductance parameters (gm_2 and gm_3) with respect to gate voltage for different temperatures. The value of gm_2 is high for high temperature (400 K) till $V_{GS} = 0.62$ V after that for low temperature it starts increasing and peak value is obtained at the $V_{GS} = 0.7$ V. The peak value reduces by 52% from 1.52×10^{-4} at 200 K to 7.27×10^{-5} at 400 K. For low second order harmonics we need less absolute value of gm_2. The gm_3 is high at high temperature (400 K) till $V_{GS} = 0.5$ V after that the low temperature (200 K) value starts increasing and positive peak is achieved at $V_{GS} = 0.64$ V. Positive peak absolute value reduces by 73.6% from 1.66×10^{-3} at 200 K to 4.39×10^{-4} at 400 K. For $V_{GS} > 0.7$ V; gm_3 decreases and least negative value is found at $V_{GS} = 1.04$ V. Negative peak absolute value reduces by 56.7% from -1.33×10^{-3} at 200 K to -5.76×10^{-4} at 400 K. For low third order harmonics we need less absolute value of the gm_3.

Figure 4c–f represent *VIP2*, *VIP3*, *IIP3,* and *IMD3* variations with respect to V_{GS} for different temperatures, respectively. *VIP2* is the measure of second order harmonic distortions of which value is needed to be higher for lower second order harmonic distortions. *VIP2* is high at higher temperatures which mean lower second order harmonic distortion is present at higher temperatures. At $V_{GS} = 0.6$ V the *VIP2* value increases by 273% from 200 to 400 K as shown in Table 4.

Absolute value of *VIP2* increases with gate voltage increase for all temperatures. Therefore, for less second order harmonic distortions, we may operate the device at higher gate voltage and a relatively higher temperature. *VIP3* is the measure of third order harmonic distortions and a higher value for less third order harmonic distortions are desirable. *VIP3* value is high at higher temperatures. For $V_{GS} = 0.6$ V, the *VIP3* value at 400 K becomes five times of the value at 200 K as shown in Table 4.
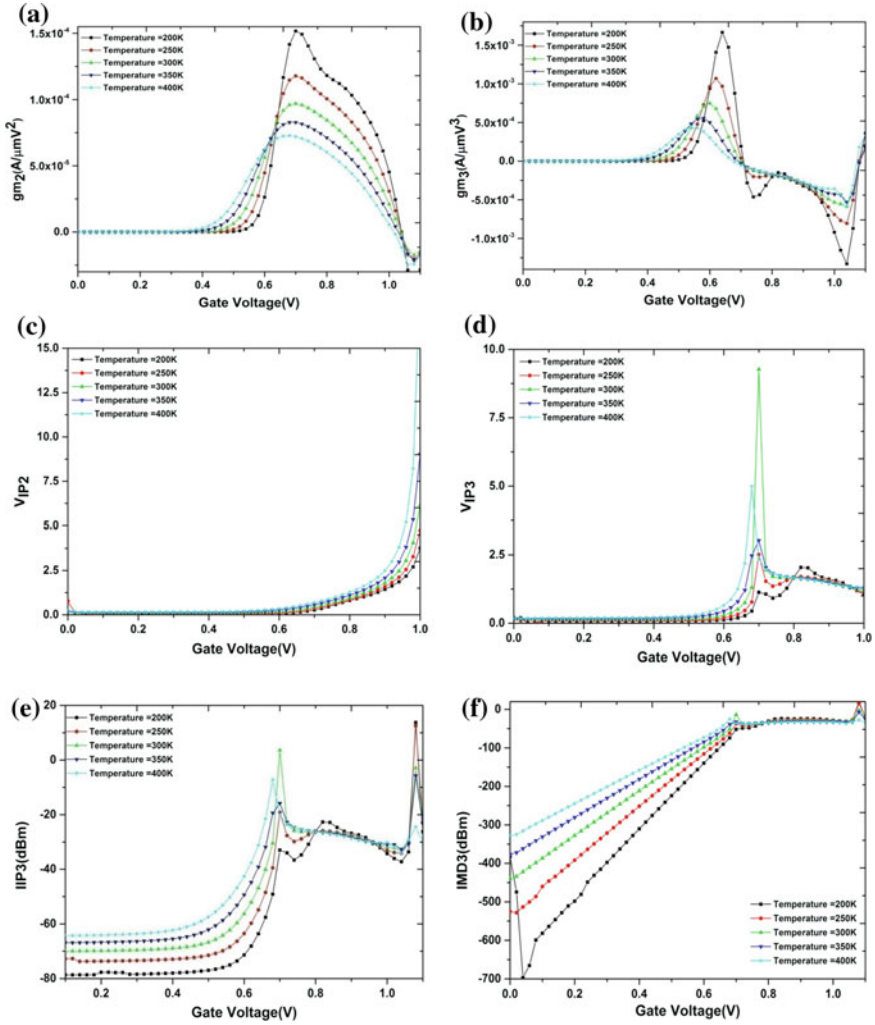
Fig. 4 **a** gm_2, **b** gm_3, **c** *VIP2*, **d** *VIP3*, **e** *IIP3* and **f** *IMD3* versus gate voltage at $V_{DS} = 1$ V for different temperatures

Table 4 Peak values of linearity distortion parameters at $V_{GS} = 0.6$ V and $V_{DS} = 1.0$ V for different temperatures

Temp. (K)	200	250	300	350	400
VIP2	0.092	0.137	0.193	0.263	0.343
VIP3	0.124	0.195	0.295	0.440	0.649
IIP3 (dBm)	−71.36	−63.52	−56.29	−49.38	−42.61
IMD3 (dBm)	−139.71	−115.80	−98.50	−84.77	−73.06

Third order input intercepts point (*IIP3*) and third order intermodulation distortions (IMD3) are also high at higher temperatures and for better linearity, their value needs to be as high as possible. At $V_{GS} = 0.6$ V the *IIP3* and *IMD3* values increase by 40.3% and 47.7%, respectively from 200 to 400 K as shown in Table 4. This is attributed to the improved gate controllability due to the use of quadruple gates, enhancing the carrier transport efficiency and hence improved transconductance gm_1.

4 Conclusion

In this present work, a device structure of junctionless quadruple gate (JLQG) has been proposed and performance for analog/RF circuits with linearity distortion analysis has been also investigated for variations of temperature. The improved sub-threshold swing (SS), I_{off}, I_{on}, I_{on}/I_{off} and DIBL have been observed due to decrease in temperatures. As the temperature decreases, analog performance parameters such as intrinsic gain (*Av*), *TGF* and peak value of early voltage (V_{EA}) increases. Decrease in temperature results in higher peak values of C_{gs}, cut-off frequency (f_T), *GFP*, *TFP*, and *GTFP*. The peak values of gm_2 and gm_3 lower as the rise in temperature, due to which, the linearity distortion parameters such as *VIP2*, *VIP3*, *IIP3,* and *IMD3* have been improved with increase in temperature and biasing voltage. Therefore, the study reveals that proposed JLQG may be used for improved analog applications.

References

1. Patterson, R.L., Dickman, J.E., Hammoud, A., Gerber, S.: Electronic components and circuits for extreme temperature environments. In: Proceedings of the IEEE Aerospace Conference, vol. 6, pp. 2543–2548 (2003)
2. Yu, B., Wang, H., Kim, H.-S., Xiang, Q., Lin, M.-R., Chang, L., Hu, C.: Proceedings of the International Symposium on VLSI Technology, Systems, and Applications, pp. 23–25 (2001)
3. Groeseneken, G., Colinge, J.P., Maes, H.E., Alderman, J.C., Holt, S.: Temperature dependence of threshold voltage in thin-film SOI MOSFETs. IEEE Electron Device Lett. **11**(8), 329–331 (1990)
4. Colinge, J.P.: Multi-gate SOI MOSFETs. Solid-State Electron. **48**, 897–905 (2004)
5. Colinge, J.P.: FinFETs and Other Multi-Gate Transistors. Springer, Berlin (2008)
6. Kumar, M.J., Orouji, A.A., Harshit, D.: New dual-material SG nanoscale MOSFET: analytical threshold-voltage model. IEEE Trans. Electron Devices **53**, 920–923 (2006)
7. Yu, B., Song, J., Yuan, Y., Lu, W.Y., Taur, Y.: A unified analytic drain-current model for multiple-gate MOSFETs. IEEE Trans. Electron Devices **55**, 2157–2163 (2008)
8. Chevillon, N., Sallese, J.M., Lallement, C., Pregaldiny, F., Madec, M., Sedlmeir, J., Aghassi, J.: Generalization of the concept of equivalent thickness and capacitances to multigate MOSFETs modeling. IEEE Trans. Electron Devices **59**, 60–71 (2012)
9. Lee, C.W., Yun, S.R.N., Yu, C.G., Park, J.T., Colinge, J.P.: Device design guidelines for nano-scale MuGFETs. Solid-State Electron. **51**, 505–510 (2007)
10. Lee, C.W., Afzalian, A., Akhavan, N.D., Yan, R., Ferain, I., Colinge, J.P.: Junctionless multigate field-effect transistor. Appl. Phys. Lett. **94**(5), 053511 (2009)

11. Colinge, J.P., Lee, C.W., Afzalian, A., Akhavan, N.D., Yan, R., Ferain, I., Razavi, P., O'Neill, B., Blake, A., White, M.: Nanowire transistors without junctions. Nat. Nanotechnol. **5**(3), 225–229 (2010)
12. Rawat, A.S., Gupta, S.K.: Potential modeling and performance analysis of junction-less quadruple gate MOSFETs for analog and RF applications. Microelectron. J. **66**, 89–102 (2017)
13. Gupta, S.K, Rawat, A.S., Verma, Y.K., Mishra, V.: Linearity Distortion Analysis of Junctionless Quadruple Gate MOSFETs for Analog Applications, pp. 1–9. Springer, Silicon (2018)
14. Silvaco: ATLAS User's Guide (2018)

A Hardware Minimized Gated Clock Multiple Output Low Power Linear Feedback Shift Register

Digvijay Singh Mehta, Varun Mishra, Yogesh Kumar Verma and Santosh Kumar Gupta

Abstract As there is rapid increase in daily used battery-powered electronics equipment, and as these battery-powered equipments are able to work for a limited amount of time before requiring to recharge, there is ever increasing demand for long battery life (as run time on a full charge) that can be achieved by either increasing the battery capacity or reducing power consumption by the devices. In this paper, a switch minimized parallel LFSR with clock gating technique is proposed, and further optimization of circuit is performed by reducing number of gates (transistor) used by the circuit. Dynamic power consumption is reduced by minimizing the switching activity factor of the circuit, for which we utilize clock gating technique. Proposed circuit power consumption is compared with previous LFSR. The proposed circuit is implemented and simulated in cadence at 180 nm channel length, which verifies further reduction in power as compared to previous technique.

Keywords PRNG · LFSR · Low power · Switching activity · Hardware minimization

1 Introduction

In current scenario, there is rapid increase in the number of battery-powered electronics equipment. These equipment demand for longer battery life or longer recharging cycle. This objective can be achieved either by increasing battery capacity or by

D. S. Mehta · V. Mishra · Y. K. Verma · S. K. Gupta (✉)
Motilal Nehru National Institute of Technology, Allahabad, UP, India
e-mail: skg@mnnit.ac.in

D. S. Mehta
e-mail: digvijaycoer@gmail.com

V. Mishra
e-mail: varun20mishra@gmail.com

Y. K. Verma
e-mail: yogeshvermaec@gmail.com

367

reducing power consumed by the circuit. Thus, low power emerges as one of the most important parameter in present VLSI design scenario.

There are many electronic equipments, which widely uses pseudo-random number generators (PRNGs), like built-in self-test (BIST) data compression circuitry, error correction circuitry, encryption circuitry, and communication circuitry. There are many techniques for generating pseudo-random number sequence. Random number generator is mainly of two types (1) true random number generator and (2) pseudo-random number generators (PRNG). True random number generators (TRNG) are those, which generate non-deterministic pattern, and that generated pattern cannot be regenerated again. TRNG generate random number as a function of physical parameter (as it depends on noise level, photoelectric effect, and other quantum phenomena, which changes randomly and one cannot predict their next value). That is why they are also known as hardware random number generator. Whereas pseudo-random number generators provide a periodic output and their output pattern can be regenerated if one have access to some of its parameters. Due to its unpredictability, TRNG are used in application requiring nondeterministic pattern like lotteries, game, gambling, etc.

As we can easily reproduce sequence generated by a PRNG, given that we have access to the information about initial condition and algorithms used by PRNG to generate the next output bit. PRNG are mostly used in the circuit where same pattern as generated by first PRNGs required again like in communication system we require the same pattern at both transmitting and receiving side, same is applicable in encryption.

Linear feedback shift register (LFSR) is the most popular technique to implement a pseudo-random pattern generator. As shown in Fig. 1, it is implemented using N number of flip-flops (FFs) connected in series with a linear feedback operation performed by XOR or XNOR gates. The bit positions in the shift register that are
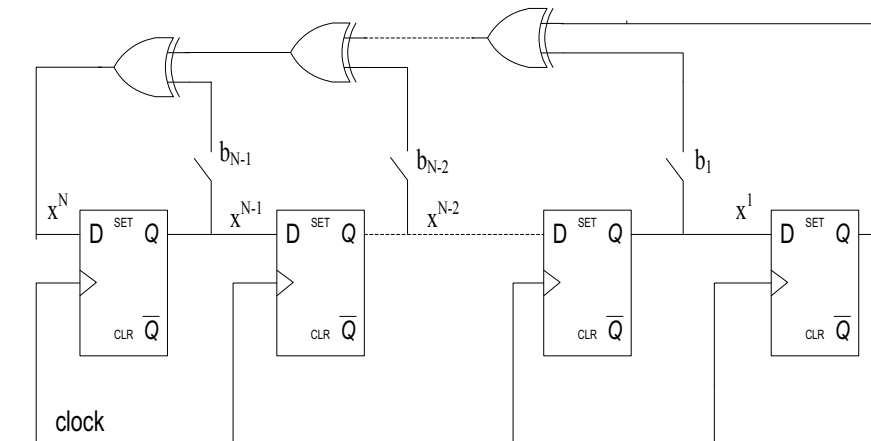
Fig. 1 Schematic diagram of serial LFSR showing multiplication coefficients for tap positions

used to generate next bit are called taps. Rightmost bit of serial LFSR provides output sequence. These taps positions are XOR'd or XNOR'D sequentially with the output bit. In addition, XOR/XNOR'D output is used as feedback to LFSR input.

In a serial LFSR at every clock cycle, clock path for each FFs toggle, results in significant amount of power consumption. Many techniques are proposed to address this issue, a parallel LFSR is proposed in which clock is only given to those FFs, which are required to store the next generated bit. Thus, there is reduction in number of FFs with active clock, which results in reduction of power consumption, but to implement a parallel LFSR we require N × M switches. For a normal LFSR it is quite large value, in [2, 3] a parallel LFSR with switch minimization procedure is proposed, by this procedure number of required switches are reduced to N + M, and a method is also proposed which produces more than one output simultaneously. In [1] further improvement is done by improving clock timing for the circuit. In [4], a simple LFSR using gated clock technique is proposed, which results in less power consumption. In this paper, we apply clock gating technique in multiple output low power parallel LFSR [1] and further optimize the circuit by reducing hardware requirement. Section 2 presents conventional LFSR and in Sect. 3 we discuss multiple output LFSR proposed in [1]. In Sect. 4, we discuss clock gating technique and propose a new circuit for LFSR in Sect. 5. Section 6 compares proposed circuit with previous LFSR.

2 Traditional LFSR

An LFSR is a shift register whose next input bit is generated by performing linear operation on some of its previously generated bit. Operation is performed on bits, which are currently stored in the shift register. The initial (starting) value of the LFSR is known as seed. A serial LFSR can easily be represented by its characteristic polynomial. Standard characteristic polynomial for an LFSR is as follows:

$$P(x) = 1 + b_1x + b_2x^2 + \ldots + b_{N-1}x^{N-1} + x^N \tag{1}$$

where power of x, i.e., N represents bit positions of tap in a shift register, and b represents whether that shift register bit is operating as tap position (used for generation of next bit) or not. If there is any connection between feedback path and certain shift register bit then $b_i = 1$, otherwise $b_i = 0$.

Output sequence produced by LFSR is pseudo-random in nature because it can easily be regenerated if one has access to information about seed value and characteristic polynomial. An N-bit LFSR can generate up to 2^{N-1} different patterns before repeating the sequence. LFSR, which can produce 2^{N-1} different patterns, are called as m-sequence LFSR.

3 Multiple Output Low Power Parallel LFSR

Conventional serial LFSR suffers from two main drawbacks: (1) for every clock cycle all the components of the circuit are clocked and (2) throughput of given circuits is limited to one bit per clock cycle. LFSR are designed using this technique when used in high-frequency applications that suffer from considerable amount of power loss. Thus, a new technique was proposed in [2] where clock is given to only those FFs, which require storing the next generated bit. In addition, as the number of FFs required to store next bit is very low compared to total number of FFs in use (depends on characteristic polynomial), there is reduction in number of FFs with active clock, resulting in reduction of power consumption. In [1, 2] many techniques are proposed to produce more than one output simultaneously. Number of outputs produced simultaneously by circuit depends on characteristic polynomial of LFSR and is given by

$$P(X) = 1 + X^{K_1} + \ldots + X^{K_{M-1}} + X^{K_M} \tag{2}$$

where M represents the number of XOR gates. An LFSR which is able to produce K_1 (highest polynomial bit) outputs for above characteristic polynomial is shown by Fig. 2.
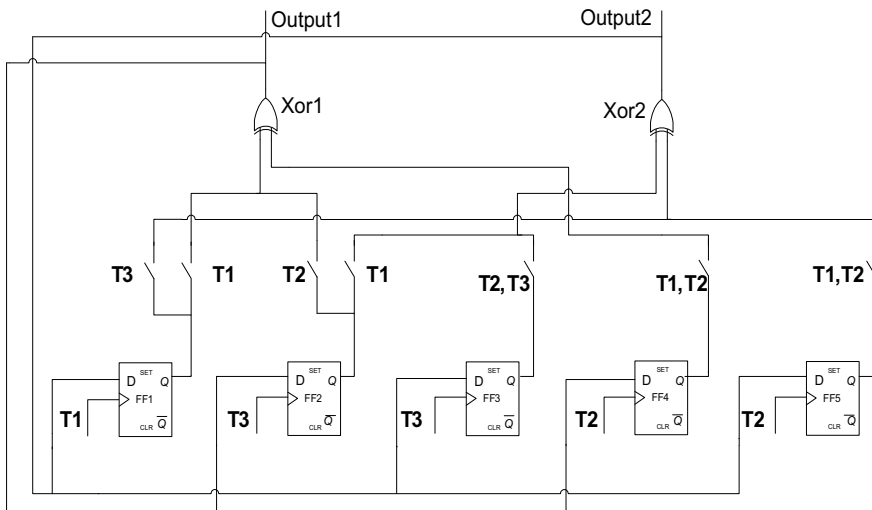
Fig. 2 Circuit diagram of multiple output low power parallel LFSR

Fig. 3 Implementation of clock gating technique for D FFs
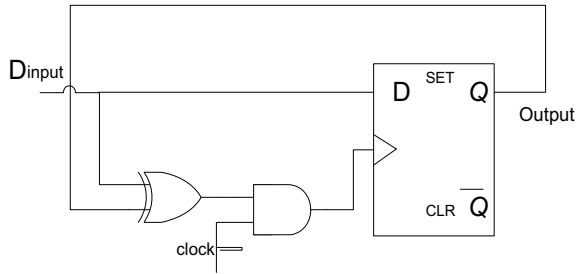
4 Clock Gated Circuits

Reduction in power consumption in digital systems can be achieved by some set of strategies called dynamic power management (DPM) [5–9]. The part of the circuit which is not performing any operation in a particular time frame is disabled in DPM. Thus, only that part of circuit consumes power, which performs some operation resulting in power savings. At the circuit level, DPM is performed by "gated clock" technique in which clock is modified in such a way that D-FFs receive clock only when transaction is required at its output whereas for other D-FFs clock is disabled [10–13].

In the circuit shown by Fig. 3, if D_{input} and output are different, then output of XOR gate becomes high, allowing clock to pass through AND gate. At the clock edge, D-FF updates its output state. When D_{input} and output of FFs are similar, XOR gate output is low, thus disabling clock for D FFs.

5 Proposed Work

A modified LFSR is implemented in which power dissipation is found to decrease by disabling the clock signal for those FFs in multiple outputs LFSR, whose output signal is same as their input signal. Modified clock signal is provided only to those FFs, which require changing their output state, which result in decrement in the number of FFs with active clock pulse. Because of disabled FFs, reduction in power consumption is achieved. In Fig. 4 circuit clock gating technique is applied to all the FFs. This required N number of each of XOR and AND gates to implement the clock gating part of the circuit. Since at a particular time only K_1 number of XOR and same number of AND gates are operating in clock gating circuits, a new circuit is proposed with less number of XOR gates (K_1 instead of N) along with AND gates replaced by switches resulting in hardware reduction.
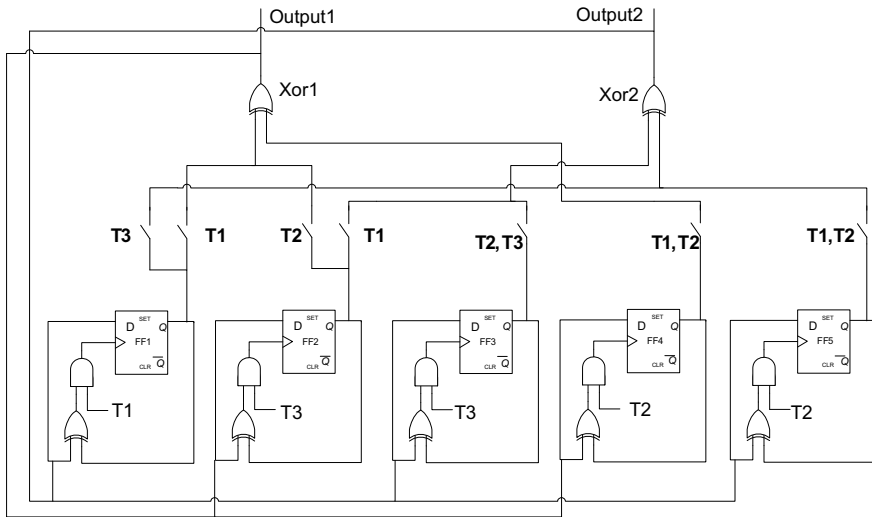
Fig. 4 Circuit diagram of clock gated multiple output low power parallel LFSR

5.1 Hardware Comparison

Hardware requirement of the proposed circuit is compared with clock gated serial LFSR [4] and multiple output parallel LFSR [1]. As proposed circuit is able to produce N outputs in $[N/K_1]$ clock cycles, $[N/K_1]$ phase generators are required better than [2] and similar to [1]. To implement a $[N/K_1]$ phase generator $\frac{1}{2}[N/K_1]$ FFs and $[N/K_1]$ 2-input NAND gates are required. The proposed circuit requires $[N/K_1]$ OR gates for control signal and average number of inputs for OR gates is $(M/D_{av})[N/K_1]$, where D_{av} is given by

$$D_{av} = (1/K_1) \sum_{i=1}^{K_1} D_i \tag{3}$$

where N is the number of FFs in LFSR, M is number of tap position, D_{av} is the average number of distinct inputs, that are input to XOR gate in clock cycles, and D_i is distinct inputs, that are input to XOR gate in ith clock cycle (Fig. 5).

The total number of switches in our design is $K_1 D_{av} + 2N$, out of which $K_1 D_{av}$ are used to connect FFs with XOR gate (because in proposed circuit K_1 XOR gate are connected with D_{av} number of switches) similar to [1] and an addition of 2 N switches due to designing of clock gating circuit (every FFs having two switches one at output for clock gating XOR gate and one for enabling gated clock to particular FFs). In the proposed circuit there are K_1, M input XOR gate similar to [1] which are used to generate the next bit. Additional K_1, 2-input XOR gates are required
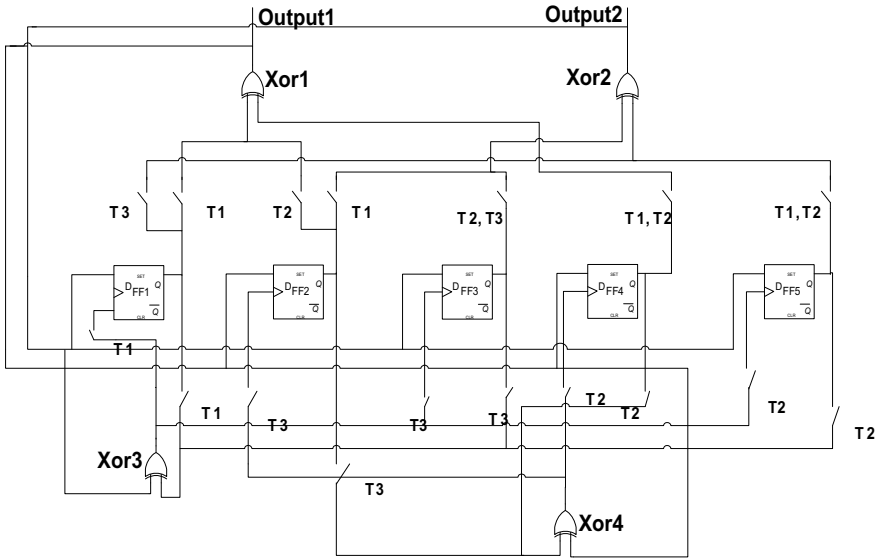
Fig. 5 Proposed parallel LFSR, number of transistors to implement clock gating technique are reduced
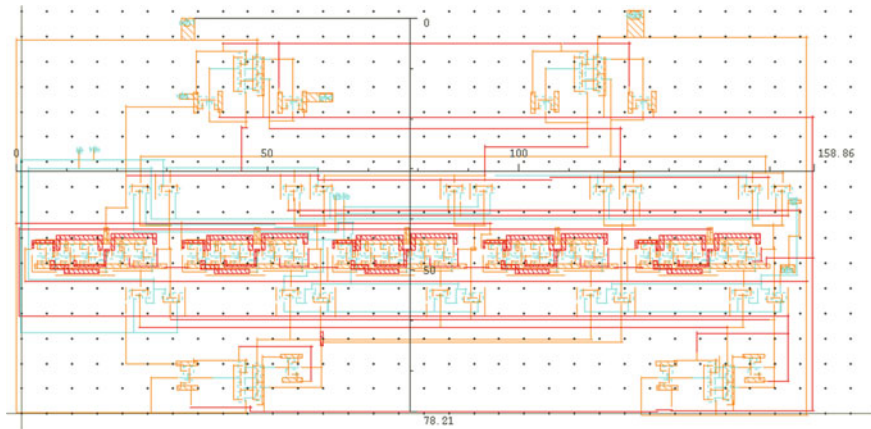
Fig. 6 Layout of proposed hardware minimized gated clock parallel LFSR

for implementing clock gating technique. Layout of the proposed circuit requires 12,424.4406 µm² area as shown in Fig. 6. Cadence Virtuoso tool is utilized while forming the layout of the proposed circuit.

5.2   Power Calculation

Dynamic power dissipation for proposed circuit is calculated as follows:

$$P = \frac{P_{PG}}{K_1} + P_{FF} + MP_{switch1} + P_{switch2} + P_{switch3} + MP_{XOR,M} + MP_{XOR,2} \quad (4)$$

where P_{PG} is power consumed by phase generator and control unit, which is same as [1]. Which includes Johnson counter and AND gates, P_{FF} is power consumed by LFSR FFs. P_{switch} is power consumed by an average switch, $P_{XOR,M}$ is power dissipated by M, input XOR gate, and $P_{XOR,2}$ is power consumed by 2, input XOR gate. $P_{switch1,}$ $P_{switch2,}$ and $P_{switch3}$ are the power consumption of switch1 (connection between D-FFs and M input XOR gate), switch2 (connection between clock gating XOR gate and D FFs clock input), and switch3 (connection between D FFs and clock gating XOR gate), respectively.

$$P_{FF}\left\{= \left(P_{FF_{CK}} + P_{FF_D}\right)\right\}$$
$$= \alpha C_{CK} V_{dd}^2 f + \alpha\left[C_D + C_{switch}\left\{\frac{K_1 D_{av} + N}{N}\right\}\right]V_{dd}^2 f \quad (5)$$

$$P_{XOR,M} = \alpha\left\{C_{XOR,M} + C_{in,XOR,2} + \frac{N}{K_1}C_{in,DFF}\right\}V_{DS}^2 f \quad (6)$$

$$P_{XOR,2} = \alpha\left\{C_{XOR,2} + C_{in,switch}\right\}V_{DS}^2 f \quad (7)$$

$$P_{switch1} = \alpha\left\{C_{switch1} + C_{in,XOR,M}\right\}V_{DS}^2 f \quad (8)$$

$$P_{switch2} = \alpha\left\{C_{switch2} + C_{in,D\_FF}\right\}V_{DS}^2 f \quad (9)$$

$$P_{switch3} = \alpha\left\{C_{switch3} + C_{in,XOR,2}\right\}V_{DS}^2 f \quad (10)$$

At every clock pulse, there are K_1 FFs which update its state but only a few are required to change its state that's why clock gating method is able to reduce power consumed by multiple output parallel LFSR [1].

5.3   Simulation and Results

Simulation has been done in Virtuoso simulator using SCL 180 nm technology file from cadence. As per Table 1 we obtain 6.76 and 53.45% power reduction in comparison to Katti's LFSR [1] and gated clock serial LFSR method [4]. Post

Table 1 Comparison of power dissipation for different PRNG technique for the polynomial $1 + x^2 + x^5$

PRNG method	Power consumption (μw)
Katti's LFSR [1]	1.034
Gated clock serial LFSR [4]	2.071
Gated clock parallel LFSR	1.375
Gated clock hardware minimized parallel LFSR	0.964

Table 2 Effect of RC-extraction in delay for proposed LFSR

Delay (ps)	Without RC-extraction	With RC-extraction
t_{PLH}	379.2	446.3
t_{PHL}	9.457	209.7
Average delay	194.3285	328

layout simulation of for hardware minimized gated clock parallel LFSR is carried using calibre. Effect of RC-extraction in rise time and fall time for given circuit is calculated. By Table 2, it can be observed that due to RC-extraction average delay of circuit is increased by 68.78%.

6 Conclusion

The proposed circuit disables the clock signals for those FF's in multiple output LFSR whose output signal is identical as their input signal. The proposed circuit is contrasted with gated clock serial LFSR and Katti's method, and significant reduction of power (i.e., reduction of 6.76% and 53.45% for Katti's and gated clock serial LFSR method, respectively) is obtained at the cost of increased area. From layout, it is observed that proposed circuit required 12,424.4406 μm^2 area. Moreover, due to RC-extraction average delay of circuit is increased by 68.78%.

References

1. Katti, R.S., Ruan, X., Khattri, H.: Multiple-output low-power linear feedback shift register design. IEEE Trans. Circuits Syst. I. **53**(7), 1487–1495 (2006)
2. Lowy, M.: Parallel implementation of linear feedback shift registers for low power applications. IEEE Trans. Circuits Syst. II. **43**(6), 458–466 (1996)
3. Hamid, M.E., Chen, C.I.H.: A note to low-power linear feedback shift registers. IEEE Trans. Circuits Syst. II. **45**(9), 1304–1307 (1998)
4. Aloisi, W., Mita, R.: Gated-clock design of linear-feedback shift registers. IEEE Trans. Circuits Syst. II. **55**(6), 546–550 (2008)

5. Koeter, J.: What's an LFSR? Texas Instruments Incorporated. http://focus.ti.com/lit/an/scta036a/scta036a.pdf (1996)
6. Praveen, J., Shanmukhaswamy, M.N.: Power reduction technique in LFSR using modified control logic for VLSI circuit. In: International Conference on Electronic Design and Signal Processing (ICEDSP), Manipal (2012)
7. Karighattam, V., Godbole, D.: High Speed and Low Power Hashing System and Method. U.S. Patent 9,960,909 (2018)
8. Baskar, S., Dhulipala, V.R.: Biomedical rehabilitation: data error detection and correction using two dimensional linear feedback shift register based cyclic redundancy check. J. Med. Imaging Health Inform. **8**(4), 805–808 (2018)
9. Shaer, L., Sakakini, T., Rouwaida, K.: A low power reconfigurable LFSR. In: 18th Mediterranean Electrotechnical Conference (MELECON), Lemesos (2016)
10. Kitsos, P., Sklavos, N., Zervas, N., Koufopavlou, O.: A reconfigurable linear feedback shift register (LFSR) for the bluetooth system. In: Proceedings of IEEE International Conference Electronics, Circuits and Systems (ICECS'01), Valletta (2001)
11. Security Requirements for Cryptographic Modules U.S. Department of Commerce, National Institute of Standard and Technology, FIPS PUB 140–142 (2001)
12. Lang, T., Musoll, E., Cortadella, J.: Individual flip-flops with gated clocks for low power datapaths. IEEE Trans. Circuits Syst. II. **44**(6), 507–516 (1997)
13. Palumbo, G., Pappalardo, F., Sannella, S.: Evaluation on power reduction applying gated clock approach. In: Proceedings of IEEE ISCAS 2002, pp. 85–88 (2002)

A Novel Dual Material Extra Insulator Layer Fin Field Effect Transistor for High-Performance Nanoscale Applications

Prateek Kishor Verma (ID)**, Varun Mishra, Yogesh Kumar Verma, Pawan K. Yadav and Santosh Kumar Gupta** (ID)

Abstract Present work introduces a Dual Material Gate Extra Insulator Layer (DMG EIL) FinFET for high-performance nanoscale applications. The Dual Material Gate screens the minimum channel potential from the drain voltage variations which reduce the short channel effects (SCEs). The EIL (high-k dielectric) injected in silicon active layer reduces the critical electric field near the drain region which reduces the hot carrier effects. The present device also offers low standby mode current useful for low power applications along with reduced SCEs and hot carrier effects (HCEs). The performance of the proposed device is compared with EIL-FinFET by investigating the output characteristics, electric field, carrier temperature and surface potential. The DMG EIL-FinFET is found to perform better compared to both—conventional FinFET and EIL-FinFET.

Keywords Silicon-on-insulator (SOI) · Dual metal gate (DMG) · FinFET · Extra insulator layer (EIL) · High-k · Hot carrier effects · Short channel effects

P. K. Verma (✉) · V. Mishra · Y. K. Verma · P. K. Yadav · S. K. Gupta (✉)
Department of Electronics and Communication Engineering, Motilal Nehru National
Institute of Technology Allahabad, Allahabad 211004, India
e-mail: prateek.kishor05@gmail.com

S. K. Gupta
e-mail: skg@mnnit.ac.in

V. Mishra
e-mail: varun20mishra@gmail.com

Y. K. Verma
e-mail: yogeshvermaec@gmail.com

P. K. Yadav
e-mail: pky.2494@gmail.com

1 Introduction

With the scaling of the metal oxide semiconductor field-effect transistors (MOSFETs) [1–4] to nanometer regime, gate control over the channel is shared by the drain and many undesirable effects such as short channel effects (SCEs) and hot carrier effects (HCEs) which become prominent. To overcome these problems, silicon-on-insulator (SOI) MOSFETs were introduced which offered many advantages over the conventional MOSFETs in terms of improved SCEs, low parasitic capacitance and higher device speed. Further, reduction in SCEs have been reported with the introduction of multiple-gate structures such as double gate (DG), fin shaped (FinFET), gate all around (GAA), etc. The multiple-gate structures offer better control over the channel region and higher transconductance [5–9]. Despite these advantages, the multiple-gate structures suffer higher average lateral electric field compared to the single-gate structures. Because of this, the effect of impact ionization near the drain region is significant at higher drain voltages leading to serious HCEs issues. To overcome the effect of HCEs, Fin Field Effect Transistor with extra insulator layer (EIL-FinFET) for high-performance nanoscale applications has been introduced [9]. In this structure, an extra insulator layer (EIL) is injected into silicon active layer. Use of EIL controls the HCE by reducing the critical electric field near the drain region. The EIL (high-k dielectric, here HfO_2) is located between the silicon active layer and drain region under the gate oxide, where the average lateral electric field in the post-saturation region is high due to the function of the gate. This insulator region (HfO_2) decreases the electric field in the channel and drain regions, especially near the Fin corners.

In this work a novel nanoscale SOI-FinFET with improved performance on HCEs and SCEs is proposed by introducing Dual Material Gate (DMG) in EIL-FinFET. The Dual Material in the gate introduces a potential step which screens the effects of drain voltage variations, lowering the SCEs along with reducing the critical electric field in the channel due to the use of EIL. This also brings about more efficiency in the operation of the proposed structure compared to EIL-FinFET.

2 Device Structure

The proposed device structure called DMG EIL-FinFET and reported EIL-FinFET are shown in Fig. 1a and b, respectively. The proposed device is gate engineered EIL-FinFET obtained by introducing a step in surface potential with the use of two metals of different work functions W1 and W2 together. A high-k dielectric material (HfO_2) has been used as EIL to reduce the surface electric field as shown by Eq. (1).

$$E_{ox} = E_{Si} \cdot \varepsilon_{Si}/\varepsilon_{ox} \tag{1}$$

Fig. 1 **a** Cross-sectional view of FinFET structure with extra Insulator Layer (EIL-FinFET) [9] and, **b** proposed dual material gate (DMG) EIL-FinFET
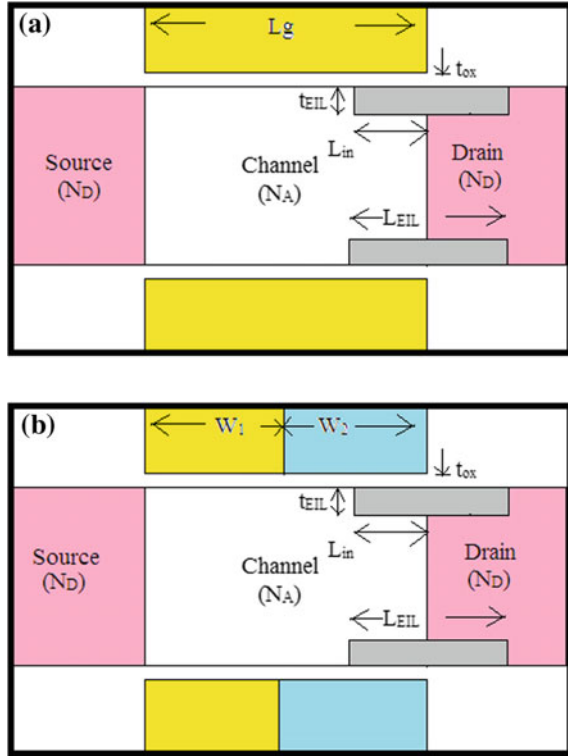
where E_{ox} and E_{Si} are the oxide and channel electric field, ε_{ox} and ε_{Si} are the dielectric constants of insulator and silicon. The use of lower work function gate material on the drain side further assists to reduce the surface electric field and also reduces the short channel effects effectively isolating the source-channel barrier from the drain voltage variations.

The device structures are implemented in SILVACO TCAD (ATLAS) simulator [10]. All other characteristics of EIL-FinFET have been verified with the published data by Orouji et al. [9]. The different device design parameters are taken as shown in Table 1. The simulations are carried out by activating the Lombardi (CVT) model [11] which is a complete model including the effects of N, T, $E_{\parallel}$ and $E_{\perp}$ and best suited for the non-planar devices. The semiconductor device simulations based on thermal equilibrium approximation are not enough for extremely scaled devices and conventional drift-diffusion models are not sufficient because the carrier temperature can be much higher than the ambient temperature. So, energy-balanced equation needs to be used for accurate prediction of the short channel characteristics. Shockley–Read–Hall (SRH) model [12, 13] clubbed with concentration-dependent mobility and Auger models [14] has been used to account for the carrier generation

Table 1 Design parameters of EIL-FinFET and proposed DMG EIL-FinFET

Parameter	Symbol	EIL-FinFET	DMG EIL-FinFET
Gate length	$L_g = (L_1 + L_2)$	30 nm	30 nm $L_1 = L_2 = 15$ nm
Gate oxide thickness (SiO$_2$)	t_{ox}	1.2 nm	1.2 nm
Buried oxide thickness (SiO$_2$)	t_{BOX}	40 nm	40 nm
Active silicon thickness	H_{fin}	30 nm	30 nm
Active silicon width	W_{fin}	10 nm	10 nm
Thickness of the extra insulator layer (HfO$_2$)	t_{EIL}	1.5 nm	1.5 nm
Extra insulator layer length (HfO$_2$)	L_{EIL}	18 nm	18 nm
Extra insulator layer length in the channel side HfO$_2$)	L_{in}	4 nm	4 nm
Channel doping concentration	N_A	10^{15} cm^{-3}	10^{15} cm^{-3}
Source/drain doping concentration	N_D	10^{20} cm^{-3}	10^{20} cm^{-3}
Gate work function	W_1	4.6 eV	4.6 eV
	W_2	–	4.4 eV

and recombination. Further, the quantum mechanical transport simulation is performed by solving a set of 3D density equations coupled with Poisson equation and electron-hole current continuity equations.

3   Results and Discussions

In the present section, the proposed device is compared with the EIL-FinFET on the basis of different performance parameters such as surface potential, critical electric field, etc. Figure 2 shows the output characteristics of DMG EIL-FinFET and EIL-FinFET at $V_{GS} = 0.75$ V with the device parameters given in Table 1 and here in Fig. 2, the thickness of the extra insulator layer (t_{EIL}) is taken as 1.5 nm. The drain current of the proposed device can be observed to be higher compared to the EIL-FinFET. Figure 3 shows the output characteristic for $L_{in} = 2$ nm at $V_{GS} = 0.75$ V and other device parameters as given in Table 1. In this case also, it can be seen that the proposed device offers higher drain current compared to EIL-FinFET.

It is very important to look at the surface electric field to get insight about the hot carrier effects (HCEs) in a device. A higher electric field at the surface leads to high energetic electrons and holes known as hot carriers, which may lead to reliability issues in a device. These hot carriers may regenerate hot carriers with impact ionization, which may leave the thin channel region to enter into the buried or gate oxide to create trap charges. This effect may worsen with elapsed time and
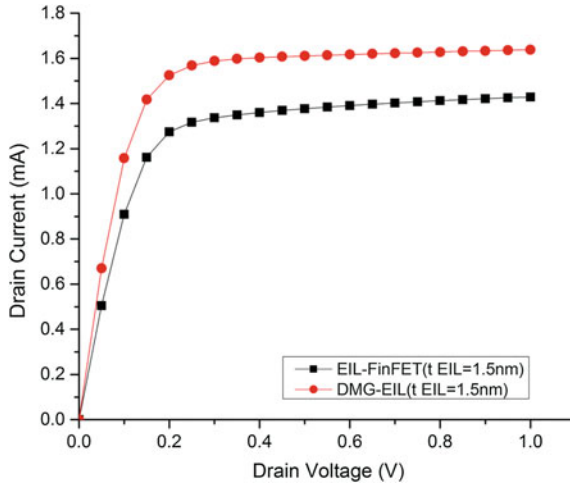
Fig. 2 Output characteristics at $V_{GS} = 0.75$ V for $t_{EIL} = 1.5$ nm
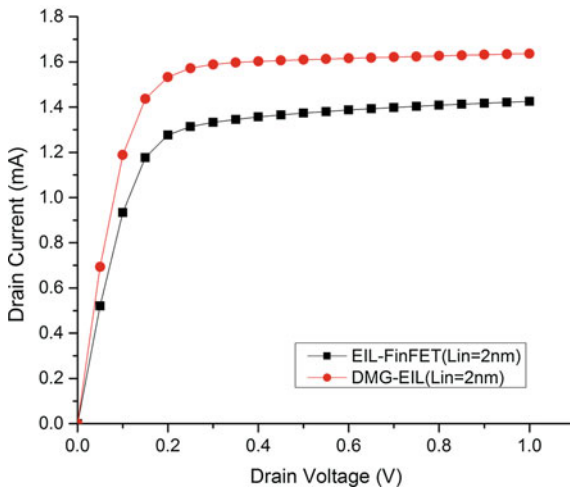
Fig. 3 Output characteristics at $V_{GS} = 0.75$ V for $L_{in} = 2$ nm

degrade the device performance by changing the threshold voltage of the device initially and may leave the device permanently damaged.

The graphical representation in Fig. 4 shows the surface electric field along the channel for the proposed device and EIL-FinFET for device parameters given in Table 1. The surface electric field of the proposed device near the drain end can be seen to be lower compared to EIL-FinFET reducing the HCEs. Figure 5 represents the surface electric field for $t_{EIL} = 3$ nm with other device parameters same as given in Table 1. Increase in thickness of EIL further reduces the surface electric field and

Fig. 4 Surface electric field along the channel at $V_{GS} = 0.15$ V and $V_{DS} = 0.75$ V
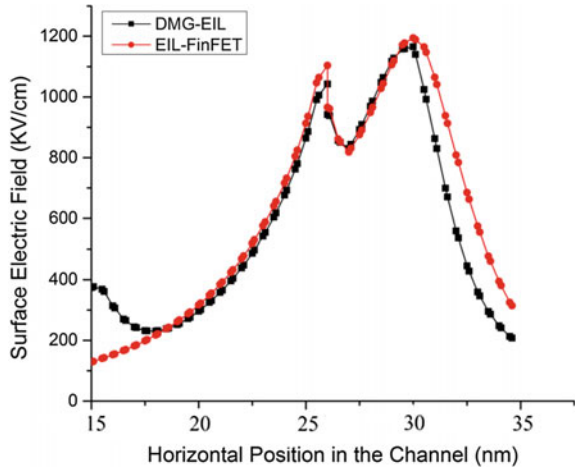
Fig. 5 Surface electric field along the channel with t_{EIL} = 3 nm at $V_{GS} = 0.15$ V and $V_{DS} = 0.75$ V
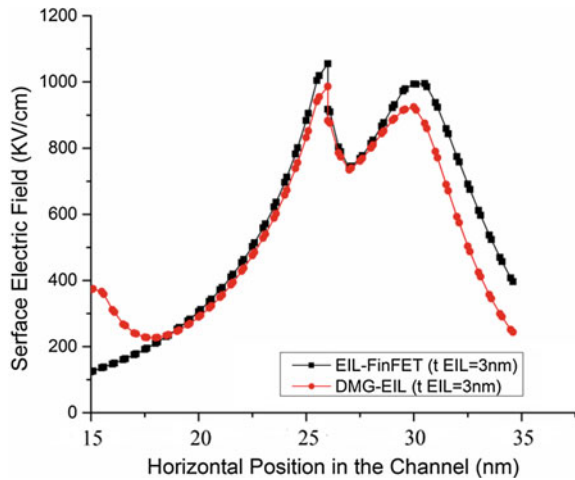
hence the HCEs. Figure 6 shows the surface electric field along the channel for L_{in} = 4 nm. In this case also, the surface electric field is lower compared to that of the EIL-FinFET. In summary, the use of DMG in EIL-FinFET helps to further reduce the HCEs.

The electron temperature has been shown in Fig. 7 for the proposed and EIL-FinFET at $V_{GS} = 0.05$ V and $V_{DS} = 0.75$ V for the device parameter given in Table 1 when the lattice temperature is 300 K. The electric field in the channel accelerates the electrons to very high speeds increasing the electron energy, which is equivalent to an increase in the equivalent temperature of the electrons. Better device characteristics such as reduced HCEs can be obtained for lower electron temperatures.

Fig. 6 Surface electric field along the channel with $L_{in} =$ 4 nm at $V_{GS} = 0.15$ V and $V_{DS} = 0.75$ V
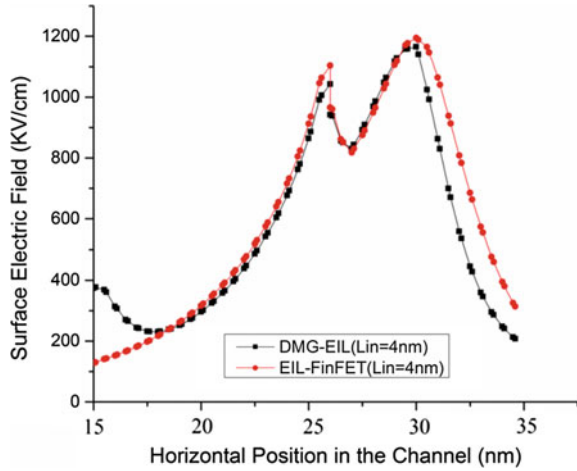
Fig. 7 Electron temperature profile along the surface of the channel at the centre of the FinFET structure at $V_{GS} = 0.05$ V and $V_{DS} = 0.75$ V
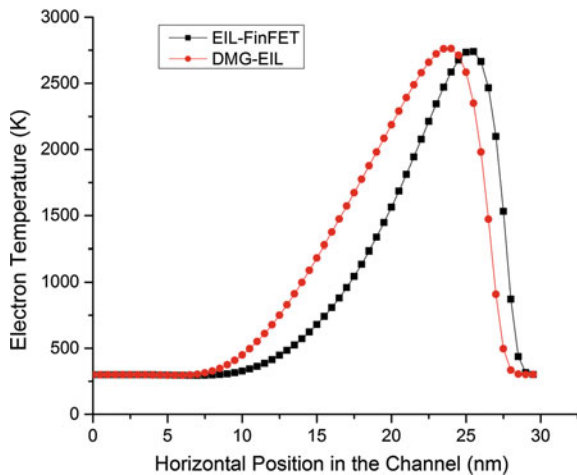
The graphical representation given in Fig. 8 shows the surface potential along the horizontal position in the channel. The introduction of the DMG in the EIL-FinFET screens the source-channel barrier from drain voltage variations effectively reducing the drain-induced barrier lowering (DIBL), one of the short channel effects. The potential barrier of the proposed device is lower compared to that of the EIL-FinFET at higher drain voltage. Hence, fewer electrons will enter the channel region from source in comparison with the EIL-FinFET, which will further reduce the leakage current in the standby mode. Due to the use of screen gates, variation in minimum surface potential is less as compared to EIL-FinFET resulting in a lower change of potential barrier in the proposed structure and the variation of standby mode current will be low.

Fig. 8 Surface potential profile along the channel for the FinFET structures at bias voltages of $V_{DS} = 0.15$ V and 1.2 V with $V_{GS} = 0.15$ V
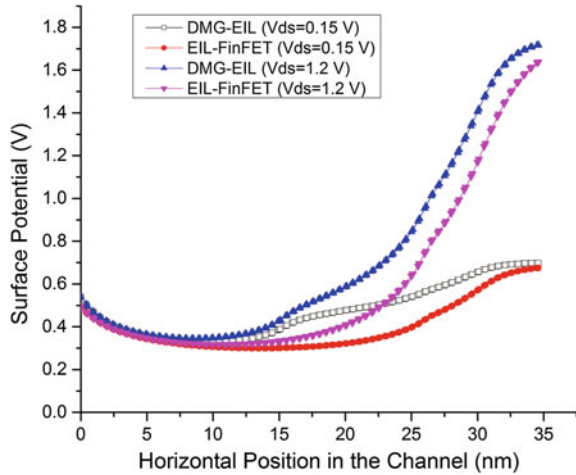
4 Conclusion

The proposed structure includes gate engineered EIL-FinFET by placing higher work function gate on the source side and lower work function gate on the drain side. The use of screen gates immunes the minimum surface potential variation in the channel region from the changes at the drain end. In addition to this, the high-k insulator material (HfO_2) between the channel and drain regions under the gate oxide, reduces the average lateral electric field. It also offers more drain current and lower electric field at the drain end. The proposed structure, DMG EIL-FinFET achieves the goal of reduced both, SCE and HCE. In addition to these advantages, the present device also offers high ON-state current and low standby mode current which are useful for low power applications.

References

1. Colinge, J.P.: Silicon-on-Insulator Technology: Materials to VLSI, 3rd edn. Kluwer Academic Publishers, USA (2004)
2. Kue, J.B., Lin, S.C.: Low-Voltage SOI CMOS VLSI Devices and Circuits. University of Waterloo NTU EE, Waterloo (2001)
3. Orouji, A.A., Kumar, M.J.: A new symmetrical double gate nanoscale MOSFET with asymmetrical side gates for electrically induced source/drain. Microelectron. Eng. **83**, 409–414 (2006)
4. Park, J.T., Colinge, J.P.: Multiple-gate SOI MOSFETs: device design guidelines. IEEE Trans. Electron Devices **49**(12), 2222–2229 (2002)
5. Colinge, J.P.: Multiple-gate SOI MOSFETs. Solid State Electron. **48**, 897–905 (2004)
6. Gupta, S.K., Baishya, S.: Analog and RF performance evaluation of dual metal double gate high-k stack (DMDG-HKS) MOSFETs. J. Nano- Electron. Phys. **5**(3), 03008 (2013)

7. Manchi, P.K., Gupta, S.K.: Corner effects in SOI tri-gate FINFET structure by using 3D process and device simulations. In: IEEE International Conference on Computer and Communication Technology (ICCCT 2010), MNNIT Allahabad, India, 17–19 September 2010, pp. 683–686 (2010)
8. Gupta, S.K., Baishya, S.: 3D-TCAD simulation study of an electrically induced source/drain cylindrically surrounding gate MOSFETs for reduced SCEs and HCEs. In: IEEE International Conference on Electronics and Computer Technology (ICECT) 2011, Kanyakumari, India, 8–10 April, pp. 429–432 (2011)
9. Orouji, A.A., Karimi, Fa.: A novel fin field effect transistor by extra insulator layer for high-performance nanoscale applications. J. Comput. Electron. **14**, 811–819 (2015)
10. Silvaco: ATLAS User's Guide (2015)
11. Lombardi, C., Manzini, S., Saporito, A., Vanzi, M.: A physically based mobility model for numerical simulation of nonplanar devices. IEEE Trans. Comput. Des. Integr. Circuits Syst. **7**, 1164–1171 (1988)
12. Shockley, W., Read, W.T.: Statistics of the recombination of holes and electrons. Phys. Rev. **87**, 835–842 (1952)
13. Hall, R.N.: Electron-hole recombination in germanium. Phys. Rev. **87**, 387 (1952)
14. Dziewior, J., Schmid, W.: Auger coefficients for highly doped and highly excited silicon Auger coefficients for highly doped and highly excited silicon. Appl. Phys. Lett. **31**, 346–348 (1977)

Performance of Double Gate Tunnel FET Devices with Source Pocket

Varun Mishra [ID]**, Yogesh Kumar Verma, Prateek Kishor Verma, Ningthoujam Qoonand Singh and Santosh Kumar Gupta**

Abstract It is known that due to several current injection mechanisms, tunnel FETs can attain a subthreshold swing less than 60 mV/dec at normal temperature, which makes it alternative for MOSFETs, particularly in low power applications. This work investigates the electrical and analog performance of dual metal double gate hetero dielectric source pocket TFETs in contrast to dual metal double gate high-k dielectric source pocket TFETs. Numerical calculations were carried out to find the characteristics of both the devices. It is observed that for hetero-dielectric structure, the OFF current decreases significantly while having the marginal effect on the ON state current in contrast to double gate high-k dielectric structure.

Keywords Dual metal double gate · Tunnel FET · Dielectric · HfO_2

1 Introduction

Tunnel FET (TFET) is also generally known as gated PIN diode due to its band-to-band tunneling mechanism (BTBT) [1] and its ability to operate at a very low supply voltage. TFET is a suitable contender to outperform MOSFET, as it provides a low OFF current, which helps in overcoming the Boltzmann tyranny (i.e., achievable subthreshold swing is below 60 mV/dec). The subthreshold swing of conventional

V. Mishra (✉) · Y. K. Verma · P. K. Verma · N. Q. Singh · S. K. Gupta
Motilal Nehru National Institute of Technology, Allahabad, Allahabad, UP, India
e-mail: rel1605@mnnit.ac.in

Y. K. Verma
e-mail: rel1607@mnnit.ac.in

P. K. Verma
e-mail: rel1513@mnnit.ac.in

N. Q. Singh
e-mail: kunan125@gmail.com

S. K. Gupta
e-mail: skg@mnnit.ac.in

MOSFETs is limited to 60 mV/dec that makes the device slower but TFETs can bring this subthreshold swing less than 60 mV/dec that makes it faster than MOSFETs [2]. At the same time, drain-induced barrier lowering (DIBL) and short channel effects (SCEs) are very low compared to MOSFETs. Formerly published reports suggest that utilizing the concept of dual metal double gate technology and high-k dielectric engineering in TFETs provides a device with improved characteristics [2, 3]. Recently TFETs with dual material double gate and with the different dielectric material (either SiO_2 or high-k) have been analyzed [2–4]. The leakage current due to very small thickness of the gate oxide can be prevented by using high-k material in TFETs [5–8]. The gate capacitance (C_{gg}) and gate-drain capacitance (C_{gd}) can also be improved by using different permittivities (k) oxide material such as a high-k oxide material on the source side and a low-k oxide material on the drain side [9, 10]. Figure 1a shows the double gate tunnel FET structure with hetero dielectric layers. Figure 1b presents the dual gate tunnel FET with high-k dielectric. Figure 1c presents the variation of energy band diagram for conventional TFET in ON state and OFF states. For ON state, when gate voltage is applied the energy band between the conduction band and valence band becomes narrow enough so that the carriers can tunnel across it.
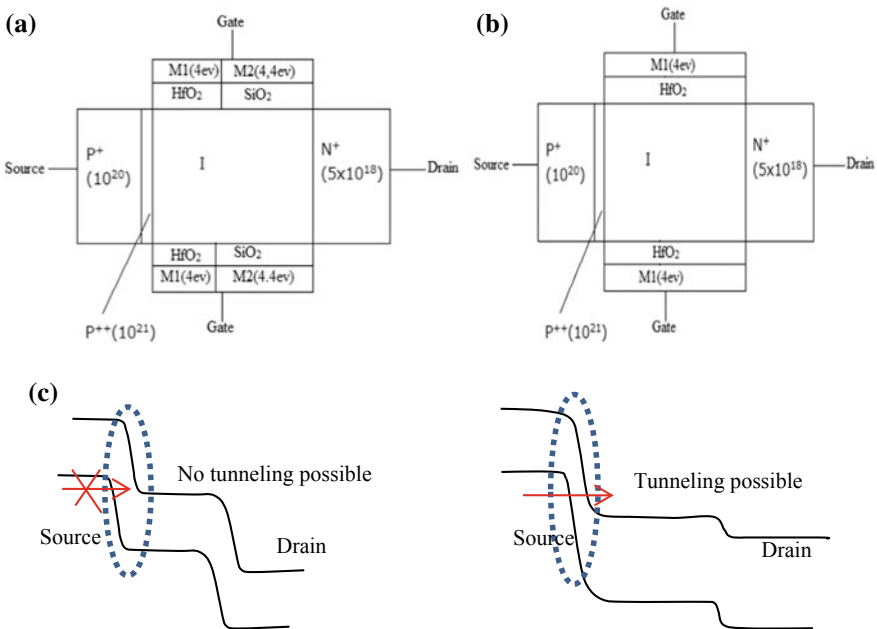
Fig. 1 **a** Dual metal double gate tunnel FET with pocket doped on source side with hetero dielectric, **b** dual gate tunnel FET with pocket doped on source side with high-k dielectric, **c** energy bands for OFF and ON states of tunnel FET
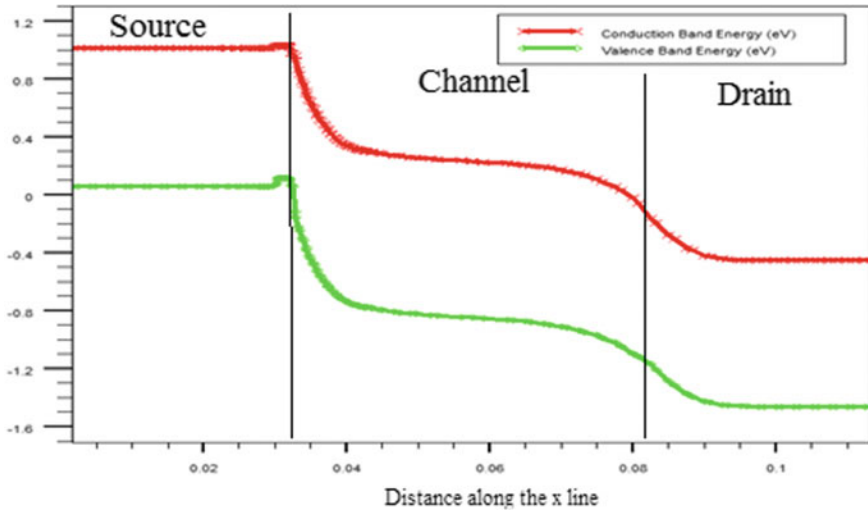
Fig. 2 Energy band diagram of dual metal gate hetero dielectric source pocket TFET with $V_{GS} = 0.5$ V

2 Device Design and Simulation

In the design process, the source is doped with 1×10^{20} cm^{-3} p^{++} type and drained with 5×10^{18} cm^{-3} n^{++} type and is intrinsic with 1×10^{15} cm^{-3} p-type and source pocket with p^{++} 1×10^{21} cm^{-3} of thickness 2 nm. Figure 2 represents the energy band diagram of dual metal double gate hetero dielectric source pocket with gate metal work function 4 eV near the source side and 4.4 eV near drain side [5, 6]. Figure 3 represents the energy band diagram of dual metal double gate high k dielectric source pocket. The simulation is done with a 2-D Silvaco ATLAS device simulator. The total channel length is 50 nm and the oxide thickness is 3 nm. In TCAD simulations, we use Shockley–Read–Hall (SRH) recombination, bandgap narrowing (BGN), concentration-dependent mobility, Auger recombination, standard band-to-band tunneling, and Lombardi (CVT) models. All the simulation is done using default parameters. The doping profiles at the junctions are nearly abrupt for all simulations.

3 Simulation Result and Discussion

The basic idea of introducing an asymmetric work function (Al = 4.0, Tungsten = 4.4) is to decrease the subthreshold slope and leakage current (I$_{OFF}$). The dual metal at the gate, along with different dielectrics provide additional flexibility to control various sections of the transfer characteristics, thus improving the TFETs performance. At OFF state, as the work function of the gate near drain increases, it
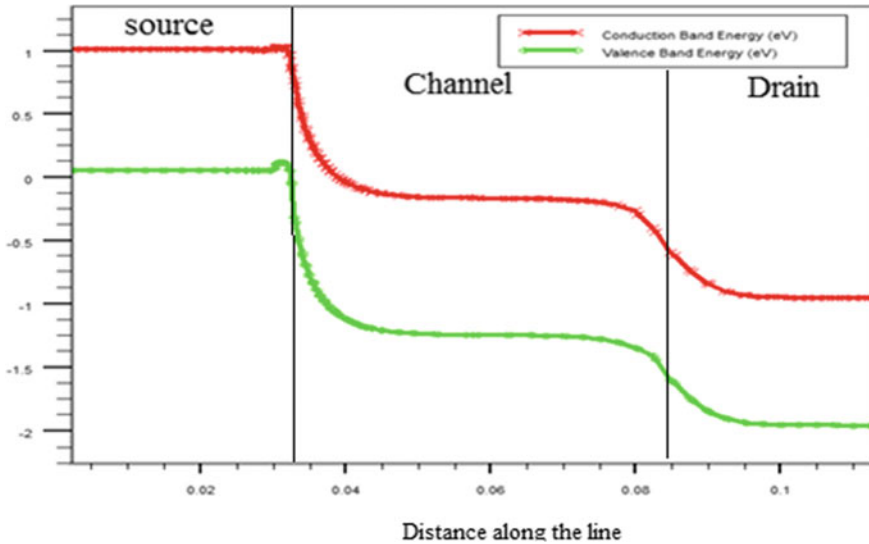
Fig. 3 Energy band diagram of dual metal gate high-k dielectric source pocket TFET with $V_{GS} =$ 0.5 V

leads to enhancement in tunneling width, which consequently suppresses the band overlap on the source side, providing the reduced OFF-state current. The OFF-state current is composed of reverse leakage current in association with the BTBT current. At ON state, enhancement in work function near the drain side does not bring any significant change in the tunnel barrier width. It can be concluded that for ON state, as the gate work function near the source side increases the current starts to decrease because of enhanced tunnel barrier width at the source/body junction. At OFF state, tunnel gate work function does not bring any significant change in tunnel barrier width.

It can be observed from Fig. 4 that the ON state current of the source pocket dual gate high-k dielectric TFET is insignificantly higher than the source pocket dual metal gate hetero dielectric TFET, but the OFF-state current of the source pocket dual metal gate hetero dielectric TFET is considerably lower and hence the I_{ON}/I_{OFF} ratio of the source pocket dual metal gate hetero dielectric TFET is very much larger than the source pocket dual gate high-k dielectric TFET. The energy band diagram presented in Figs. 2 and 3 clearly states that at lower V_{GS} for dual gate high-k dielectric TFET, significant bending will surely increase the OFF-state current. It is observed that the subthreshold slope of the source pocket dual metal gate hetero dielectric TFET is less than the source pocket dual gate high-k dielectric TFET, because for lower gate to source voltage, drive current is very low and when the device is ON the drain current increases quickly. On the drain side of source pocket dual metal gate hetero dielectric TFET, use of low-k material leads to lowering of the OFF-state current. Thus, I_{ON}/I_{OFF} is very high and the DIBL is low for source pocket dual metal gate

Fig. 4 Transfer characteristics of the proposed devices for $V_{DS} = 0.5$ V
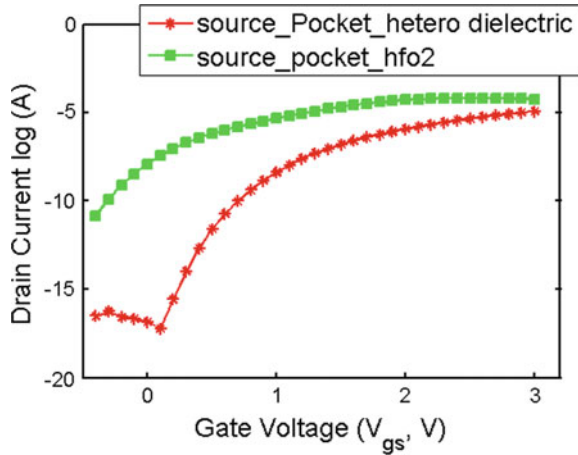
hetero-dielectric TFET as compared to dual metal high-k dielectric source pocket TFET.

Table 1 presents the numerically calculated value of DIBL for source pocket dual gate high-k dielectric TFET and source pocket dual metal gate hetero dielectric TFET. The work function of a gate electrode employed near the drain side is optimized at 4.4 eV; on increasing the work function above the optimized value, the threshold voltage increases. On increasing the drain to source bias, due to 4.4 eV gate electrode near the drain side, the barrier width does not change significantly. Hence, higher DIBL of high-k dielectric TFET makes it an inferior device in contrast to hetero dielectric TFETs.

Figure 5 shows that the threshold voltage is lesser in the case of source pocket dual metal gate hetero dielectric TFET in contrast to source pocket dual gate high-k dielectric TFET. V_{TH} variation with channel length is insignificant for source pocket dual material gate hetero dielectric TFET but presents a staggered variation in case of source pocket dual gate high-k dielectric TFET. Hence, the source pocket dual material gate hetero dielectric TFET is superior and provides better electrical characteristics over the previously discussed TFETs. Due to higher gate metal of 4.4 eV near the drain side on source pocket dual material gate hetero dielectric TFET, the threshold voltage is low.

The transconductance of a device is a very important parameter; it signifies the amplification property of a device so the value of transconductance should be as

Table 1 DIBL for two types of pocket source tunnel FET

Types of tunnel FET	Gate oxide used	Measured DIBL (mV/V)
Source pocket dual gate high-K dielectric TFET	HfO_2	120
Source pocket dual metal gate hetero dielectric TFET	HfO_2 and SiO_2	50

Fig. 5 Threshold voltage variation against channel length
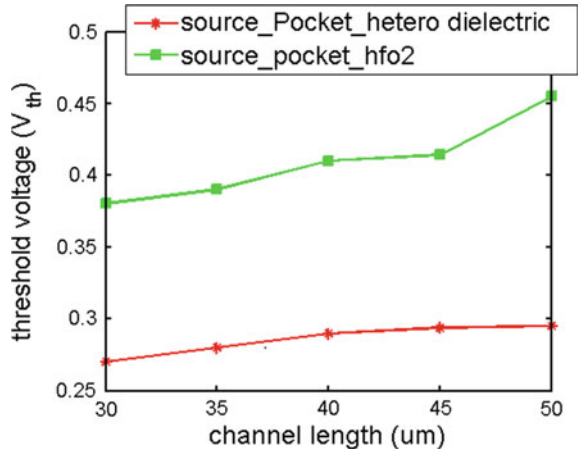
high as possible. Transconductance of a device is defined as the ratio of variation of drain current to the variation of gate voltage. Figure 6 presents the transconductance (g_m) variation with respect to the gate to source voltage (V_{GS}) for source pocket dual gate high-k dielectric TFET and source pocket dual material gate hetero dielectric TFET. Transconductance in the case of source pocket dual metal gate hetero dielectric TFET is more than source pocket dual gate high-k dielectric TFET, because of high-k dielectric on source side and low-k on drain side that enhances the electric field on the source side and consequently leads to increase in the tunneling process.

Figure 7 presents the drain conductance (G_d) variation with respect to V_{GS}. A little increase in drain conductance is noticed for dual gate high-k dielectric TFET in contrast to source pocket dual material gate hetero dielectric TFET. Because of high-k dielectric employed all over the channel, an increase in the electric field is observed

Fig. 6 Tranconductance (g_m) w.r.t. gate-to-source voltage for $V_{DS} = 0.5$ V
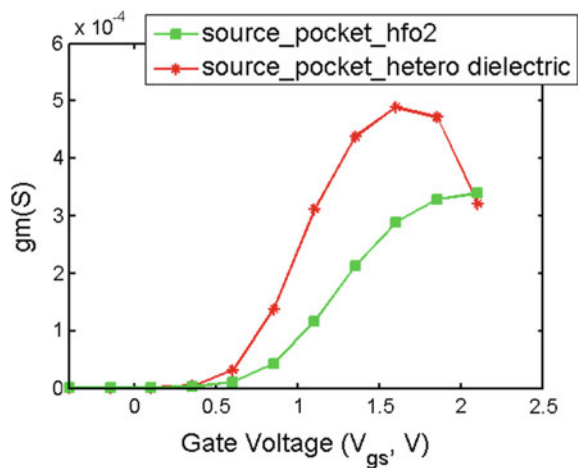
Fig. 7 Variation of drain
conductance against
gate-to-source voltage
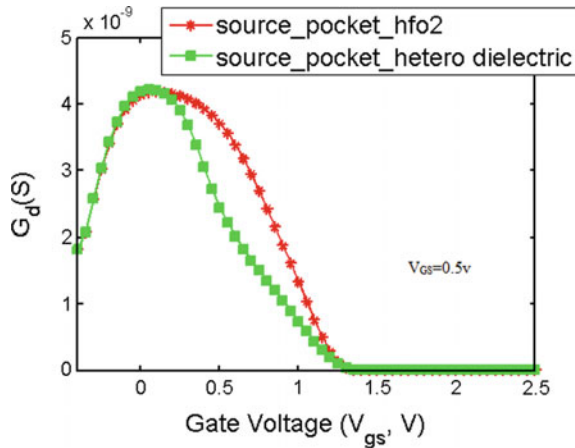
whereas for source pocket dual material gate hetero dielectric TFET, electric field
near drain–channel junction is lower than source–channel junction.

Gate-to-drain capacitance (C_{gd}) and gate-to-source capacitance (C_{gs}) variation
with respect to V_{GS} are presented in Figs. 8 and 9, respectively. It is observed from
Fig. 9 that C_{gs} reduces with V_{GS}, so the primary component that influences the
total gate capacitance (C_{gg}) is C_{gd}. Suppression of potential barrier at drain–channel
junction leads to increase in C_{gd} with V_{GS}. Control of gain over drain current is
increasing (i.e., I_{on}/I_{off} increases). Source pocket dual material gate hetero dielectric
TFET has better transconductance than source pocket dual gate high-k dielectric. It
can be seen from Fig. 9 that C_{gs} decrease as V_{GS} is increased. Here, C_{gs} of source
pocket dual gate high-k dielectric TFET is higher than source pocket dual metal gate
hetero dielectric TFET. Figure 10 represents the g_m/I_D variation w.r.t. gate–source

Fig. 8 Gate to drain
capacitance variation against
V_{GS} for $V_{DS} = 0.5$ V
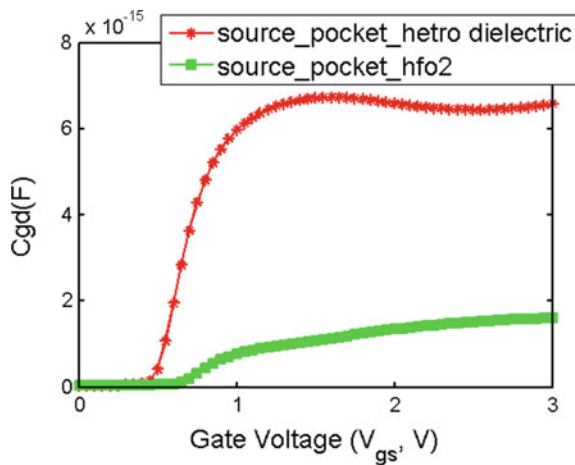
Fig. 9 Gate to source capacitance variation against V_{GS} for $V_{DS} = 0.5$ V
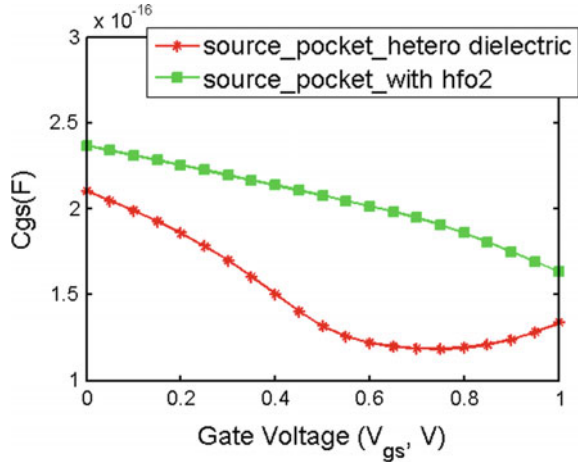
Fig. 10 Variation of g_m/I_D against V_{GS}
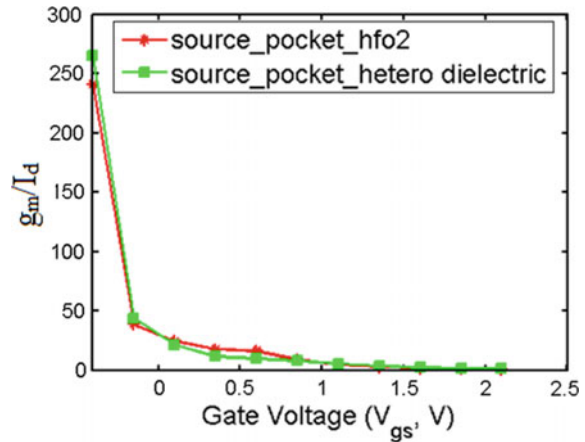
voltage. The device efficiency (g_m/I_D) is a parameter that presents how precisely the current is utilized in acquiring a certain value of transconductance. High device efficiency of a device shows that a lower bias is required in the realization of circuits. It is clear that both g_m and g_m/I_D are directly dependent on the drain current. The g_m/I_D of both the structures are almost the same; there are not many changes but there is slight improvement in the source pocket dual material gate hetero dielectric TFET.

4 Conclusion

It is observed that an additional pocket doped in the source side improves the performance of double gate TFETs. Work function engineering of gate metals varies the threshold voltage and it increases the ON current of the TFET, but the effect was significant when the work function of the gate near the source side is kept low. The device parameters such as work function of gate metals and channel length are optimized in order to provide better analog and electrical performance of the device. Source pocket doped hetero dielectric TFET has a lesser threshold voltage and lower OFF state current in contrast to pocket doped on source side with high-k dielectric TFET. A significant decrease in DIBL is observed for hetero dielectric TFET in contrast to high-k dielectric TFET. Transconductance shows better in the case of hetero dielectric TFET. The drain conductance (G_d) is higher for high-k dielectric TFET because of high-k dielectric employed all over the channel. C_{gs} and C_{gd} capacitances were higher for hetero dielectric TFETs in contrast to high-k dielectric TFET. Hence, source pocket hetero dielectric TFET proves to be a suitable candidate for future semiconductor technology.

References

1. Kao, K.-H., Verhulst, A.S., Vandenberghe, W.G., Sorée, B., Groeseneken, G., De Meyer, K.: Direct and indirect band-to-band tunneling in germanium-based TFETs. IEEE Trans. Electron Devices **59**(2), 292–301 (2012)
2. Boucart, K., Ionescu, A.M.: Double-gate tunnel FET with high-κ gate dielectric. IEEE Trans. Electron Devices **54**(7), 1725–1733 (2007)
3. Mishra, V., Verma, Y.K., Verma, P.K., Gupta, S.K.: EMA-based modeling of the surface potential and drain current of dual-material gate-all-around TFETs. J. Comput. Electron. **17**(4), 1596–1602 (2018)
4. Choi, W.Y., Park, B.-G., Lee, J.D., Lui, T.-J.K.: Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Device Lett. **28**(8), 743–745 (2007)
5. Anghel, C., Chilagani, P., Amara, A., Vladimirescu, A.: Tunnel field effect transistor with increased ON current, low-k spacer and high-k dielectric. Appl. Phys. Lett. **96**(12), 122104 (2010)
6. Gupta, S.K., Kumar, S.: Analytical modeling of a triple material double gate TFET with hetero-dielectric gate stack. Silicon 1–15 (2018)
7. Chang, H.-Y., Adams, B., Chien, P.-Y., Li, J., Woo, J.C.S.: Improved subthreshold and output characteristics of source-pocket Si tunnel FET by the application of laser annealing. IEEE Trans. Electron Devices **60**(1), 92–96 (2013)
8. Jhaveri, R., Girish, N.V., Woo, J.C.S.: Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor. IEEE Trans. Electron Devices **58**(1), 80–86 (2011)
9. Jagdesh Kumar, M., Saurabh, S.: Novel attributes of dual material gate nanoscale tunnel field-effect transistor. IEEE Trans. Electron Devices **58**(2), 404–410 (2011)
10. Gupta, S.K., Rawat, A.S., Verma, Y.K., Mishra, V.: Linearity distortion analysis of junctionless quadruple gate MOSFETs for analog applications. Silicon 1–9 (2018)

The Parameters Affecting Graphene Conductivity for Sensor and High-Frequency Application

Sharad Kumar Yadav and Richa Singh

Abstract In this paper, we simulated the ac-conductivity equation of graphene and discussed various parameter effects on graphene conductivity in gigahertz and terahertz region for sensing and high-frequency application. We discussed the adsorb charge mechanism which creates the n-type and p-type doping in intrinsic graphene and shifts the Fermi level away from the Dirac point. We increase the carrier concentration from $1.3 * 10^{13}$ cm^{-2} to $2.2 * 10^{13}$ cm^{-2} which increased the real part of conductivity from 0.049 to 0.064 S, similarly when the relaxation time varied from 1 to 1.2 ps the real part conductivity increases from 0.049 to 0.059 S (till gigahertz frequency); this alteration is the key for sensing application. The effect of small temperature variation shows a non-distensible change in conductivity with the simulated equation although its effect and application are discussed in detail. The calculation shows dispersionless behavior of the conductivity till the gigahertz frequency. In the terahertz region, the real and imaginary part of conductivity varies with frequency which offers dynamic control of tuning the antenna for high-speed data rate at such a higher frequency where the meta-based antenna suffers.

Keywords Graphene · Dirac point · Sensor · High-frequency conductivity

1 Introduction

The lightweight, flexible and in some cases optically transparent 2-D layered material-based devices are gaining attention in the current nanomaterial research arena [1]. There are a range of 2-D layered materials as graphene [2], borophene [3], phosphorene [4], germanene [5], silicene [6], various transition metal dichalcogenides [7], and several oxide materials [8]. These 2-D materials offer very attractive

S. K. Yadav (✉) · R. Singh (✉)
ECE Department, Motilal Nehru National Institute of Technology, Allahabad, Allahabad, UP, India
e-mail: sharadsws@gmail.com

R. Singh
e-mail: singhricha51@gmail.com

electronic and optoelectronic behaviors and depending on the number of layers and their order of stacking, the optical and electronic (bandgap) alteration is very significance and hence the same material could be a conductor or a semiconductor, optically transparent or semi-transparent or non-transparent [9].

Graphene is the first known 2-D layered van der Waals material. Single-layer graphene (SLG) has zero bandgap near the Dirac point (E_d) and shows linear dispersion away from E_d [10]. The linear dispersion of SLG leads to quasiparticle which is analogous to massless Dirac fermions and propagates with the Fermi velocity (V_F) which is about 1/300 of the light speed (C) [2], hence SLG offers higher carrier mobility even at room temperature. These properties open a range of applications for SLG-based devices as in sensors, in high-frequency devices (up to 100 GHz) [11], in terahertz plasmonic [12] and in tunable nanoantenna [13].

In the case of graphene sensor devices, the key parameter is the carrier concentration. As the sensing atoms/molecules land on graphene, they should change the doping profile of graphene and as a result, the Fermi level moves either in occupied (π) or unoccupied (π^*) state, i.e., away from the E_d as shown in Fig. 1. There are several doping methods reported to move the E_F of graphene from its E_d [14, 15]. The ability to adsorb higher band energy atoms/molecules on graphene sheet is an attractive technique to shift E_F away from zero density of state (DOS) although the amount of charge transfer on the sheet directly affects the sensing efficiency [16]. Example, the NH_3, which when land on graphene sheet, donate electrons (i.e., offer n-type doping) and the NO_2 adsorb the electrons and create hole (i.e., p-type doping) [16], and because of this, the E_F moves away from the E_d and hence the conductivity of graphene varies which is the key to develop graphene sheet for sensor application.

When it comes to the high-frequency application, the metal-based tunable antenna often suffers from enhanced grain boundary scattering. The graphene sheet overcomes this problem and offers frequency tunability in the terahertz range. In GHz frequency range, the metallic electrical loss increases with the frequency. Alshehri et al. [18] reported the experiment using coplanar waveguide (CPW) (millimeter
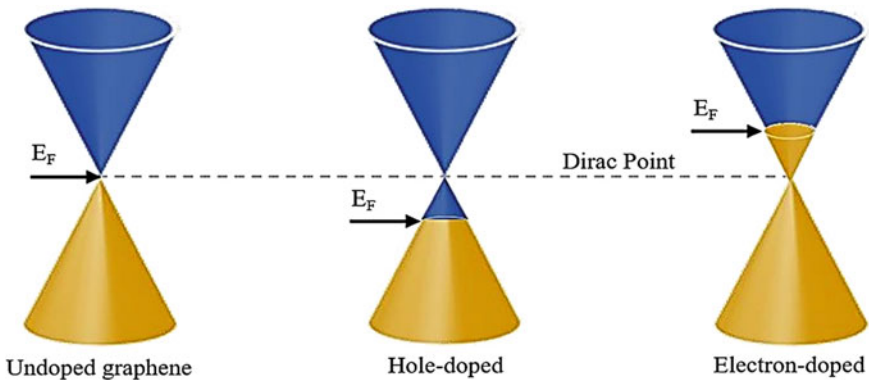
Fig. 1 Graphene bands in the presence of n-type and p-type dopants [17]

long) that 0.4 dB/mm insertion loss at 20 GHz increases by 0.1 dB/mm with each GHz increment in frequency. The reported metallic conduction is mainly constrained to thin surface layer because of the roughness loss and the skin effect where the roughness loss contributes most, which in case of graphene would not be significant [19]. Mihnev et al. [20] reported the effect of carrier relaxation (dynamics) in highly and lightly doped graphene. In the terahertz region, the carrier relaxation critically depends on the Fermi level and is primarily governed by the efficient interaction between carrier–carrier and carrier–phonon [20].

2   Computational Detail

The net carrier density of graphene is given as [9]

$$n - p = sgn(E_F)\frac{(E_F^2)}{(\pi(\hbar v_F)^2)} \tag{1}$$

where $(n - p)$ is the net carrier density (cm^{-2}) which is taken as 1.3×10^{13} cm^{-2}, the E_F is the Fermi energy level (eV), $\hbar$ is plank constant 1.05×10^{-34} J s^{-1} and V_F is the Fermi velocity (~10^6 m s^{-1}).

As there is a direct relation between the carrier concentration and Fermi level, the change in Fermi level will shift the Fermi level.

The conductivity equation as a function of frequency is [13]

$$\sigma_s(\omega) = \frac{(2ie^2 K_B T)}{\pi \hbar^2 (\omega + \frac{i}{\tau})} ln\left(2\cosh\left(\frac{E_F}{2K_B T}\right)\right)$$
$$+ \frac{e^2}{4\hbar}\left(\begin{array}{c} \frac{1}{2} + \frac{1}{\pi}tan^{-1}\left(\frac{\hbar\omega - 2E_F}{2K_B T}\right) \\ - \frac{i}{2\pi}ln\left(\frac{(\hbar\omega + 2E_F)^2}{(\hbar\omega - 2E_F)^2 + 4(K_B T)^2}\right) \end{array}\right) \tag{2}$$

where "σ" is the conductivity, "ω" is the frequency, "K$_B$" is the Boltzmann constant (1.38×10^{-23} J K^{-1}), "T" is the temperature (in K), "τ" is the relaxation time and "e" is the electron charge (1.6×10^{-19} C).

From the above two equations (Eqs. 1 and 2), it is clear that the carrier concentration, temperature, and the relaxation time are the three key parameters on which the graphene conductivity depends.

Equation (2) is written in two parts, part one in red color and part two in black. Part one of this equation accounts for the intraband conduction and part two for the interband conduction. When the Fermi energy level is less than half of the photon energy, part two (interband transition) of Eq. (2) dominates. As the Fermi level crosses half of the photon energy, the interband transition becomes negligible due to the Pauli-blocked state. When the Fermi energy level goes higher, the photon energy—part one

(intraband conduction) of Eq. (2)—dominates. At higher frequency, the gap between the Fermi energy level and the Dirac level becomes higher and only those photons which have higher energy ($>2\ E_F$) will be adsorbed and contribute in the conductivity [13].

3   Result and Discussion

The effect of three parameters (carrier concentration, temperature, and relaxation time) is shown in Figs. 2 and 3. In the low-frequency region (<100 GHz), the conductivity shows a dispersion-less behavior. As the carrier concentration increases
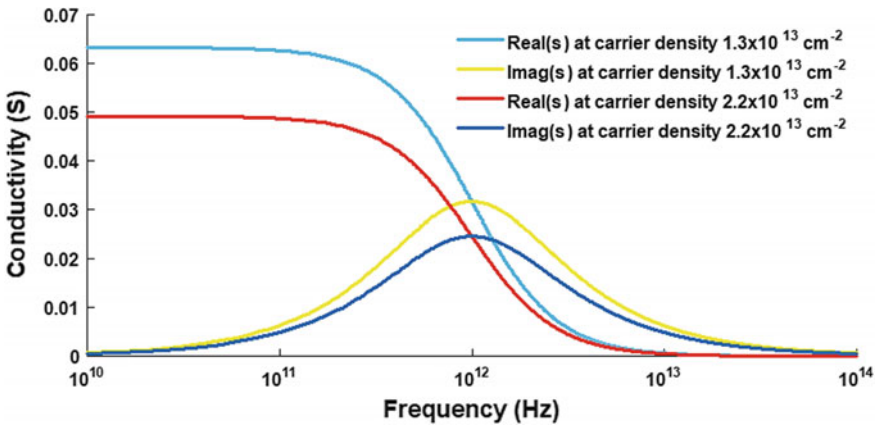
Fig. 2   The change in ac-conductivity with doping concentration 1.3×10^{13} cm^{-2} and 2.2×10^{13} cm^{-2} of graphene
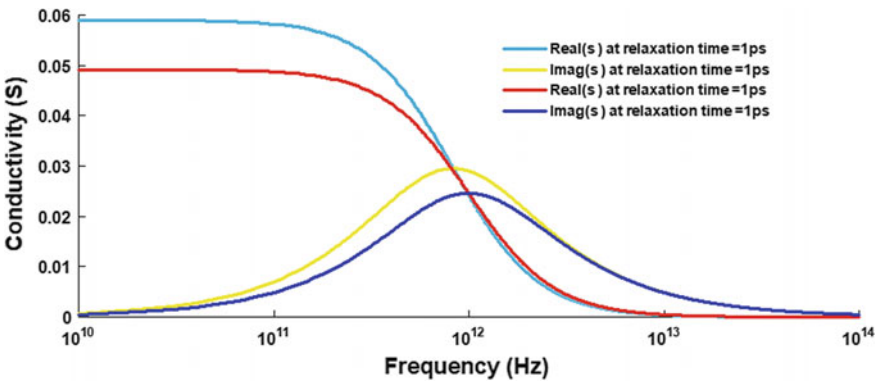
Fig. 3   The change in ac-conductivity with relaxation time 1 and 1.2 ps of graphene
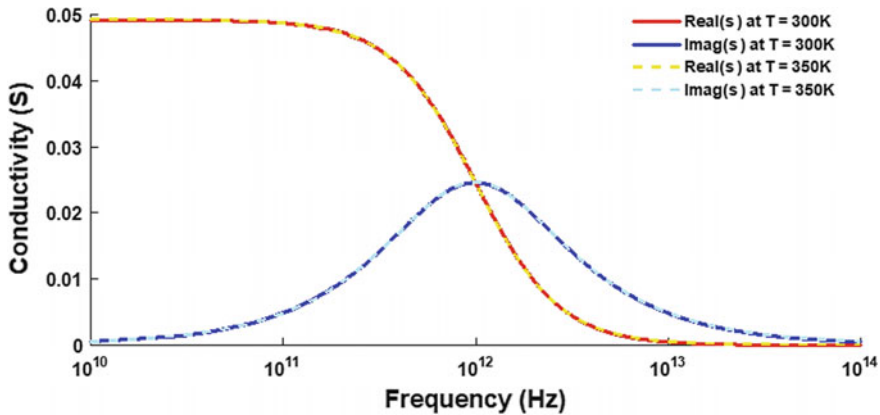
Fig. 4 The ac-conductivity of graphene at temperature 300 and 350 K

from 1.3×10^{13} cm^{-2} to 2.2×10^{13} cm$^{-2,}$ the real part conductivity increases from 0.049 to 0.064 S and the imaginary part is zero (till GHz region) at T = 300 K and τ = 1 ps, as shown in Fig. 2. A similar type of trend could be observed in Fig. 3 where carrier relaxation time is increased from 1 to 1.2 ps at carrier concentration 1.3×10^{13} cm^{-2} and T = 300 K; the real part conductivity increases from 0.049 to 0.059 S and the imaginary part is zero (till GHz region). These values further can be verified with the dc conductivity equation given as [9] (Fig. 4)

$$\sigma(0) = \frac{e^2 E_F \tau}{\pi \hbar^2} = 118 E_F \tau \tag{3}$$

The effect of low-temperature variation on graphene conductivity with the given conductivity equation is not distinguishable (i.e., the conductivity plots with smaller variation in temperature overlap with each other). G. Liu et. al. recently reported that graphene conductivity is inversely proportional to T^2, i.e., the nonlinear relation between conductivity and temperature which in case of metal is linear [21]. Hence, the graphene could have potential application in thermistor where a nonlinear relation is required between resistance and temperature. The conductivity (σ) of graphene is proportional to the mobility (μ) [22]. As the temperature increases, the μ decreases and hence the conductivity decreases (μ proportional to T^{-4}). Although apart from temperature, the SLG mobility depends on several scattering mechanisms like the defect, phonon-boundary, edge, surface, and interface [23] it is still a challenge to detect which scattering effect dominates and how it affects the μ as the temperature varies.

Figure 2 depicts that at a lower frequency (<100 GHz), as the carrier concentration increases, the conductivity increases which makes graphene a suitable material for sensing applications. In the introduction section, the theory behind this is explained with the band structure (Fig. 1) of graphene. The real part conductivity is flat (constant) between 0 to about 100 GHz and the imaginary part is zero. This shows even in

GHz range that graphene is not affected by skin effect or kinetic inductive impedance where metal suffers from these. This behavior could be an asset for the broadband application like $\lambda/4$-filter, $\lambda/2$-coupler, high-speed switching, frequency modulators, and mm-wave links for faster data rate.

In terahertz (THz) range, the intraband transition yield localized the highly confined graphene plasmons. The real and imaginary part of conductivity varies in THz range which is analogous to the frequency-dependent reflective index and dielectric. At THz frequency, when the E_F is away from the E_d, only those photons which have $\hbar\omega > 2|E_F|$ get absorbed and they contribute for conductivity. Due to such properties in THz range, in addition to lightweight and flexibility of graphene, it is made the most suitable material for high-speed data transfer tuneable antenna devices, portable and wearable communication devices, body sensors, and low-power wireless networks. Graphene-based higher frequency devices (operates in GHz and THz) could play a vital role in the development of portable medical diagnostic devices and security screening.

4   Conclusion

The effect of carrier density, temperature, and relaxation time on graphene conductivity has been simulated using MATLAB and its application in sensors and high-frequency devices have been discussed. Despite having zero bandgaps, it has been shown how the sensing object could dope the graphene and shift the Fermi level away from Dirac point which eventually alters the conductivity. This particular property of graphene makes it suitable for sensing application. In the gigahertz region, the frequency-independence (dispersionless) behavior of conductivity put graphene forward for broadband application. When it comes to terahertz region, it is shown that the real and imaginary part of conductivity varies and this variation in conductivity offers control on tuning with very confined scattering at such a high frequency which is remarkable since the conventional metallic material suffers from skin effect, enhanced grain boundary scattering, and roughness loss while tuning at this region of frequency. So this large tuning range makes graphene more suitable for plasmonic antennas.

References

1. Novoselov, K.S., Fal′ko, V.I., Colombo, L., Gellert, P.R., Schwab, M.G., Kim, K.: Nature **490**, 192 (2012)
2. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A.: Science (80-) **306**, 666 LP (2004)
3. Piazza, Z.A., Hu, H.-S., Li, W.-L., Zhao, Y.-F., Li, J., Wang, L.-S.: Nat. Commun. **5**, 3113 (2014)

4. Li, L., Yu, Y., Ye, G.J., Ge, Q., Ou, X., Wu, H., Feng, D., Chen, X.H., Zhang, Y.: Nat. Nanotechnol. **9**, 372 (2014)
5. Bampoulis, P., Zhang, L., Safaei, A., van Gastel, R., Poelsema, B., Zandvliet, H.J.W.: J. Phys. Condens. Matter **26**, 442001 (2014)
6. Tao, L., Cinquanta, E., Chiappe, D., Grazianetti, C., Fanciulli, M., Dubey, M., Molle, A., Akinwande, D.: Nat. Nanotechnol. **10**, 227 (2015)
7. Wang, Q.H., Kalantar-Zadeh, K., Kis, A., Coleman, J.N., Strano, M.S.: Nat. Nanotechnol. **7**, 699 (2012)
8. Sun, Z., Liao, T., Dou, Y., Hwang, S.M., Park, M.-S., Jiang, L., Kim, J.H., Dou, S.X.: Nat. Commun. **5**, 3813 (2014)
9. Samuels, A.J., Carey, J.D., Appl, A.C.S.: Mater. Interfaces **7**, 22246 (2015)
10. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I., Grigorieva, I.V., Dubonos, S.V., Firsov, A.A.: Nature **438**, 197 (2005)
11. Lin, Y.-M., Dimitrakopoulos, C., Jenkins, K.A., Farmer, D.B., Chiu, H.-Y., Grill, A., Avouris, P.: Science (80-) **327**, 662 LP (2010)
12. Low, T., Avouris, P.: ACS Nano **8**, 1086 (2014)
13. Yao, Y., Kats, M.A., Genevet, P., Yu, N., Song, Y., Kong, J., Capasso, F.: Nano Lett. **13**, 1257 (2013)
14. Zhao, L., He, R., Rim, K.T., Schiros, T., Kim, K.S., Zhou, H., Gutiérrez, C., Chockalingam, S.P., Arguello, C.J., Pálová, L., Nordlund, D., Hybertsen, M.S., Reichman, D.R., Heinz, T.F., Kim, P., Pinczuk, A., Flynn, G.W., Pasupathy, A.N.: Science (80-) **333**, 999 LP (2011)
15. Pi, K., Han, W., McCreary, K.M., Swartz, A.G., Li, Y., Kawakami, R.K.: Phys. Rev. Lett. **104**, 187201 (2010)
16. Leenaerts, O., Partoens, B., Peeters, F.M.: Phys. Rev. B **77**, 125416 (2008)
17. Giubileo, F., Di Bartolomeo, A.: Prog. Surf. Sci. **92**, 143 (2017)
18. Alshehri, A.H., Jakubowska, M., Młożniak, A., Horaczek, M., Rudka, D., Free, C., Carey, J.D., Appl, A.C.S.: Mater. Interfaces **4**, 7007 (2012)
19. Alshehri, A.H., Jakubowska, M., Sloma, M., Horaczek, M., Rudka, D., Free, C., David Carey, J.: Appl. Phys. Lett. **99**, 153109 (2011)
20. Mihnev, M.T., Kadi, F., Divin, C.J., Winzer, T., Lee, S., Liu, C.-H., Zhong, Z., Berger, C., de Heer, W.A., Malic, E., Knorr, A., Norris, T.B.: Nat. Commun. **7**, 11617 (2016)
21. Davaji, B., Cho, H.D., Malakoutian, M., Lee, J.-K., Panin, G., Kang, T.W., Lee, C.H.: Sci. Rep. **7**, 8811 (2017)
22. Yin, Y., Cheng, Z., Wang, L., Jin, K., Wang, W.: Sci. Rep. **4**, 5758 (2014)
23. Compton, O.C., Nguyen, S.T.: Small **6**, 711 (n.d.)

Numerical Measurement of Oscillating Parameters of IMPATT Using Group IV and Group III–V Materials

Girish Chandra Ghivela [iD]**, Prince Kumar and Joydeep Sengupta**

Abstract With the help of numerical approach, we have determined the oscillating parameters of Double Drift Region (DDR) Impact Avalanche Transit Time (IMPATT) diode oscillator using different semiconducting materials at Ka band (26.5–40 GHz). The materials used are silicon, germanium, gallium arsenide, indium phosphide, and wurtzite gallium nitride. Avalanche region of IMPATT behaves as an LC parallel circuit. Therefore, inductance, capacitance, and resonant frequency are computed in the avalanche zone by taking all materials individually as substrate elements. Numerically measured inductance and capacitance profiles are in good agreement with earlier reported experimental curves.

Keywords Ka band · Avalanche · Drift · IMPATT · Microwave · Oscillator

1 Introduction

Impact Avalanche Transit Time (IMPATT) diode is capable of generating sufficient power at microwave, millimeter-wave, and submillimeter-wave zones [1]. However, its operation is severely affected by a high level of phase noise that they generated [2]. Dynamic negative resistance property of IMPATT comes from the phase shift between the ac current and voltage. This phase shift is caused by (i) the finite build-up time of the avalanche current called avalanche delay and (ii) the transit time delay—time required by carriers to cross the drift region [2]. The negative resistance arises from these delays, as reported by several published articles [3–12]. These delays affect the phase noise. But the phase noise varies from one IMPATT to the other depending upon which semiconductor material is taken as the base substrate material in the doping perspective. Therefore, in our analysis, we have taken Si, Ge, GaAs, InP, WzGaN as base substrate materials in IMPATT to study the oscillating parameters. The avalanche region behaves as an LC parallel circuit [2]. So

G. C. Ghivela (✉) · P. Kumar · J. Sengupta
Electronics and Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur 440010, India
e-mail: girishvnit2012@gmail.com

the inductance, capacitance, and resonant frequency of IMPATT are computed in the avalanche region along with resonant frequencies.

2   Modeling and Design Parameters

The inductance, capacitance, and resonant frequency of IMPATT are computed through numerical approach by taking one-dimensional DDR IMPATT diode as shown in Fig. 1 at frequency ranging from 26.5 to 40 GHz (Ka band) [13–16]. In the schematic of IMPATT, X_0 is the position of field maximum, W is total layer width which is obtained by summing drift layer width at n-side (d_n), drift layer width at p-side (d_p), and avalanche layer width (X_A), and J_0 represents total current density. Modeling of IMPATT starts with solving of Poisson and current continuity equations, which are given by Eqs. (1) and (2), respectively [17, 18]

$$\frac{\partial E(x)}{\partial x} = \frac{q}{\varepsilon}\left[N_D - N_A + p(x) - n(x)\right] \tag{1}$$

and

$$-\frac{\partial J_n}{\partial x} = \frac{\partial J_p}{\partial x} = \alpha_n J_n + \alpha_p J_p \tag{2}$$

where E(x) is the electric field profile over the length x, q is the charge of electron, and $\varepsilon = \varepsilon_0 \varepsilon_s$ represents absolute permittivity of the semiconductor, N_D and N_A are the ionized donor and acceptor densities, respectively; electron and hole densities at
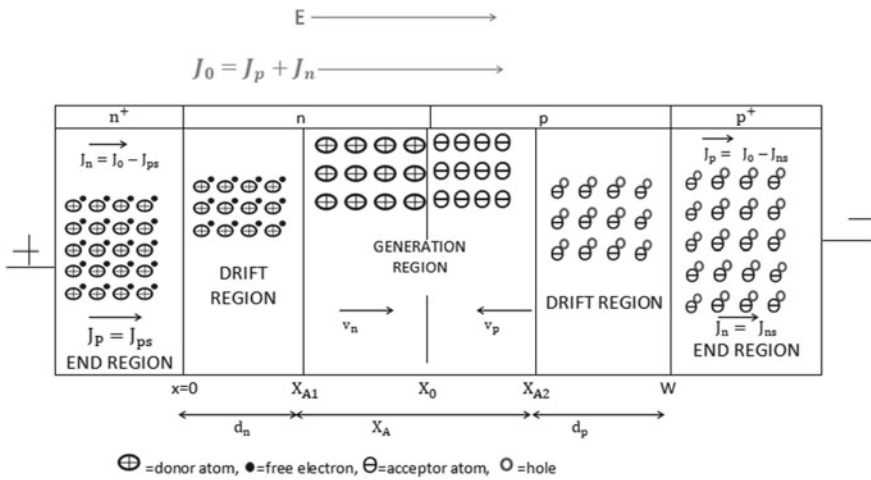
Fig. 1 Schematic of one-dimensional DDR IMPATT

any point x are taken as n(x) and p(x), respectively; $J_n = q\,n\,|v_n|$ and $J_p = q\,n\,|v_p|$ are current densities contributed by electrons and holes, α_n and α_p are the electron and hole ionization rates with v_n and v_p as drift velocities of electrons and holes [17].

Figure 2 represents the equivalent circuit of IMPATT [19]. G and B represent the diode conductance and susceptance, respectively. R_s is the series resistance of the device, g is the load conductance, and L is the circuit inductance. C_p and L_p are the package capacitance and package inductance, respectively. In the avalanche region, the behavior of inductance and capacitance can be modeled in parallel combination, but in the drift region, the drift inductance is in series with drift capacitance [12]. However, the effect of drift inductance is less compared to drift capacitive impedance and can be neglected.

The avalanche response time can be computed by solving the relation given by

$$\tau_A \frac{\partial J}{\partial t} = -(J_p - J_n)|_0^{X_A} + 2J \int_0^{X_A} \alpha \, dx \tag{3}$$

where $\tau_A = X_A/v_s$ is the transit time across the avalanche region, v_s is for saturation velocity, and J is the sum of J_n and J_p.

In the avalanche region, the inductance and capacitance are given as [2]

$$L_A = \frac{\tau_A}{2\,J_0\alpha'A}$$

$$= \frac{\tau_A}{2\,J_s\alpha'A}(1 - \int_0^W \langle\alpha\rangle \, dx)$$
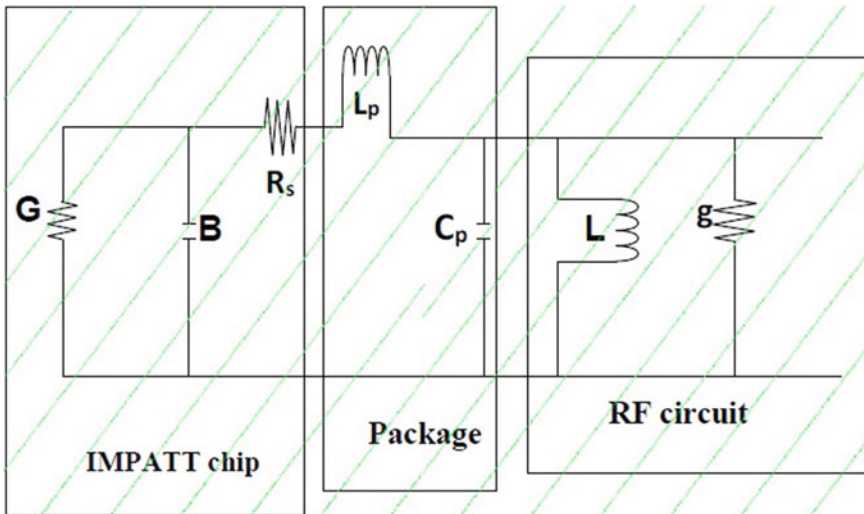
Fig. 2 Equivalent circuit of IMPATT diode with packaging

Table 1 Design parameters

Material	J_0 (A m^{-2})	N_D (m^{-3})	N_A (m^{-3})	v_n (ms^{-1})	v_p (ms^{-1})
Si	1.9×10^9	1×10^{17}	1×10^{17}	1×10^5	0.75×10^5
GaAs	4.6×10^9	1×10^{18}	1×10^{18}	0.8×10^5	0.8×10^5
InP	2.1×10^9	1.5×10^{17}	1.5×10^{17}	0.6×10^5	0.76×10^5
Ge	4.4×10^9	1.3×10^{17}	1.3×10^{17}	0.6×10^5	1×10^5
WzGaN	5×10^8	1.2×10^{19}	1.2×10^{19}	2×10^5	2.5×10^5

$$= \frac{\tau_A}{2\, J_s \left(\frac{\partial \alpha}{\partial E}\right) A} (1 - \int_0^W \langle \alpha \rangle \, dx) \tag{4}$$

$$C_A = \frac{\varepsilon_s A}{X_A} \tag{5}$$

and the resonant frequency of this LC combination is given by

$$\omega_r = 2\,\pi\, f_r = \sqrt{\frac{2\, \alpha' v_s J_0}{\varepsilon_s}} \tag{6}$$

where J_0 is the direct current density under the dc condition and related to thermally generated reverse saturation current density J_s, $\alpha' = (\partial \alpha / \partial E)$, A and α are diode area and ionization integrand, respectively [2]. In the drift region, response time is given by

$$\tau_D = \frac{(W - X_A)}{v_s} \tag{7}$$

and the drift capacitance is

$$C_{Drift} = \frac{A\, \varepsilon_s}{(W - X_A)} \tag{8}$$

Recently reported design parameters for the abovementioned materials are used in our numerical study as listed in Table 1 [20].

3 Results and Discussion

The behavior of the avalanche region as oscillator can be analyzed from the numerically obtained inductance and capacitance profiles as shown in Figs. 3 and 4. The nonlinear variation of inductance and capacitance profiles comprising an antiresonant behavior in the avalanche region. Our numerically obtained results are in good agreement with the experimentally reported inductance and capacitance profiles by
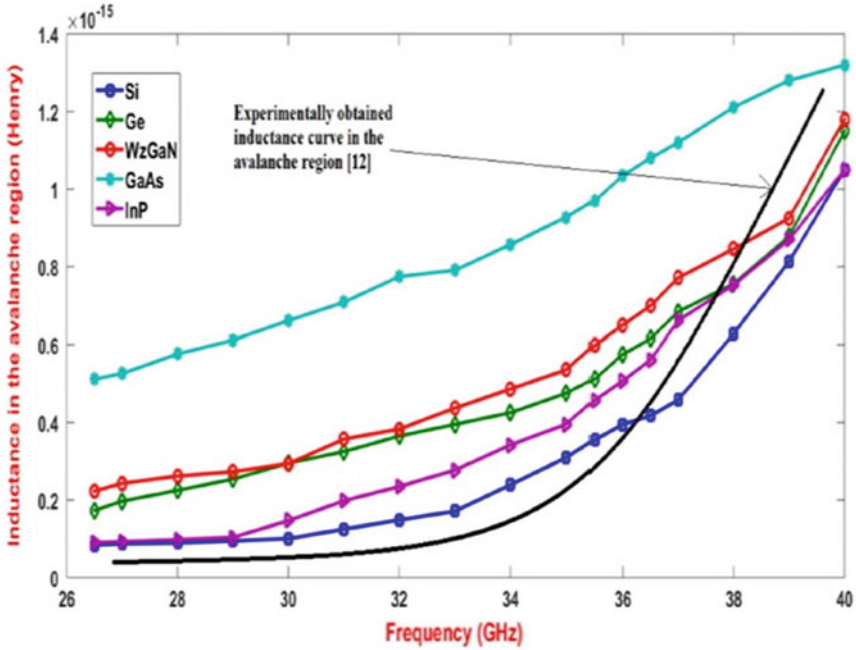
Fig. 3 Avalanche inductance profiles over the Ka band frequency regime
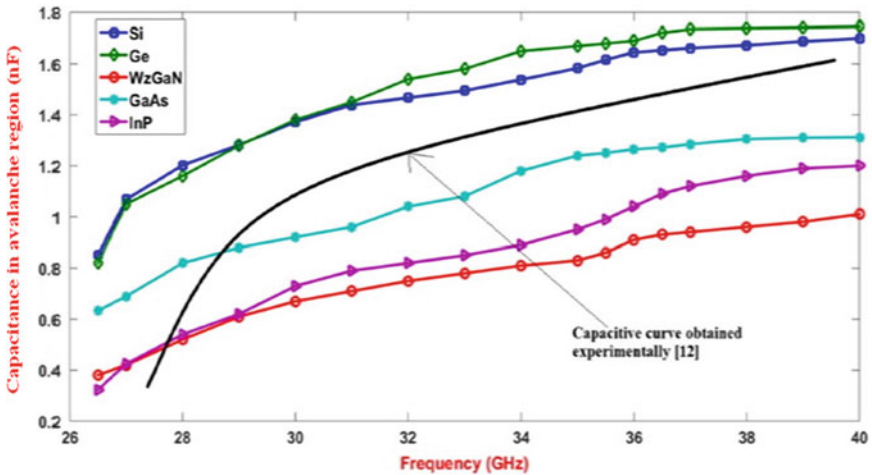
Fig. 4 Avalanche capacitance profiles with variation of frequency

[12]. These experimental curves are shown by thick black curves in Figs. 3 and 4. Inductance in the avalanche region is more for the GaAs-based DDR IMPATT as shown in Fig. 3. However, from the other materials-based DDR IMPATTs, inductance values in the avalanche zone are nearly close to each other at the corresponding frequency. From the plot of Fig. 4, we can observe that capacitance in the avalanche zone is more for Ge and less for WzGaN-based IMPATT.

In the Ka band, simulated drift capacitances and resonant frequencies for all the materials-based IMPATT are given in Table 2 and Table 3, respectively. The drift capacitance is increasing with the increase in frequency for all the materials. Ge-based IMPATT is having higher drift capacitance variation from 0.0785 to 0.1213 nF and its WzGaN counterpart is having lower variation from 0.0161 to 0.0242 nF. InP is having lower values of resonant frequencies compared to others at the corresponding operating frequencies.

Table 2 Numerically measured drift capacitance

f (GHz)	Si	Ge	WzGaN	GaAs	InP
	C_{Drift} (nF)	C_{Drift} (nF)	C_{Drift} (nF)	C_{Drift} (nF)	C_{Drift} (nF)
26.5	0.0505	0.0785	0.0161	0.0528	0.0766
27	0.0515	0.0801	0.0163	0.0538	0.0783
28	0.0532	0.0834	0.0169	0.0559	0.0819
29	0.0551	0.0866	0.0172	0.0581	0.0824
30	0.0569	0.0889	0.0182	0.0601	0.0846
31	0.0588	0.092	0.0185	0.0623	0.0875
32	0.0607	0.0949	0.0194	0.0644	0.0907
33	0.0626	0.0981	0.0198	0.0665	0.0397
34	0.0645	0.1012	0.0207	0.0686	0.097
35	0.0664	0.1039	0.0209	0.0706	0.0996
35.5	0.0673	0.1054	0.0215	0.0716	0.1008
36	0.0682	0.1068	0.0217	0.0724	0.1024
36.5	0.0693	0.1086	0.0221	0.0737	0.1048
37	0.0703	0.1103	0.0224	0.0749	0.1068
38	0.0724	0.1138	0.0229	0.0773	0.1106
39	0.0745	0.1175	0.0235	0.0798	0.1147
40	0.0766	0.1213	0.0242	0.0822	0.1192

Table 3 Numerically measured resonant frequency

f (GHz)	Si	Ge	WzGaN	GaAs	InP
	ω_r ($\times 10^7$ rad/s)	ω_r ($\times 10^7$ rad/s)	ω_r ($\times 10^7$ rad/s)	ω_r ($\times 10^7$ rad/s)	ω_r ($\times 10^7$ rad/s)
26.5	10.61	8.452	10.72	5.556	1.851
27	10.46	8.327	10.23	5.474	1.82
28	9.688	8.08	9.662	5.18	1.776
29	9.081	7.875	9.452	5.007	1.592
30	8.518	7.277	9.01	4.774	1.46
31	8.137	7.019	8.75	4.579	1.397
32	7.916	6.711	8.37	4.447	1.353
33	7.713	6.513	8.12	4.291	1.313
34	7.534	6.3	7.928	4.070	1.267
35	7.284	5.876	7.675	3.856	1.202
35.5	7.133	5.717	7.356	3.732	1.164
36	6.982	5.555	6.129	3.544	1.149
36.5	7.05	5.6	6.054	3.612	1.165
37	7.029	5.6	6.012	3.666	1.169
38	6.995	5.616	5.96	3.697	1.155
39	6.972	5.714	5.851	3.739	1.151
40	6.959	5.774	5.803	3.78	0.818

4 Conclusion

The oscillating parameters of IMPATT are measured through numerical approach. A comparison among them is presented so that we can have an idea of comparative oscillating behavior. This will further help in the determination of high power generation capability by the materials through the study of avalanche response time.

References

1. Midford, T.A., Bernick, R.L.: Millimeter wave CW IMPATT diodes and oscillators. IEEE Trans. Microw. Theory Tech. **27**, 483–492 (1979). https://doi.org/10.1109/TMTT.1979.1129653
2. Sze, S.M., Ng, K.K.: Physics of Semiconductor Devices, pp. 466–488. Wiley, New Jersey (2007)
3. Read, W.: A proposed high-frequency negative-resistance diode. Bell Syst. Tech. J. **37**, 401–446 (1958). https://doi.org/10.1002/j.1538-7305.1958.tb01527.x

4. Johnston, R.L., DeLoach Jr., B.C., Cohen, B.G.: A silicon diode microwave oscillator. Bell Syst. Tech. J. **44**, 369–372 (1965). https://doi.org/10.1002/j.1538-7305.1965.tb01667.x

5. Lee, C.A., Batdorf, R.L., Wiegmann, W., Kaminski, G.: The read diode—an avalanching, transit-time, negative resistance oscillator. Appl. Phys. Lett. **6**, 89–91 (1965). https://doi.org/10.1063/1.1754180

6. Shockley, W.: Negative resistance arising from transit time in semiconductor diode. Bell Syst. Tech. J. **33**, 799–826 (1954). https://doi.org/10.1002/j.1538-7305.1954.tb03742.x

7. DeLoach Jr., B.C.: The IMPATT story. IEEE Trans. Electron Devices **23**, 657–660 (1976). https://doi.org/10.1109/T-ED.1976.18469

8. Misawa, T.: Negative resistance in p-n junction under avalanche breakdown conditions, Part I. IEEE Trans. Electron Devices **13**, 137–143 (1966). https://doi.org/10.1109/T-ED.1966.15647

9. Misawa, T.: Negative resistance in p-n junction under avalanche breakdown conditions, Part II. IEEE Trans. Electron Devices **13**, 143–151 (1966). https://doi.org/10.1109/T-ED.1966.15648

10. Gummel, H.K., Blue, J.L.: A small signal theory of avalanche noise in IMPATT diodes. IEEE Trans. Electron Devices **14**, 569–580 (1967). https://doi.org/10.1109/T-ED.1967.16005

11. Sze, S.M., Ryder, R.M.: Microwave avalanche diodes. Proc. IEEE **59**, 1140–1154 (1971). https://doi.org/10.1109/PROC.1971.8360

12. Gilden, M., Hines, M.E.: Electronic tuning effects in the read microwave avalanche diode. IEEE Trans. Electron Devices **13**, 169–175 (1966). https://doi.org/10.1109/T-ED.1966.15652

13. Acharyya, A., Banerjee, J.P.: Prospects of IMPATT devices based on wide bandgap semiconductors as potential terahertz sources. Appl. Nanosci. **4**, 1–14 (2014). https://doi.org/10.1007/s13204-012-0172-y

14. Ghivela, G.C., Sengupta, J., Mitra, M.: Ka band noise comparison for Si, Ge, GaAs, InP, WzGaN, 4H-SiC based IMPATT diode. Int. J. Electron. Lett. **7**, 107–116 (2019). https://doi.org/10.1080/21681724.2018.1460869

15. Sengupta, J., Ghivela, G.C., Gajbhiye, A., Mitra, M.: Measurement of noise and efficiency of 4H-SiC IMPATT diode at Ka band. Int. J Electron. Lett. **4**, 134–140 (2016). https://doi.org/10.1080/21681724.2014.966774

16. Sengupta, J., Ghivela, G.C., Mitra, M.: Dynamic characterization and noise analysis of 4H-SiC IMPATT diode at Ka band. Int. J. Soft Comput. Eng. **4**, 145–149 (2014)

17. Ghivela, G.C., Sengupta, J.: Prospects of impact avalanche transit time diode based on chemical vapor deposited diamond substrate. J. Electron. Mater. **48**, 1044–1053 (2019). https://doi.org/10.1007/s11664-018-6821-5

18. Sengupta, J., Ghivela, G.C., Gajbhiye, A., Jothe, B., Mitra, M.: Temperature dependence of 4H-SiC IMPATT diode at Ka band. Int. J. Electr. Electron. Comput. Syst. **19**, 1–5 (2014)

19. Cullen, A.L., Forrest, J.R.: Analytic theory of the IMPATT diode and its application to calculations of oscillator locking characteristics. Proc. IEE **121**, 1467–1474 (1974). https://doi.org/10.1049/piee.1974.0308

20. Ghivela, G.C., Sengupta, J., Mitra, M.: Space charge effect of IMPATT diode using Si, Ge, GaAs, InP, WzGaN, 4H-SiC at Ka band. IETE J. Educ. **58**, 61–66 (2017). https://doi.org/10.1080/09747338.2017.1378132

Analysis of SRAM Cell for Low Power Operation and Its Noise Margin

Sunil Kumar Ojha, O. P. Singh, G. R. Mishra and P. R. Vaya

Abstract In recent years improvement in the design of SRAM cell increases drastically. The two major factors which have to be taken care are power dissipation and the noise margin of the SRAM cell. The power is subdivided into two groups that are switching power and standby power. The leakage current plays an important role in the power dissipation and has to be taken care. The noise margin also categorized as read margin and write margin. The purpose of this paper is to analyze and optimize the SRAM cell operation with respect to power and also measure the noise margin. The results show the low power operation of SRAM cell with relatively improved noise margin.

Keywords SoC · SRAM cell · Subthreshold leakage · Noise margin · Current ratio of MOSFET · Short channel effect

1 Introduction

Memories (Specially SRAMs) are widely used in the electronic subsystems. According to the current trends nearly all NOCs and SOCs use 70% of their silicon area for the memory, hence it became more important to give attention to the reliability of the memory on-chip. The stability of the memories is major concern for the inte-

S. K. Ojha (✉) · O. P. Singh · G. R. Mishra
Department of Electronics & Communication Engineering, Amity School of Engineering & Technology (ASET), Amity University, Lucknow Campus, Lucknow, Uttar Pradesh, India
e-mail: sunilrirt07@gmail.com

O. P. Singh
e-mail: opsingh@amity.edu

G. R. Mishra
e-mail: grmishra@amity.edu

P. R. Vaya
Department of Electrical Engineering, Indian Institute of Technology Madras (IITM), Chennai, India
e-mail: prvaya@gmail.com

413

grated circuit designers. The stability of memory cell determines the sensitivity of the cell to process variation. Various major efforts have been already in effect for the analysis of noise margin of SRAM cell and to effectively calculate the stability of overall design. The read and write stability are two major factors which decide the reliability of any memory. Since the process technology trends are changing very rapidly and it is nearly moving towards less than 10 nm process node; therefore it is a mandatory need to perform the static noise margin of the memory cell. Further, with the reduction of process technology nodes many other parameters are affected which includes the power supply modification, threshold voltage reduction, leakage current, energy consumption, short channel effect, on/off ratio of the MOSFET and many others. A brief explanation has been provided for all the above said factors in subsequent parts of this paper and finally the design model of SRAM cell is analyzed by considering the effect of all the related parameters.

2 Minimum Energy Consumption Modeling

Under this topic modeling of drain current in the subthreshold region is discussed. To do the analysis of minimum energy the current model will be appropriate for calculations. The developed model for the required minimum energy point allows immediate calculation of main factors and the basic key parameters on the system energy.

Subthreshold leakage current model:-
During the subthreshold operation the channel of the transistor is not inverted due to the diffusion of the current flow. Equation (1) is a basic equation for modeling subthreshold current and total off current.

$$I_{D:Sub-Threshold} = I_0 \exp\left(\frac{V_{gs} - V_T}{nV_{th}}\right). \tag{1}$$

Below Eq. (2) shows some basic equation with V_{DS} roll-off:

$$I_{D:Sub-Threshold} = I_0 \exp\left(\frac{V_{gs} - V_T}{nV_{th}}\right)\left(1 - \exp\left(\frac{-V_{ds}}{V_{th}}\right)\right). \tag{2}$$

where I_0 is the drain current when $V_{GS} = V_T$ given by Eq. (3).

$$I_0 = \mu_0 C_{ox}\left(\frac{W}{L}\right)(n - 1)V_{th}^2. \tag{3}$$

It is assumed that the total drain current in subthreshold equals subthreshold current. As expected for diffusion current ($I_{D:Sub-Threshold}$) shows that I_D depends exponentially on V_{GS}. V_T is transistor threshold voltage. N is the subthreshold slope vector ($n = 1 + \frac{C_d}{C_{ox}}$) and V_{th} is thermal voltage.

$$V_{th} = \frac{kT}{q}$$

3 Short Channel Effect

It leads to an increase in leakage current. The short channel is an effect in MOS transistors where the channel length of the transistors is of the same order as the depletion layer width of source and drain junction. When the length of the channel is reduced to improve the operating speed and number of component per chips, the short channel effects arise.

The major reasons for the short channel effect are due to two main phenomena:-

1. The limitation of the electron drift in the channel.
2. The change in the threshold voltage due to the short channel length.

$$V_T(\text{Short Channel}) = V_{T0} - \Delta V_{T0}. \tag{4}$$

V_{T0} Zero-Bias threshold voltage
ΔV_{T0} Threshold voltage shifts (reduction) due to the short channel effect

$$V_T = V_{T0} + \gamma \left(\sqrt{|-2\varphi_F + V_{SB}|} - \sqrt{2\varphi_F} \right). \tag{5}$$

$$\Delta V_{T0} = \frac{1}{C_{ox}} \sqrt{2q\epsilon_{si}N_A |2\varphi_F|} . \frac{x_j}{2L} \left[\left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1 \right) + \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1 \right) \right]. \tag{6}$$

$$\gamma = \frac{\sqrt{2qN_A\epsilon_{si}}}{C_{ox}}. \tag{7}$$

$$V_{T0} = \varphi_{GC} - 2\varphi_F - \frac{Q_{Bo}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}}. \tag{8}$$

4 Leakage Current

The leakage current is divided into two main subgroups the first group is leakage current due to reverse bias pn-junction current, and leakage current due to subthreshold channel conduction current.

Current–voltage equation of n-channel MOS:
For n-channel MOSFET:

$$I_D = 0; \; V_{GS} = V_T$$

$$I_{D(Lin)} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L} \left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2 \right].$$
$$: V_{GS} \geq V_T \text{ and } V_{DS} < V_{GS} - V_T \tag{9}$$

$$I_{D(Sat)} = \frac{\mu_n C_{ox}}{2} \cdot \frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}).$$
$$: V_{GS} \geq V_T \text{ and } V_{DS} \geq V_{GS} - V_T \tag{10}$$

In CMOS logic, the leakage current is the only source of static power dissipation.

If Ion (on current) increases the operating speed of the circuit also increases; therefore it is advisable to use a small V_T.

When $V_{GS} < V_T$: The n-channel MOSFET will be in the off state. Nevertheless, an unwanted leakage current may flow between the drain and source. The current observes at $V_{GS} < V_T$ is called the subthreshold current. This leads to the MOSFET off state current I_{off}.

I_{off} is the I_D measured at $V_{GS} = 0$ and $V_{DS} = V_{DD}$.

It is mandatory to maintain I_{off} as smaller value so that it helps in minimizing the consumption of static power if the circuit is in the standby (Fig. 1).

It is plotted between V_{GS} and I_D. The current flow at V_{GS} V_T is called subthreshold current (Fig. 2).

When VT becomes higher than VGS then electron concentration w.r.t. inversion reduces but this also might result a small leakage current to flow from source to the drain.

V_T Roll-Off: Short channel MOSFET is hard to turnoff: V_T must not be made much low; doing so results I_{off} would be very large. The channel length must not
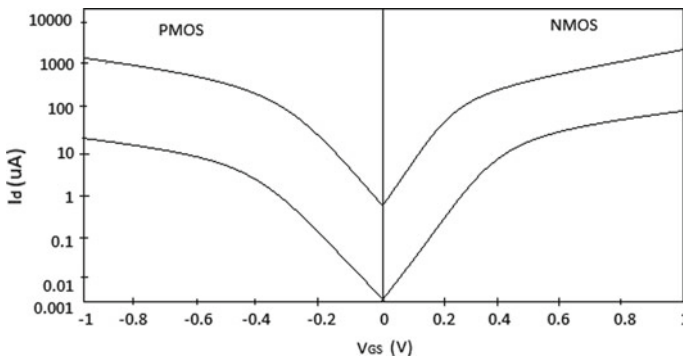
Fig. 1 The typical subthreshold current plot
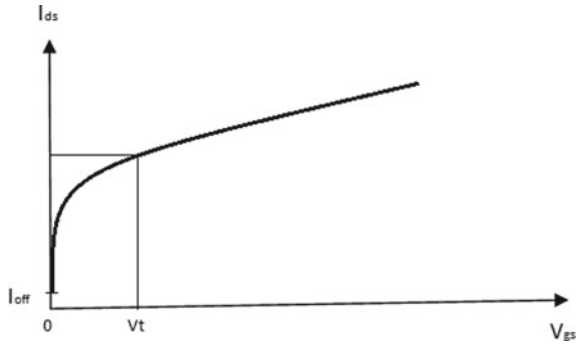
Fig. 2 Subthreshold I–V with V_T and I_{off}

be very short since V_T reduces if any reduction in L. when V_T reduces below a certain level the I_{off} increases at large extent and also the channel length will not be appropriate (Fig. 3).

On/Off Current Ratio

Ion is specified as the I_{Dsat} for particular gate voltage (usually max V_{GS}) $V_{DS} = V_{GS}$. I_{off} is the I_{DSS} (Leakage current) where $V_{GS} = 0$ and $V_{DS} = $ max. I_{on}/I_{off} is the figure of merit for having high performance (more I_{on}) and low leakage current (less I_{off}) for CMOS transistors. Typically more gate controls lead to more I_{on}/I_{off}. Also, The I_{on}/I_{off} ratio reduces relative to strong inversion in the subthreshold region.
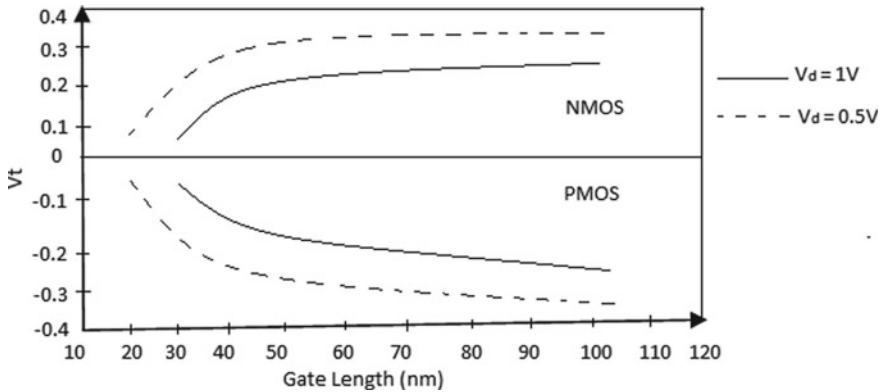
Fig. 3 VT decreasing with decreasing Lg
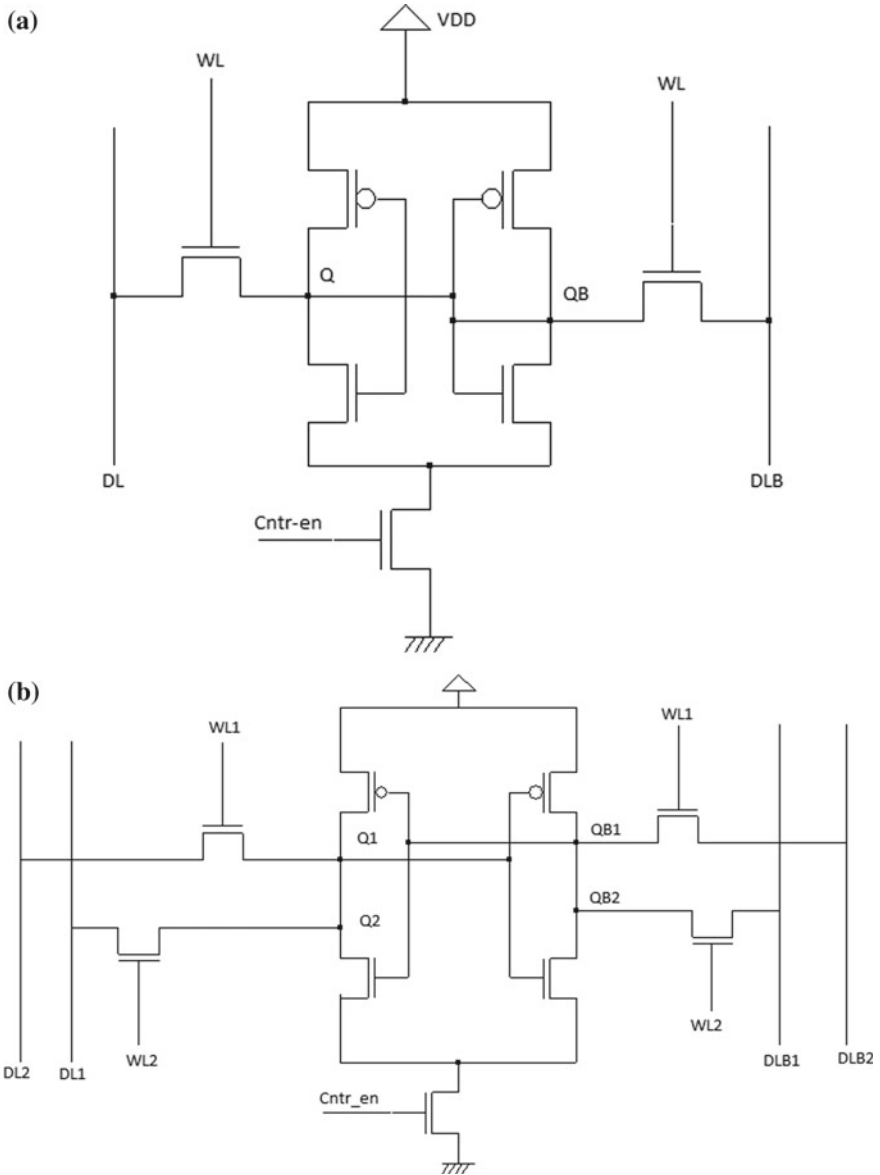
Fig. 4 **a** Single port SRAM cell with control enable, **b** double port SRAM cell with control enable

5 SRAM Cell

Figure 4a shows the single-port SRAM cell. The single-port SRAM cell consists of six transistors in which two are PMOS and remaining four are NMOS, but the below SRAM cell has one extra NMOS transistor for control enable signal this particular NMOS will act as a switch and it will be helpful in leakage current reduction in the fact that when the signal "Cntr-en" is high the switch is activated and the cell will function normally but when the signal is low this will provide a high resistance path and will reduce the leakage current.

The double port SRAM cell consists of eight transistors in which two are PMOS and the remaining six are NMOS. The extra NMOS switch is also added here and it will act as per the above circuit, i.e., it will help in reduction of leakage current. Double port SRAM cell is helpful in allowing multiple reads or writes operation to be performed at the same time. Now a day's most CPU processors consist of double ported SRAM in it. Figure 4b shows the double port SRAM cell.

6 Static Noise Margin

For the proper read operation, the data line is precharged to the power supply. The approach includes an SRAM cell is presented as two equivalent inverters as shown in Fig. 5. The two appropriate noise sources is inserted between input and output node. The connected noise voltage sources must have the same value and should act together to disturb the present stable state of the cell. By applying such noise source
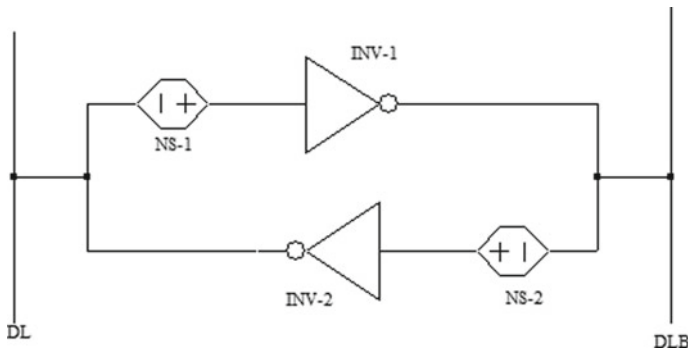
Fig. 5 Circuit for calculating SNM of SRAM cell

the simulator gives the input–output voltage curve and that curve represents the noise margin for worst case. This arrangement of noise source and its application to the input of each inverter makes the value of obtained SNM to be the worst-case SNM. The noise margin setup is done using 7 nm standard process technology node and the corresponding waveform is obtained using HSPICE simulation.

7   Simulation and Results

The proposed SRAM cells are simulated using HSPICE simulator and the process node for transistors used is chosen as standard 7 nm. Figure 6a shows the values of the supply voltage and the word line voltage.

Figure 6b shows the values of the data line voltages. The data line voltages must be opposite to each other.

By the application of the above supply, the resultant waveform of internal storage nodes is shown in Fig. 6c.

Figure 6d shows the leakage current and power dissipation for double port SRAM cell, while the Fig. 6e shows the leakage current and power dissipation for a single port SRAM cell.

Figure 7a shows the switching power of the SRAM cell, i.e., how much power the cell consumes when it makes switching from 0 to 1 or vice versa.

Figure 7b–e Shows the read and write curve for single and double port SRAM cell.

All the input–output supply voltages and currents, corresponding power dissipation and leakage currents are tabulated in the Table 1. The table also includes the noise margin values for the SRAM cells. The correction factor is shown so that the design must sustain the values in the different working scenarios such as different process, voltage, and temperature conditions.
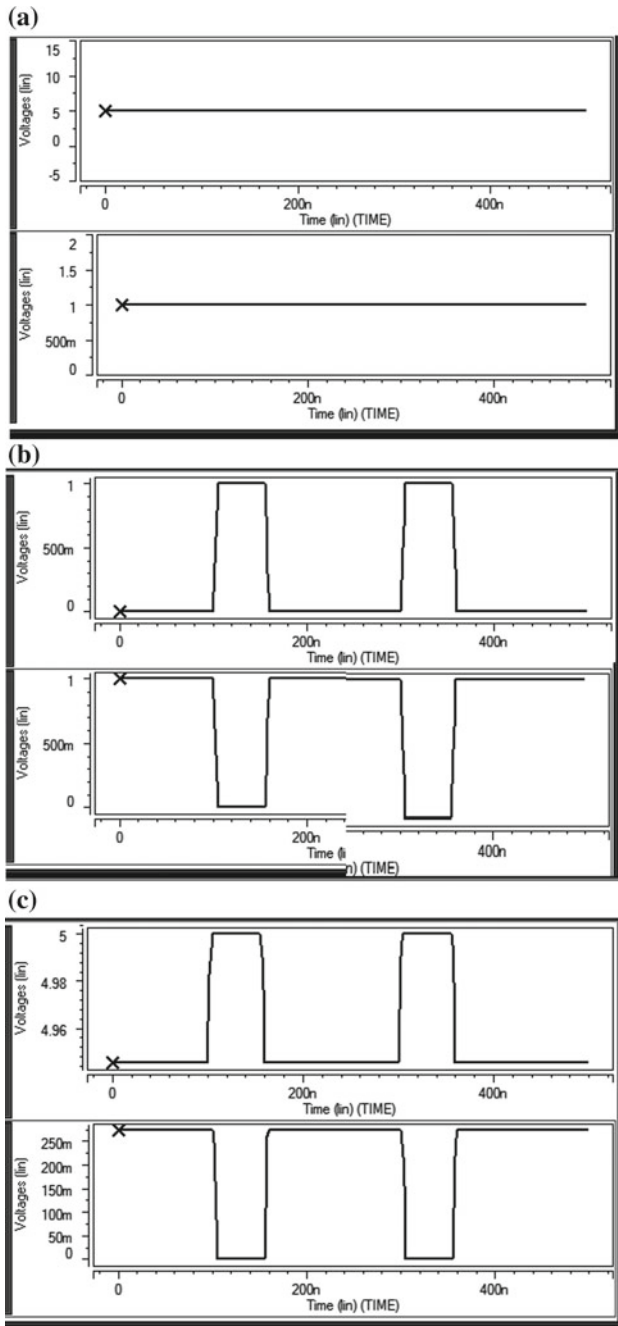
Fig. 6 **a** Input supply voltages, **b** data line voltages, **c** signals at internal nodes, **d** leakage current and power dissipation for double port SRAM cell, **e** leakage current and power dissipation for single port SRAM cell
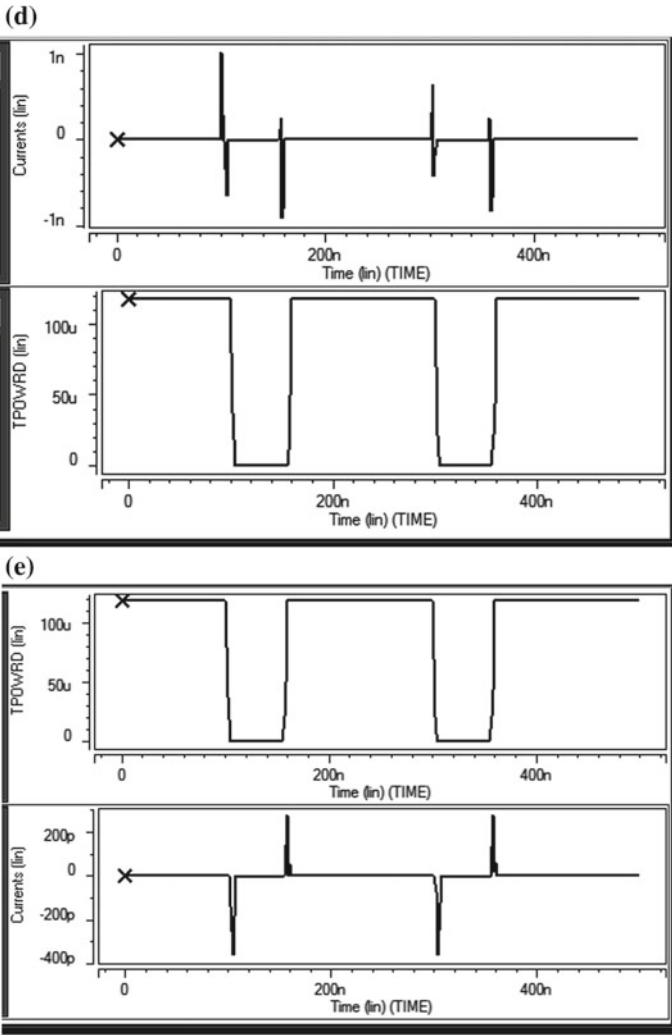
Fig. 6 (continued)
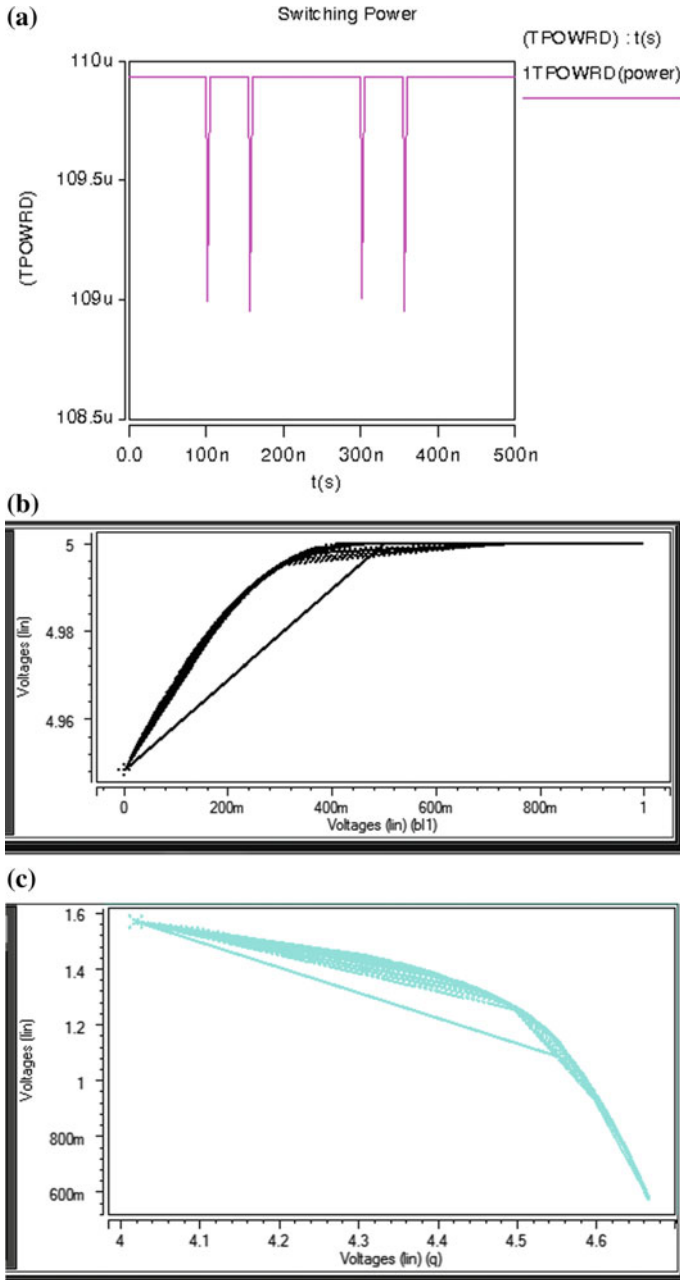
Fig. 7 **a** Switching power, **b** read margin curve for single port SRAM cell, **c** write margin curve for single port SRAM cell, **d** read margin curve for double port SRAM cell, **e** write margin curve for double port SRAM cell

(d)

(e)
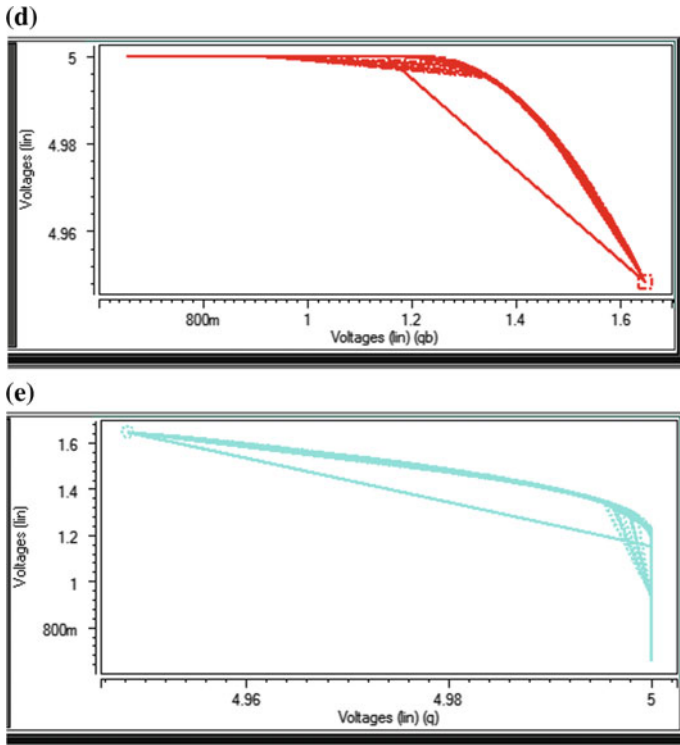
Fig. 7 (continued)

Table 1 Results and Values

Sr. no.	Signal name	Signal type	Values	Correction factor
1	V_{dd} (Max.)	Input voltage supply	5 V DC	NA
2	V_{dd} (Min.)	Input voltage supply	1.8 V DC	NA
3	Word line (WL)	Input voltage supply	1 V DC	NA
4	Data line (DL)	Input voltage supply	1 V pulse	NA
5	Data line prime (DLB)	Input voltage supply	1 V pulse	NA
6	Internal storage (Q and QB)	Output voltage	0.67 V pulse	NA
7	Power dissipation	Output	110 μW	+10% −10%

Table 1 (continued)

Sr. no.	Signal name	Signal type	Values	Correction factor
8	Leakage current	Output current	200 pA	+10% −10%
9	Read margin	Output voltage	290 mV	+5% −5%
10	Write margin	Output voltage	240 mV	+5% −5%
11	SNM	Output relative voltage	320 mV	+5% −5%

8 Conclusion

This paper has presented the technique to design SRAM cell for power efficient applications. The presented design is also useful to improve the SNM of the cell; this noise margin improvement helps in enhancing the overall system performance. All the design presented in this paper is implemented using HSPICE simulator and the process technology is selected as 7 nm. The proposed design will be useful in SOC and NOC and various other integrated chips. The improved noise margin observed as 320 mV with the minimal supply voltage of 1.8 V. The power dissipation is reported as 110 µW which is very efficient in low power applications.

References

1. Pavlov, A., Sachdev, M.: CMOS SRAM Circuit Design and Parametric Test in Nano-scaled Technologies: Process-Aware SRAM Design and Test. Springer Science (2008)
2. Farkhani, H., et al.: A new asymmetric 6T SRAM cell with a write assist technique in 65 nm CMOS technology. Elsevier Microelectron. J. **45**, 1556–1565 (2014)
3. Hassanzadeh, S., et al.: A novel low power 8T-cell sub-threshold SRAM with improved Read-SNM. In: International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS), pp. 35–38, March (2013)
4. Dhilleswararao, P., et al.: High SNM 32 nm CNFET based 6T SRAM Cell design considering transistor ratio. In: International Conference on Electronics and Communication System (lCECS-2014), pp. 1–6, Feb (2014)
5. Wang, W., et al.: High SNM 6T CNFET SRAM cell design considering nanotube diameter and transistor ratio. In: IEEE International Conference on Electro/Information Technology (EIT), pp. 1–4, May (2011)
6. Madiwalar, B., et al.: Single bit-line 7T SRAM cell for low power and high SNM. In: International Multi-conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), pp. 223–228, March (2013)
7. Sil, A., et al.: A novel high write speed, low power, read-SNM-free 6T SRAM cell. In: Midwest Symposium on Circuits and Systems, pp. 771–774, Aug (2008)
8. Sil, A., et al.: A novel 8T SRAM cell with improved read-SNM. In: IEEE Northeast Workshop on Circuits and Systems, pp. 1289–1292, Aug (2007)

9. Solanki, S., et al.: A low-leakage single-ended 6T SRAM cell. In: 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET), pp. 698–702, Nov (2010). ISSN 2157-0477

10. Chen, S.-Y., et al.: Single-ended disturb-free 5T loadless SRAM Cell using 90 nm CMOS process. In: IEEE International Conference on IC Design & Technology (ICICDT), pp. 1–4, June (2012). ISBN 978-1-4673-0146-6

11. Giraud, B., et al.: A novel 4T asymmetric single-ended SRAM cell in sub-32 nm double gate technology. In: IEEE International Symposium on Circuits and Systems, pp. 1906–1909, May (2008). ISBN 978-1-4244-1683-7

12. Singh, J., et al.: A subthreshold single ended I/O SRAM cell design for nanometer CMOS technologies. In: IEEE International SOC Conference, pp. 243–246, Sept (2008). ISBN 978-1-4244-2596-9

13. Kushwah, C.B., et al.: A single-ended with dynamic feedback control 8T subthreshold SRAM cell. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. 1–5 (2015). ISSN: 1063-8210

14. Hobson, R.F.: A new single-ended SRAM cell with write-assist. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **15**(2), 173–181 (2007). ISSN 1063-8210

15. Yang, Y., et al.: Single-ended 9T SRAM cell for near-threshold voltage operation with enhanced. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. (99), 1–5 (2014). ISSN: 1063-8210

16. Li, Q., et al.: A 5.61 pJ, 16 kb 9T SRAM with single-ended equalized bitlines and fast local write-back for cell stability improvement. In: Proceedings of the European Solid-State Device Research Conference (ESSDERC), pp. 201–204, Sept (2012). ISBN 978-1-4673-1707-8

17. Wen, L., et al.: Single-ended, robust 8T SRAM cell for low-voltage operation. Microelectron. J. **44**(8), 718–728 (2013)

18. Anand, N., et al.: Highly stable subthreshold single-ended 7T SRAM cell. In: 2nd International Conference on Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN), pp. 1–4, Dec (2014). ISBN 978-1-4799-6985-2

19. Ichikawa, T., et al.: A new analytical model of SRAM cell stability in low-voltage operation. IEEE Trans. Electron Devices **43**(1), 54–61 (2002). ISSN: 0018-9383

20. Chang, L., et al.: An 8T-SRAM for variability tolerance and low-voltage operation in high-performance caches. Solid-State Circuits **43**(4), 956–963 (2008). ISSN: 0018-9200

21. Mann, R.W., et al.: Impact of circuit assist methods on margin and performance in 6T SRAM. Solid State Electron. J. **54**, 1398–1407 (2010)

22. Zhang, K., et al.: A 3-GHz 70 Mb SRAM in 65 nm CMOS technology with integrated column-based dynamic power supply. In: IEEE International Solid-State Circuits Conference, pp. 474–611, February (2005)

23. Kaur, R., et al.: A 6T SRAM cell based pipelined 2R/1W memory design using 28 nm UTBB-FDSOI. In: 28th IEEE International System-on-Chip Conference (SOCC), pp. 310–315, September (2015)

24. Sinangil, M.E., et al.: A 45 nm 0.5 V 8T column-interleaved SRAM with on-chip reference selection loop for sense-amplifier. In: IEEE Asian Solid-State Circuits Conference, pp. 225–228, November (2009)

25. Wang, D.P., et al.: A 45 nm dual-port SRAM with write and read capability enhancement at low voltage. In: IEEE International SOC Conference, pp. 211–214, September (2007)

26. HSPICE user manual simulation and analysis

27. Predictive Technology Model (PTM). http://ptm.asu.edu/

Fabrication of Nano-petals $Zn_{0.97}Cu_{0.03}O$ Thin Film and Application in Methane Sensing

Brij Bansh Nath Anchal, Preetam Singh and Ram Pyare

Abstract This paper presents the fabrication of nano-petals $Zn_{0.97}Cu_{0.03}O$ thin film using electro-spin technique and sensing of methane. Doping of 3% Cu into ZnO was carried out by chemical route and spin coated thin film was grown on a glass substrate. Analysis of XRD, SEM, EDX and AFM characterizations were performed for surface morophology and grain sizes of nano-petals. Study of sensitivity for 1000, 1200 and 1400 ppm methane as better sensing at low operating temperature 125 °C.

Keywords XRD · SEM · EDX · AFM · Nano-petals · Thin film · Sensor

1 Introduction

The band gap of Zinc oxide is as large as 3.37 eV. It is hexagonal wurtzite structured crystalline, high thermal stabled n type semiconductor material. Development of nanostructures in the ZnO based thin film, used in gas sensing [1]. A. Yu et al., reported Fe doped ZnO thin film based improved the gas sensing properties [1]. Recent literature reported to different ZnO based nanostructures such as nanorods, nanofiberes and nanorod array based sensors enhance sensitivity [2–7]. The ZnO nanorods based ethanol sensor in operating range of 5–500 ppm concentrations of alcohol vapours at an operating temperature of 370 °C [2]. The ZnO nanofibers based sensor shown high sensitive in H_2 sensing for 0.1–10 ppm concentrations at 350 °C [3]. Earlier literature reported that enhance sensitivity by development of nano material based structure and transition metals doping [1]. Doping of Cu ions in the form of cuprous and cupric ions in lattice and decreased the depletion layer and enhance sensitivity in the presence of reducing gases. The depletion layers created at boundaries of grains and particles, while porosity increased the adsorption sites. These parameters cause surface reduction [7–10]. The film deposition techniques as chemical, physical, evaporation, thermal, dip, drop, spray and spin coatings. The spin

B. B. N. Anchal (✉) · P. Singh · R. Pyare
Department of Ceramic Engineering, Indian Institute of Technology (BHU), Varanasi 221005, India
e-mail: bbnanchal.rs.cer12@iitbhu.ac.in

coating technique is better for the deposition of thin films. This process is low power consumption over the physical techniques and preferential thickness controlled [11–13]. Analysis of crystalline, surface morphology, compositions and roughness of thin film by the scientific procedure of X-ray diffraction (XRD), Scanning electron microscope (SEM), Energy dispersive spectroscopy (EDX), Atomic force analysis (AFM) respectively, provided from RIGAKU DIFFRACTOMETER (Smart Lab 9 kW, Target: Cu Kα, λ = 1.54 Å), EVO-18 Research ZEISS, NT-MDT- NTEGRA. The spin coating used by equipment of DELTA SCIENTIFIC EQUIPMENT; SPIN COATER MODEL DELTA SPIN-1 and measurement of resistance using digital multimeter as FLUKE-107 600 V. This paper reported to 1000, 1200, 1400 ppm methane detecting at 125 °C. Lower explosive limit (LEL) of methane is 5%, So, we decided research below LEL due to safety purpose [14].

2   Experimental

Zinc Carbonate ($ZnCO_3·2ZnO·3H_2O$) was dissolved in the aqua regia solution and stirred at 120 °C. After 15 min Copper (II) sulphate [$CuSO_4·5H_2O$] solution was added into the solution and stirred after half an hour melamine was added and stirred till the solution turns transparent homogeneous and slightly yellow. The next step processed to spin coating. Drop casted on glass substrate, which hold on spin coater and spun at 2500 rpm for 30 s, thereafter keep on 150 °C hot plate for 5 min and it repeat 10 times. Finally, deposited substrate was kept at 400 °C for 60 min. Further silver paste using metal mask of paste method. There are used substrate as glass cuted in the size of 2.5 × 2.5 cm and cleaned in ultrasonic bath.

3   Results and Discussions

XRD analysis confirms the presence of crystallized and linear grains thin film. The peaks noted were (100), (002), (101), (102), (110), (103), (200), (112), (201) and (004) which was confirmed by JCPDS no. 36-1451 (Fig. 1). These peaks indicates zinc oxide based material. The SEM analyses confirm nano-petals present are linear and the film is dense and porous. The petals shapes were in nano and micro levels with a thickness of 44, 67, 89 nm, etc. These are suitable forchemisorptions process. These factors increased sensitivity in the presence of reducing gases (Fig. 2). EDX analysis confirms to the compositions of copper and zinc oxide on the binding energy range of 10 keV with overlapped peak (Fig. 3). Three dimensional AFM, confirmed the film to be smooth, uniform and the roughness was measured to be 300 nm (Fig. 4). The characterizations confirm that nano-petal based film is suitable in sensing of gases.

When methane flowed in the assembled sensing set up and exposed on sensor then we found that resistance decreases due to the reduction of the surface. Assembled sensing set up shown in Fig. 5 (Schematic diagram). Gas sensor recovered original
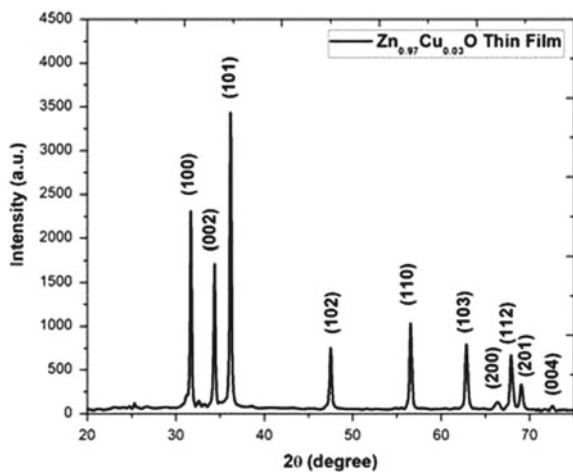
Fig. 1 XRD of nano-petals $Zn_{0.97}Cu_{0.03}O$ thin film
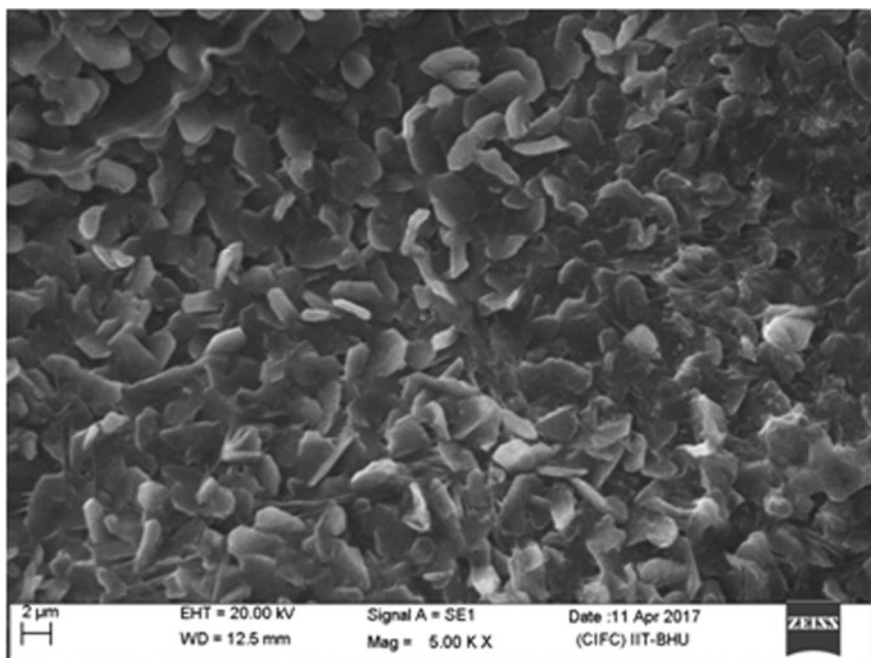
Fig. 2 SEM image of nano-petals $Zn_{0.97}Cu_{0.03}O$ thin film
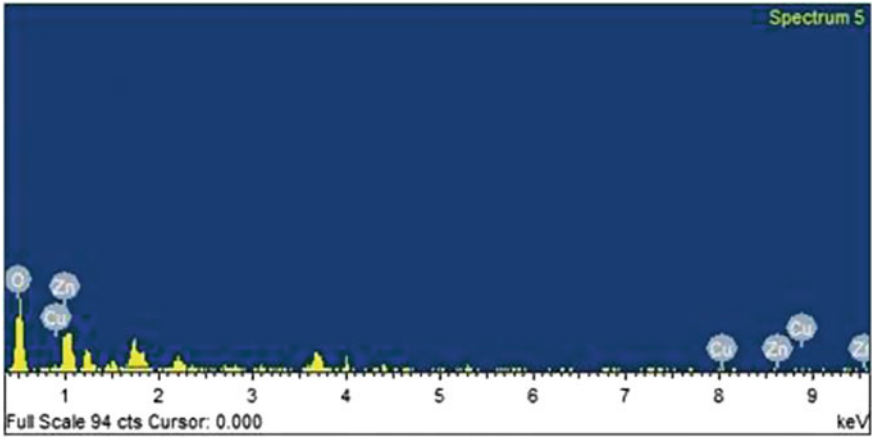
Fig. 3 EDX of nano-petals $Zn_{0.97}Cu_{0.03}O$ thin film

Fig. 4 Three-D AFM of nano-petals $Zn_{0.97}Cu_{0.03}O$ thin film
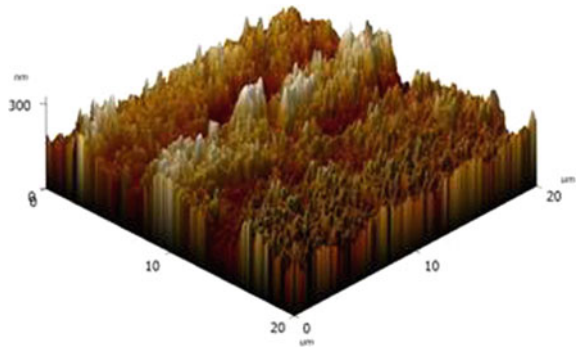
Fig. 5 Schematic diagram of assembled gas sensing set up
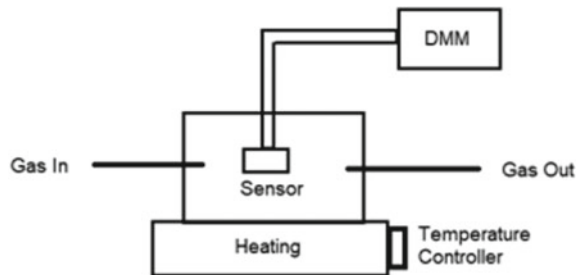
state when removed methane from sensing set up. The sensitivity determined as Eq. (1) [10, 11] are following:-

$$\text{Sensitivity (\%) } S = \left[(R_a - R_g)/R_a\right] \times 100 \tag{1}$$

where R_a is the resistance of thin film in the air and R_g in the presence of gas.

Measurement of resistance for nano-petal based thin film by multimeter in three steps. In the first step the resistance of thin film in air.was measured Similarly the resistance was measured in the presence of methane and, also resistance was measured with increasing concentrations of methane (second step). Lastly measured resistance in the condition of removal of methane from testing set up (step 3). We observed that resistance of thin film decreased and was stable in the presence of methane. When was methane removed from the testing set up, then resistance increased and recovered original resistance of thin film. So, sensitivity is the function of concentration and sensitivity increases with concentration of methane. Nano-petal based thin film sensor detected 1000–1400 ppm concentration of methane at an operating temperature of 125 °C in the experiment. Determined lowest sensitivity is 16% for 1000 ppm and highest sensitivity is 40% for 1400 ppm (Fig. 6), while average sensitivity is 26.66% for 1000–1400 ppm methane. Average response and recovery time are 33.33 s and 42.33 s respectively. Thereafter selectivity analysis in separate detection of ammonia then found that selectivity response is better for 1000–1400 ppm concentration of methane at operating temperature 125 °C. Methane reaction with oxygen species are following [9, 10].
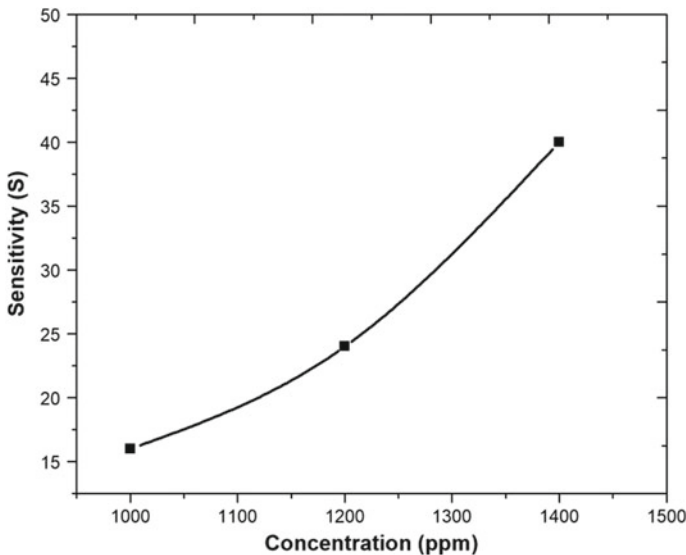
Fig. 6 Sensitivity of nano-petal $Zn_{0.97}Cu_{0.03}O$ thin film sensor

$$CH_4 + 4O^-(ads) \rightarrow CO_2 + 2H_2O(gas) + 4e^- \tag{2}$$
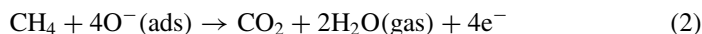
4 Conclusions

Fabrication of nano-petals of $Zn_{0.97}Cu_{0.03}O$ thin film using the technique of spin coating via chemical route. The characterizations of thin film confirmed that material is crystalline, nano-micro petals found in morphology and roughness is 300 nm. The nano structured petals in film improved to the methane sensing properties and detected to 1000–1400 ppm concentration of methane at a low operating temperature of 125 °C and selectivity with ammonia for 1000–1400 ppm. The lowest sensitivity is 16% for 1000 ppm and highest sensitivity is 40% for 1400 ppm. Spin coating process is easy and cost-effective.

References

1. Yu, A., Qian, J., Pan, H., Cui, Y., Xu, M., Tu, L., Zhou, X.: Micro-lotus constructed by Fe-doped ZnO hierarchically porous nanosheets: preparation, characterization and gas sensing property. Sens. Actuators B: Chem. **158**(1), 9–16 (2011)
2. Yin, M., Liu, M., Liu, S.: Diameter regulated ZnO nanorod synthesis and its application in gas sensor optimization. J. Alloy. Compd. **586**, 436–440 (2014)
3. Katoch, A., Choi, S.W., Kim, H.W., Kim, S.S.: Highly sensitive and selective H_2 sensing by ZnO nanofibers and the underlying sensing mechanism. J. Hazard. Mater. **286**, 229–235 (2015)
4. Wang, L., Kang, Y., Liu, X., Zhang, S., Huang, W., Wang, S.: ZnO nanorod gas sensor for ethanol detection. Sens. Actuators B: Chem. **162**(1), 237–243 (2012)
5. Hassan, J.J., Mahdi, M.A., Chin, C.W., Abu-Hassan, H., Hassan, Z.: A high-sensitivity room-temperature hydrogen gas sensor based on oblique and vertical ZnO nanorod arrays. Sens. Actuators B: Chem. **176**, 360–367 (2013)
6. Wei, A., Pan, L., Huang, W.: Recent progress in the ZnO nanostructure-based sensors. Mater. Sci. Eng. B **176**(18), 1409–1421 (2011)
7. Gruber, D., Kraus, F., Müller, J.: A novel gas sensor design based on $CH_4/H_2/H_2O$ plasma etched ZnO thin films. Sens. Actuators B: Chem. **92**(1–2), 81–89 (2003)
8. Takata, M., Tsubone, D., Yanagida, H.: Dependence of electrical conductivity of ZnO on degree of sintering. J. Am. Ceram. Soc. **59**(1–2), 4–8 (1976)
9. Motaung, D.E., Mhlongo, G.H., Kortidis, I., Nkosi, S.S., Malgas, G.F., Mwakikunga, B.W., Kiriakidis, G.: Structural and optical properties of ZnO nanostructures grown by aerosol spray pyrolysis: candidates for room temperature methane and hydrogen gas sensing. Appl. Surf. Sci. **279**, 142–149 (2013)

10. Choudhary, M., Mishra, V.N., Dwivedi, R.: Effect of temperature on palladium-doped tin oxide (SnO_2) thick film gas sensor. Adv. Sci. Eng. Med. **5**(9), 932–936 (2013)
11. Rambu, A.P., Doroftei, C., Ursu, L., Iacomi, F.: Structure and gas sensing properties of nanocrystalline Fe-doped ZnO films prepared by spin coating method. J. Mater. Sci. **48**(12), 4305–4312 (2013)
12. Gao, F., Liu, X.Y., Zheng, L.Y., Li, M.X., Bai, Y.M., Xie, J.: Microstructure and optical properties of Fe-doped ZnO thin films prepared by DC magnetron sputtering. J. Cryst. Growth **371**, 126–129 (2013)
13. Anchal, B.B.N., Singh, P., Pyare, R.: A nano-wrinkled $Zn_{0.92}Fe_{0.08}O$ thin film developed using a high-RPM electro-spin patterning technique via Sol-Gel route for methane sensing. ChemistrySelect **3**(42), 11881–11889 (2018)
14. No, E.E.S.: Best Practice Guidance for Effective Methane Drainage and Use in Coal Mines (2010)

Study and Analysis of Low Power Dynamic Comparator

Ritesh Kumar Kushwaha, Prem Kumar and P. Karuppanan

Abstract Nowadays, everything will be digital for ease of processing, but the real world is in analog. Various techniques has been used to convert the analog signal to digital, and the fastest analog-to-digital converter (ADC) is Flash type ADC among all. It requires a high-speed comparator, which compares the analog input signals and produces the digital output through a predetermined threshold voltage accordingly. However, high-speed comparator increases the transistor and hence correspondingly increases large area, supply voltage, power, etc. Also, the system with low power dissipation speeds reduction for many transistors with a high offset voltage. Therefore, designing an ADC system that requires less power with faster operation is a great concern. Diverse kinds of comparators are available in the present scenario. In this paper, it has analyzed Conventional dynamic comparator, Double tail dynamic comparator, Doubletail comparator with enhanced latch regeneration and proposed comparator. The projected comparator with enhanced latch regeneration speed is better than the previous two conventional dynamic comparators in terms of power and speed.

Keywords VLSI design · Dynamic comparator · Power · Speed

1 Introduction

In today's world, battery-operated devices demands are increasing rapidly that gives major importance toward low power approaches for high-speed applications. With the rapid growth of microchip technology engineering typically portable electronic

R. K. Kushwaha · P. Kumar · P. Karuppanan (✉)
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India
e-mail: karuppanan1982@gmail.com

R. K. Kushwaha
e-mail: riteshkushwaha04@gmail.com

P. Kumar
e-mail: prem.8293@gmail.com

systems, which is wireless communication devices, battery-powered medical devices, etc. In all these applications, battery lives are one of the most important factors. Therefore, the need for low power and low voltages is essential. IC's cost can also reduce by using thicker oxide layer or higher supply voltages Normally, In the IC's where analog-to-digital converter or digital-to-analog converter plays an important role, low power and low voltage comparator are an essential and necessary thing. A comparator is an important part of ADC and so consumes large power in the device. It is an essential or primary concern to achieve low power and the high-speed comparator. Therefore, reducing Kickback noise enhanced the performance of the comparator. They generate it because of the output signal to the input. The feedback mechanism is responsible for backflow, which includes both positive and negative feedback [1–3].

In a recent article [4], enhanced comparator circuit was proposed by including additional circuitry and using 0.8 V a supply boosted comparator, the supply voltage and consuming power was about 15.54 μW. The dynamic comparator structure contains two tail transistors that make its operation faster. Similarly, in the article [2], the double tail dynamic comparator is designed. This comparator contains two tails which give enough current for faster decision-making. In a supply boosted comparator [5], two differential amplifier was used. One differential amplifier was used only for boosting techniques, but another differential amplifier has been used for regular comparator circuits. The resolution and gain of this comparator have improved regarding the previous one. Latch comparator offset is analyzed [2] in terms of its load capacitor mismatch. The offset is a very critical parameter in comparator design. If mismatch load is used then the offset voltage of the comparator is high. The supply boosting technique [6] is also the common choice in comparator design. It is more suitable for a low clock and a low power comparator circuit. In this comparator, a charge pump circuit is used. This charge pump circuit boosted the clock during the comparison phase. Because of this comparator circuits become faster than the previous conventional dynamic and the double tail dynamic comparator.

In this paper, Conventional dynamic comparator, double tail dynamic comparator, and dynamic comparator have been studied and analyzed with enhanced latch regeneration speed. The paper is organized as follows: In Sect. 2, clock-based conventional dynamic comparator design has been explained. In Sect. 3, simulation results and discussion have been done. Finally, the conclusion is given in Sect. 4.

2 Clocked Based Dynamic Comparator

In this section, different procedure of the conventional and double tail, dynamic comparators have been discussed.

2.1 Clocked Based Conventional Dynamic Comparator

This section briefly describes the delay and power consumption of different dynamic comparators.

In Fig. 1, the circuit design of conventional dynamic comparator is revealed. This comparator static power consumption is zero and if clk = 0 then tail transistor M_{tail} is "off". Transistors (M_7–M_8) is reset by temporary switched, Output voltages (O_{utn}, O_{utp}) to V_{DD}. Likewise, when clk = V_{DD}, M_{tail} is "on" and transistors (M_7, M_8) are "off". In this condition, O_{utp} and O_{utn}, started discharging by dissimilar discharging rates, corresponding to the input voltages (INN and INP). Let, $V_{INN} < V_{INP}$; then O_{utn} discharges lesser than O_{utp}, hence O_{utp} drops to $V_{DD} - |V_{thp}|$ earlier than O_{utn}. As a result, transistor M_5 (PMOS) will turn "on" that leads the latch regeneration due to end to end inverters (M_3, M_5 and M_4, M_6). Also, when for $V_{INN} > V_{INP}$; circuit operation is vice versa.

The conventional dynamic comparator has two types of delay such as discharge delay (t_0) and latching delay (t_{latch}). The delay t_0 occurs owing to the capacitive discharge of the load capacitance (C_L) until the time required to the transistor (M_5/M_6) turns "on". For, $I_{NP} > I_{NN}$ the drain current (I_2) of transistor M_2 gets discharged more rapidly of O_{utp} node than O_{utn} node. The O_{utn} node is driven by M_1 having a lesser drain current. Subsequently, the t_0 is derived from:

Fig. 1 Single tail dynamic comparator
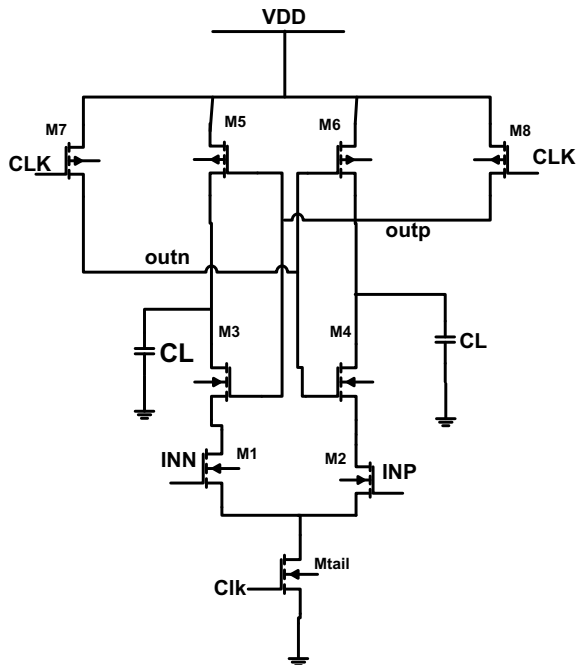
$$t_o = \frac{C_L |V_{thp}|}{I_2} \tag{2.1}$$

where t_o is the delay due to all transistor except latch, C_L is the load capacitance, V_{thp} is threshold voltage of PMOS transistor and I_2 is Current which is contributing to tail current.

The latching delay (t_{latch}) is the delay due to cross-coupled inverters in the circuit and given by [7].

$$t_{latch} = \frac{C_L}{gm_{eff}} * \ln\left(\frac{\Delta V_{out}}{\Delta V_0}\right) \tag{2.2}$$

$$t_{delay} = t_0 + t_{latch} \tag{2.3}$$

where t_{delay} is a total delay, t_o is the delay because of all transistor except latch, t_{latch} is a delay because of latch, C_L is a load capacitance, gm_{eff} is effective transconductance, ΔV_{out} is output voltage difference. The ΔV_0 is the measured output voltage difference at the dropping output. Hence, load capacitance (C_L) is directly proportional, input difference voltage (V_{in}) is indirectly proportional to the total delay (t_{delay}) of the comparator. Therefore, the decreased value of V_{cm}, increases the delay (t_0) for slighter bias current (I_{tail}) henceforth, the increased value of initial voltage difference (V_0) is obtained that decreasing t_{latch}.

For fast speed and switching operation, V_{cm} requires only 70% V_{in} [7]. Also, because of several stacked transistors, a sufficiently high supply voltage is needed for a proper delay time because of transistors M_3 and M_4 initially contributed for positive feedback until transistors M_5 or M_6 get started. If the supply voltage is low, then a very small amount of gate-source voltage (transconductance) is produced, which creates a larger latch delay time. To obtained enhanced Gm/I ratio and long integrated interval, lower tail current is desired to retain the differential couple in weak inversion [8, 9]. Contrary, for the fast generation, requires a large tail current. Moreover, M_{tail} drives normally in triode region hence tail current depends on V_{cm} and it is not suitable for regeneration. Another disadvantage of such a circuit is that the tail transistor (M_{tail}) has only one current path for the latch and the differential amplifier.

The power of the comparator is given by

$$P_{total} = P_{static} + P_{dynamic} \tag{2.4}$$

$$P_{static} = I_{leakage} * V_{dd} \tag{2.5}$$

where $I_{leakage}$ is leakage current of transistor and V_{dd} supply voltage

$$P_{dynamic} = C_L * V_{dd}^2 * f \tag{2.6}$$

where C_L load capacitance and f is the frequency of the pulse.

2.2 Clocked Based Double Tail Dynamic Comparator

In Fig. 2, a double tail comparator is shown. This comparator requires very low supply voltage compared to the older conventional dynamic comparator due to less stacking. The advantage of the double tail is to produce huge current during latching and immediate latching, also it does not depend on V_{cm} and for low offset, a very less current at the input stage [4]. When clk = zero, transistor M_{tail1}/M_{tail2} are "off" hence M_3 and M_4 begins to charge node fp, fn to V_{DD} and transistor M_{R2}, M_{R1} to ground. On the other hand when clk = V_{DD} both tail transistors (M_{tail1}/M_{tail2}) get "on" hence M_3 and M_4 turns "off" which leads to discharge the voltage at node fp, fn at the rate of $IM_{tail1}/Cfn(p)$ hence intermediate differential voltage $V_{fn(p)}$ is established which offer protection among input/output, so, reduces the noise [2]. It can be easily seen in the Fig. 2, which in evaluation phase O_{utp}, O_{utn} both node discharge and phase gets started. Therefore, any phase O_{utn} or O_{utp} energies to high and alternative remain in the same stage. The time delay is analyzed by following equations:

$$t_{delay} = t_0 + t_{latch} \tag{2.7}$$

Fig. 2 Dynamic comparator (double tail)
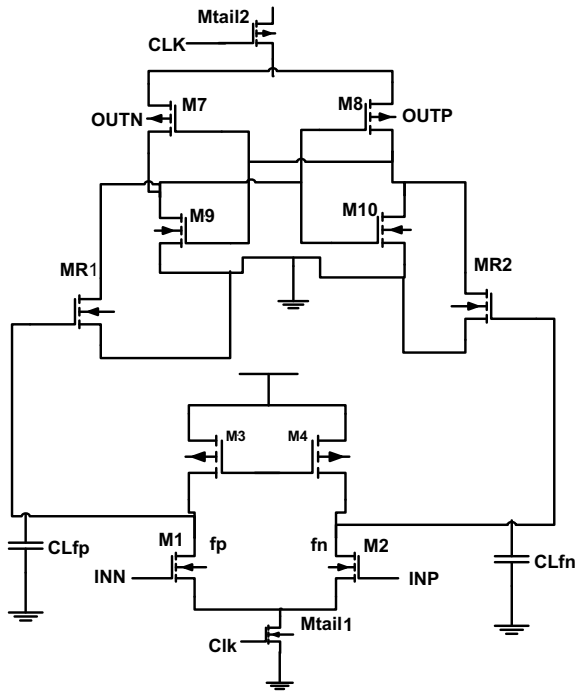
$$t_{latch} = \frac{C_L}{gm_{eff}} * \ln\left(\frac{V_{DD}/2}{\Delta V_O}\right) \tag{2.8}$$

$$t_o = \frac{C_L|V_{thn}|}{I_2} \tag{2.9}$$

where t_{delay} is the total delay, t_o is the delay due to all transistor except latch, t_{latch} is the delay due to latch, C_L is the load capacitance, gm_{eff} is the effective transconductance, t_o is the delay due to all transistor except latch and I_2 is the current which is contributing to tail current.

2.3 Double Tail Dynamic Comparator with Enhanced Latch Regeneration Speed

Figure 3, the modified circuit diagram of the comparator is shown. When clk = zero, transistor M_{tail1}, M_{tail2} are "off" hence M_3 and M_4 begins to charge node fp, fn to V_{DD} and control transistor (M_{c1}, M_{c2}) turn "off". When clk = V_{DD}, then each tail transistor M_{tail1}, M_{tail2} turns "on", and so transistors M_3, M_4 flip "off". But still, transistors M_{c1}, M_{c2} are at "off" condition due to node fn, fp at V_{DD}. Also, the voltage at fn, fp begin to drop with dissimilar rates in keeping with the input voltages. In case of $V_{INN} < V_{INP}$, 'fp' decreases and become lesser than 'fn'.

When node fn remains decreasing, then respective PMOS transistor M_{c1} flip to "on", and simultaneously fp node reaches to V_{DD}. So every other M_{c2} stays "off", permitting fn to be absolutely discharged. The transconductance characteristics of comparators are mentioned in [5]. It is worth notices that static power consumption is observed due to transistor M_{c1} is activated. To conquer this problem, modified circuitry have been suggested as shown in Fig. 3, the switches of NMOS are added with Msw1 and Msw2.

The nodes of fn and fp were excited by the 'VDD' while reset operation, and fn, fp begins to fall with exclusive discharging rates so, comparator identifies quicker discharging node. The operation of the latch is emulated by control transistor switches.

The Delay is analyzed by the following equations:

$$t_{delay} = t_0 + t_{latch} \tag{2.10}$$

$$t_{latch} = \frac{C_L}{gm_{eff} + gm_{r1,2}} * \ln\left(\frac{V_{DD}/2}{\Delta V_0}\right) \tag{2.11}$$

$$t_o = 2 * \frac{V_{thn} \times C_{Lout}}{I_{tail2}} \tag{2.12}$$

where t_{delay} is the total delay, t_o is the delay because of all transistor except latch, t_{latch} is the delay due to latch, C_L is the load capacitance, gm_{eff} effective transconductance,
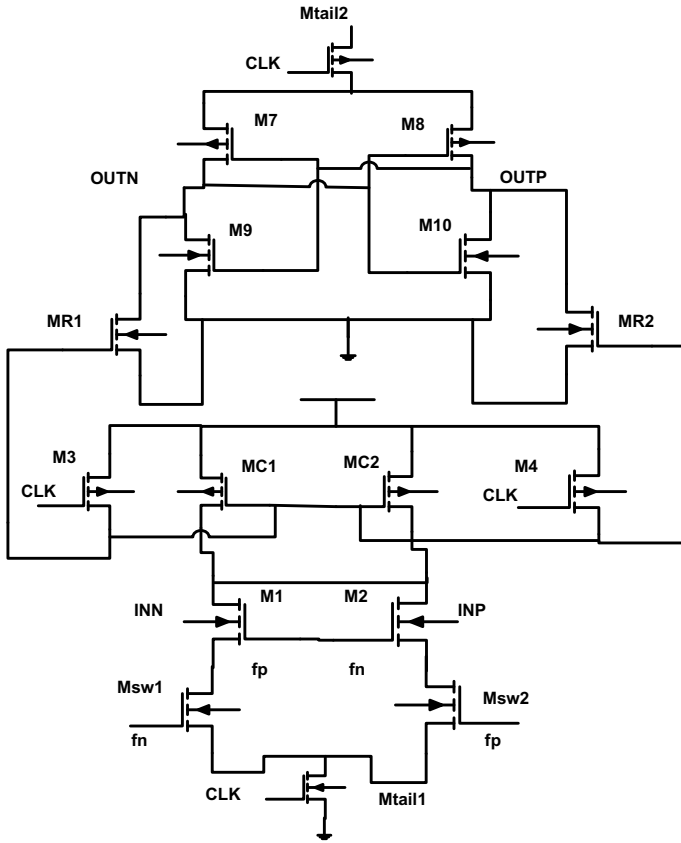
Fig. 3 Double tail dynamic comparator with enhanced latch regeneration speed

$gm_{r1,2}$ is the latch transconductance, V_{thn} is NMOS threshold voltage, and I_2 is the current which is contributing to the tail current.

2.4 Proposed Comparator

Recently, many technique have been suggested to moderates the power consumption efficiently for all the transistor. The subthreshold current can be reduced considerably by decreasing the V_{DS}. The V_{DS} can be reduced by increasing source terminal voltage hence, additional transistors labeled PMOS have been used to enhance the source terminal voltage of MOSFET. As the additional PMOS transistor is at the cutoff position, it produces leakage currents, consequently source voltage increases and due to this V_{DS} decreases. In Fig. 4, four extra PMOS transistor have been used and due to this, the power of the circuit decreases up to some extent.
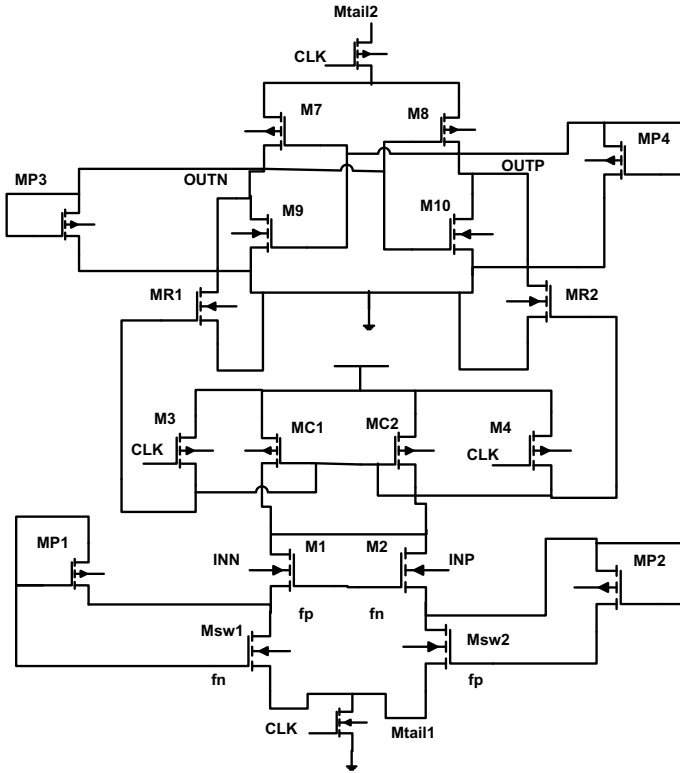
Fig. 4 Proposed comparator circuit

When clk = zero, transistor M_{tail1}, M_{tail2} are "off", hence M_3 and M_4 begins to charge node fp, fn to V_{DD} and control transistor (M_{c1}, M_{c2}) turn "off". When clk = V_{DD}, then each tail transistor M_{tail1}, M_{tail2} turns "on", and so transistors M_3, M_4 flip "off". But still, transistors M_{c1}, M_{c2} are at "off" condition due to node fn, fp at V_{DD}. Also, the voltage at fn, fp begin to drop with dissimilar rates in keeping with the input voltages. Let us anticipate a case while $V_{INN} < V_{INP}$, therefore fp decreases lesser then fn. When node fn remains decreasing, then respective PMOS transistor M_{c1} flip to "on", and simultaneously fp node reaches to V_{DD}. So every other M_{c2} stays "off" permitting fn to be absolutely discharged.

3 Simulation Results and Discussion

For simulation purpose, the cadence virtuoso gpdk 180 nm technology has been used. Firstly the circuit is designed, and then the delay and the power of each comparator has been analyzed one by one.
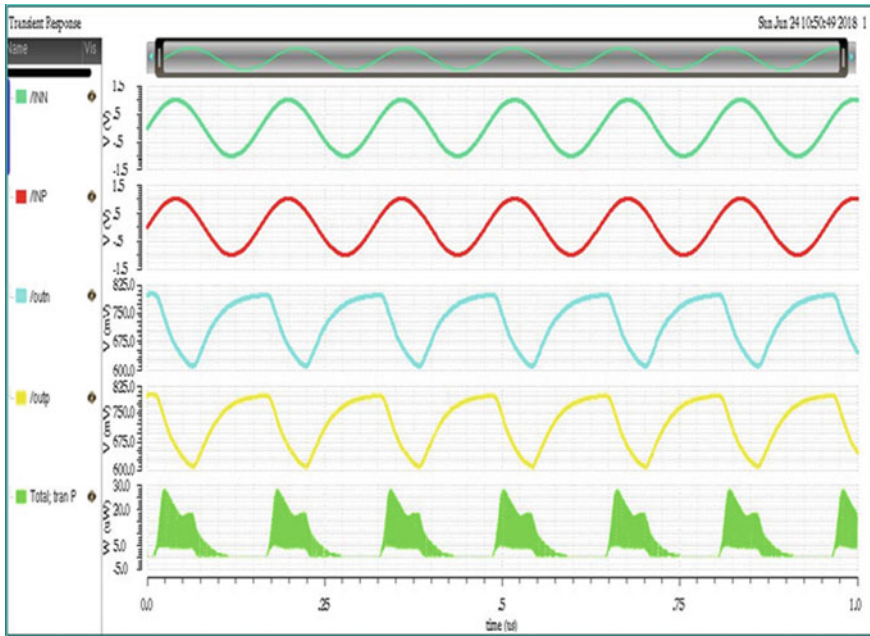
Fig. 5 Simulation of conventional dynamic comparator

Conventional dynamic comparator circuit simulation is presented in Fig. 5. The sinusoidal input of 702.5 mV and 607.5 mV to differential input has been given. Here, the 1 pF capacitor is used at the output node and both outputs are analyzed in terms of their respective input signals. From the graph, it can be easily seen that the charging and discharging behavior of output voltage according to supplied input. Transient power is also analyzed in this circuit so it can get a brief analysis of the conventional dynamic comparator circuits.

Double tail dynamic comparator circuit simulation is presented in Fig. 6. The sinusoidal input of 702.5 mV and 607.5 mV to differential input has been given. Here, a 1 pF capacitor is used at the output node and both outputs are analyzed in terms of their respective input signal. Here, it can see from the graph that outputs are charging and discharging according to supplied input. Output plot is more accurate than the conventional dynamic comparator. This due to two tail present in this circuit. The extra tail provides enough current for faster decision-making in the comparator circuit. The transient power is also analyzed in this circuit so that it can get a brief analysis of the conventional dynamic comparator circuit.

Double tail dynamic comparator circuit simulation is presented in Fig. 7 for DC input. The sinusoidal input of 702.5 mV and 500 mV as a reference voltage to differential inputs has been given. Here, the 1 pF capacitor is used at the output node. The supplied constant voltage source to one of its inputs, hence it can get a better understanding of output response. Here both outputs are analyzed in terms of their
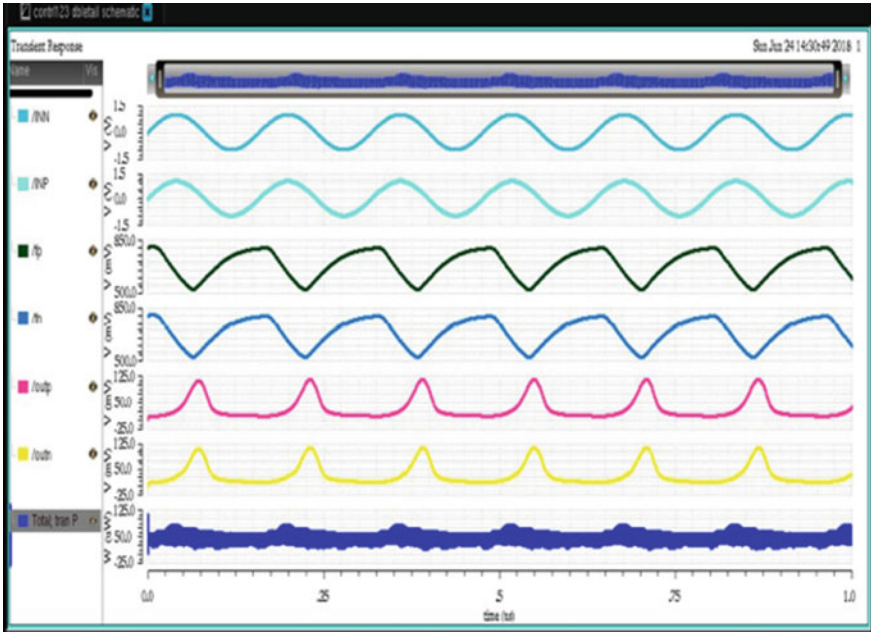
Fig. 6 Double tail dynamic comparator
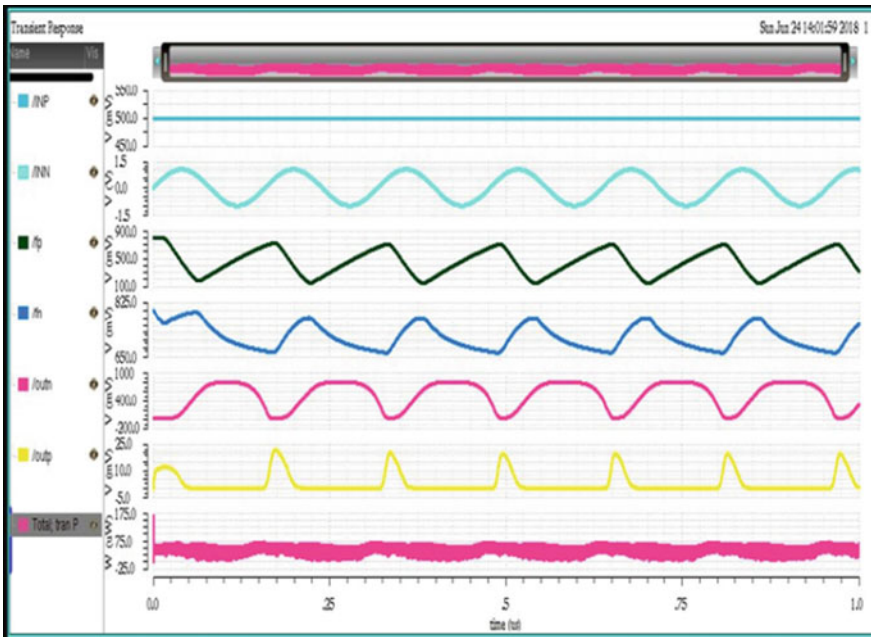
Fig. 7 Double tail comparator when one of it's input is d.c. voltage
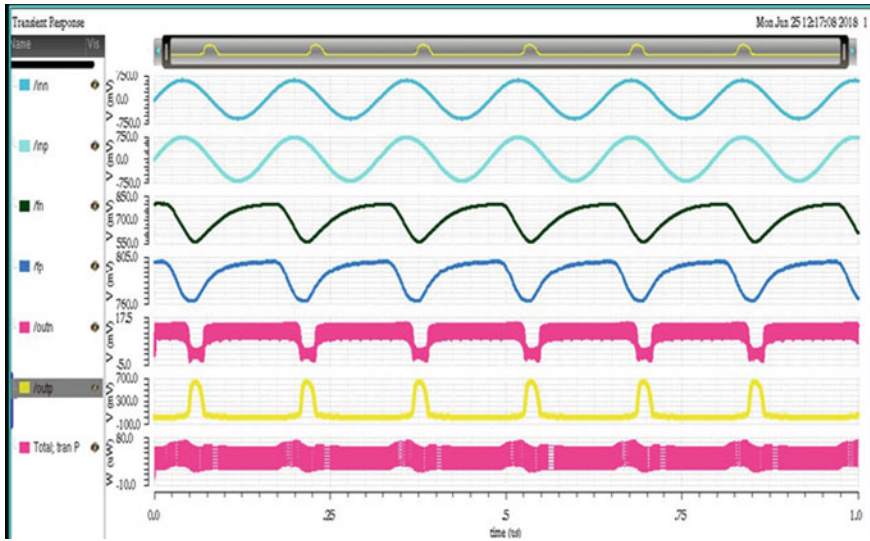
Fig. 8 Double tail comparator with enhanced latch regeneration speed

respective input signals. Here, it can be seen from the graph that both the outputs are charges and discharges according to supplied input.

Double tail dynamic comparator with enhanced latch regeneration speed circuit simulation is presented in Fig. 8. The supplied sinusoidal input of 702.5 mV and 607.5 mV to differential input. Here, the 1 pF capacitor is used at the output node. Here both outputs are analyzed in terms of their respective input signals. In this comparator, two latch has been used, due to this decision-making operation becoming faster. As seen from the corresponding output graph. Here, it can be seen from the graph that both the outputs are charges and discharges according to supplied inputs. The output plots are more accurate than double tail dynamic comparator because two tail and two latches are present in this circuit. Extra latch provides enough current for faster decision-making in the comparator circuit.

Dynamic comparator with enhanced latch regeneration speed circuit simulation is presented in Fig. 9 for d.c. input, The sinusoidal input of 702.5 mV and 500 mV constant d.c. voltage to differential input is provided. Here, the 1 pf capacitor at the output node is used and both outputs are analyzed in terms of their respective input signals. The constant voltage source is supplied to its any input, so that gets a better understanding of output response. In this comparator, two latch has been used, due to this decision-making operation becoming faster. Here, it can be seen from the graph that both output charges and discharges according to supplied input. It can be observed that the output plot is more accurate than the double tail dynamic comparator. This comparator uses two latch circuit, hence this decision-making operation becomes faster.
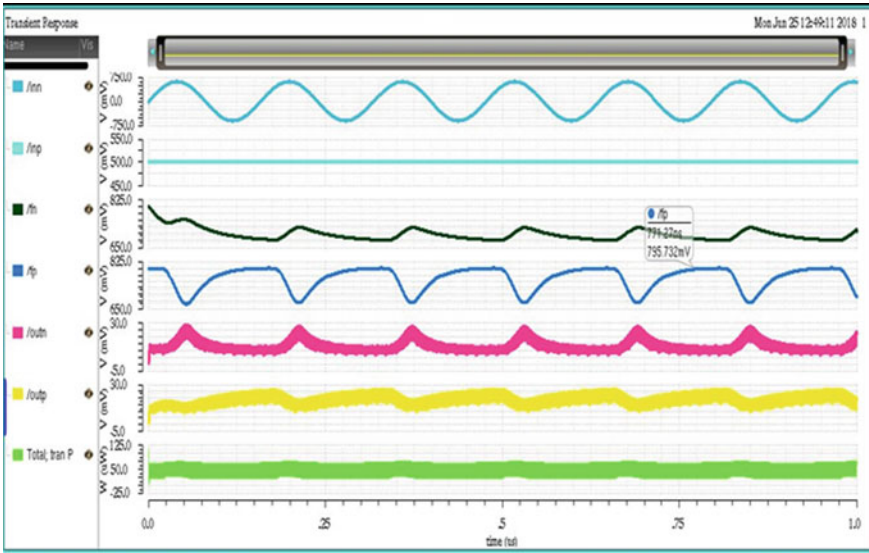
Fig. 9 Dynamic comparator with enhanced latch regeneration speed when one of its input is d.c. voltage
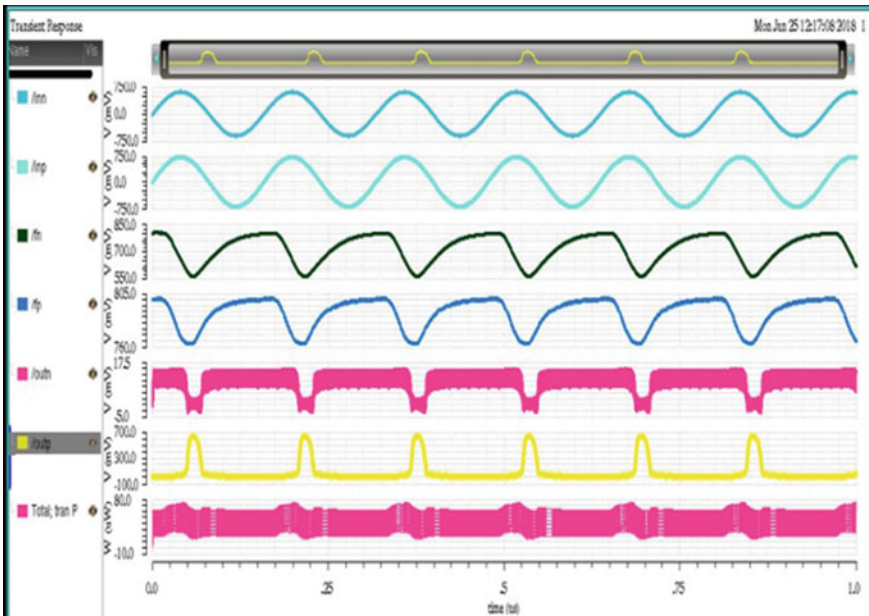
Fig. 10 Proposed comparator when sinusoidal input is applied
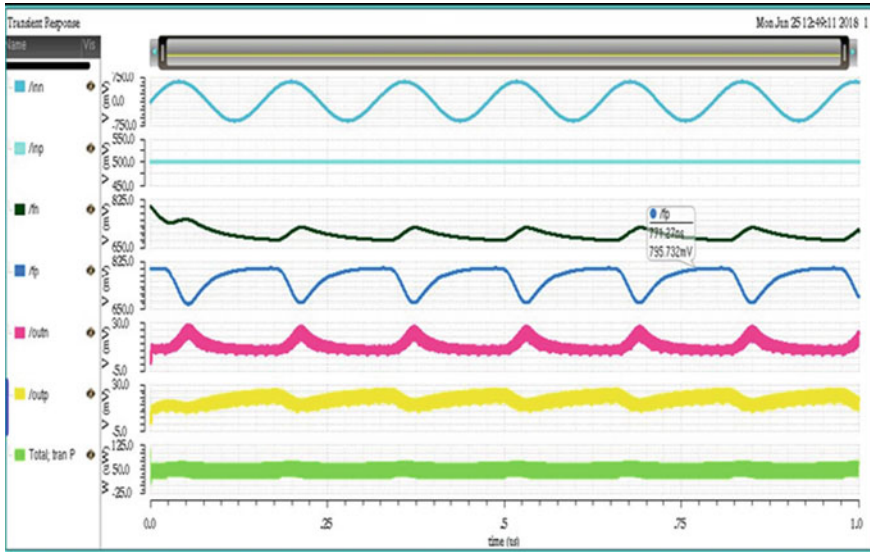
Fig. 11 Proposed dynamic comparator circuit when one of its input is d.c. voltage

Proposed dynamic comparator circuit simulation is presented in Fig. 10. The feed sinusoidal input of 702.5 mV and 607.5 mV to the differential input. Here, the 1 pF capacitor is used at the output node. The proposed circuit is power efficient, but the simulation output is similar to enhanced latch regeneration speed comparator. In this comparator, two latch has been used, due to this decision-making operation becoming faster. In the graph, it is clearly seen from the corresponding outputs. The output charges and discharges according to supplied inputs. Hence, the output plot is more accurate than double tail dynamic comparator owing to two tails and latches present in this circuit. The extra latches provide enough current for faster decision-making in the comparator circuit.

Proposed dynamic comparator circuit simulation is presented in Fig. 11 for d.c. input. The assumed sinusoidal input of 702.5 mV and 500 mV constant d.c. voltage to the differential input. Here, the 1 pF capacitor is used at the output node. Here both outputs are analyzed in terms of their respective input signal node. In this Simulation Output is the same as enhanced latch regeneration speed comparator. But this proposed circuit is power efficient. The supplied voltage source is constant to one of its input so that it can get a better understanding of output response. In this comparator two latch has been used, since this decision-making operation becomes faster.

It can be seen from the corresponding output graph that output plot is more accurate than the double tail dynamic comparator because of two tail and two latches are present in this circuit. Extralatch provides enough current for faster decision-making in the comparator circuit. Transient power is also analyzed in this circuit so

Table 1 Performance comparison

Design specification	Conventional dynamic comparator	Double tail dynamic comparator	Double tail comparator with enhanced latch regeneration	Proposed comparator
Technology	180 nm	180 nm	180 nm	180 nm
Supply	0.8 V	0.8 V	0.8 V	0.8 V
Slew rate	1.3 kV/μs	3.25 kV/μs	7.61 kV/μs	7.25 kV/μs
Rise time (clock)	0.08 ns	0.08 ns	0.08 ns	0.08 ns
Fall time (clock)	0.08 ns	0.08 ns	0.08 ns	0.08 ns
Delay	13.8 ns	5 μs	4.2 μs	4.26 μs
Power consumption	15.54 μw	29.21 μw	28.23 μw	24.64 μw
Max. sampling frequency	800 MHz	1.6 GHz	2.2 GHz	2.24 GHz

that it can get a brief analysis of dynamic comparator circuit with enhanced latch regeneration speed.

In, double tail dynamic comparator with enhanced latch regeneration speed circuit simulation the given sinusoidal input of 702.5 mV and 607.5 mV to the differential input. Here, two latches and the 1 pF capacitor is used at the output node hence the decision-making operation becomes faster. It can be seen from the corresponding output graph. The output plot is more accurate than the double tail dynamic comparator because two tails and two latches are present in this circuit. Extra latches provide enough current for faster decision-making in the comparator circuit. Transient power is also analyzed in this circuit so it can get a brief analysis of dynamic comparator circuit with enhanced latch regeneration speed.

The performance comparison is illustrated in Table 1. This table gives a detailed analysis of all the comparator circuits in terms of slew rate, rise time, fall time, delay and power consumption, and maximum sampling frequency.

4 Conclusion

In this paper, comprehensive delay analysis for various varieties of clocked dynamic comparators and expressions were derived and conferred. Various structures of conventional dynamic comparator such as conventional double tail dynamic comparators, and the double tail dynamic comparator with enhanced latch regeneration and proposed comparator have been analyzed in terms of slew rate, rise and fall times, power consumption, delay, and maximum sampling frequency. The simulated results confirmed that the projected comparator have significantly reduced the delay and the

power consumption as compared to the double tail dynamic comparator and Double tail comparator with enhanced latch regeneration. All comparators have been designed and simulated in cadence virtuoso.

References

1. Goll, B., Zimmermann, H.: A comparator with reduced delay time in 65-nm CMOS for supply voltages down to 0.65 V. IEEE Trans. Circuits Syst. II Express Briefs **56**(11), 810–814 (2009)
2. Nikoozadeh, A., Murmann, B.: An analysis of latch comparator offset due to load capacitor mismatch. IEEE Trans. Circuits Syst. II Express Briefs **53**(12), 1398–1402 (2006)
3. Nath Mandal, D., et al.: Analysis and design of low voltage low power dynamic comparator with reduced delay and power. Int. J. Eng. Res. Gen. Sci. **2** (2014)
4. Babayan-Mashhadi, S., Lotfi, R.: Analysis and design of a low-voltage low-power double-tail comparator. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **22**(2), 343–352 (2014)
5. Ay, S.U.: A sub-1 volt 10-bit supply boosted SAR ADC design in standard CMOS. Analog Integr. Circuits Signal Process. **66**(2), 213–221 (2011)
6. Mesgarani, A., et al.: Supply boosting technique for designing very low-voltage mixed-signal circuits in standard CMOS. In: 2010 53rd IEEE International Midwest Symposium on Circuits and Systems. IEEE (2010)
7. Ay, S.U.: A sub-1 volt 10-bit supply boosted SAR ADC design in standard CMOS. Analog Integr. Circuits Signal Process **66**(2), 213–221 (2011)
8. Blalock, B.J., et al.: Body-driving as a low-voltage analog design technique for CMOS technology. In: 2000 Southwest Symposium on Mixed-Signal Design (Cat. No. 00EX390). IEEE (2000)
9. Maymandi-Nejad, M., Sachdev, M.: 1-bit quantiser with rail to rail input range for sub-1 V $\Delta\Sigma$ modulators. Electron. Lett. **39**(12), 894–895 (2003)

Tuned Universal Filter Design Using Single Differential Difference Current Conveyor for Sub-GHz Frequency Band

**Shalini Mishra, Devarshi Shukla, Vijaya Bhaduaria
and Santosh Kumar Gupta**

Abstract In this paper, the proposed configuration is realizing High-Pass, Band-Pass, Low-Pass, and Notch filters simultaneously using a single Differential Difference Current Conveyor (DDCC) as an active block. First, the circuit is implemented using passive resistors and then, in proposed configuration, passive resistors are replaced by active resistors. Proposed configuration occupied less area, provide better noise immunity, and offer programmability at the cost of slight increase in Total Harmonic Distortion. The circuit is simulated in cadence virtuoso 180 nm technology and checked for programmability. It is found that cutoff frequency varied from 13.9 to 28.34 MHz by varying the gate voltage from 1.1 to 1.5 V of the active resistance.

Keywords DDCC · THD · Universal filter

1 Introduction

Traditionally, Operational Amplifiers (op-amps) were used to design various filters. The use of op-amps has various disadvantages including limited bandwidth, limited slew rate, and complex circuitry. This led to discovery of many active blocks [1]. One such active block is Differential Difference Current Conveyor (DDCC) [2–7]. DDCC active block overcomes the disadvantages of op-amps.

Introduced in 1996, DDCC binds advantages of Differential Difference Amplifier (DDA) [8], like high-input impedance, low-output impedance, and capability of

S. Mishra
Qualcomm, Bengaluru, India
e-mail: shalmish@qti.qualcomm.com

D. Shukla (✉) · V. Bhaduaria · S. K. Gupta
Department of Electronics and Communication Engineering, MNNIT Allahabad, Allahabad, India
e-mail: devarshinitrr@gmail.com

V. Bhaduaria
e-mail: vijaya@mnnit.ac.in

S. K. Gupta
e-mail: skg@mnnit.ac.in

performing arithmetic operations, and Second Generation Current Conveyor (CCII) [9] like high gain and bandwidth. Second Generation Current Conveyor (CCII) has various applications like the design of different types of Filters, Amplifiers, and Oscillators. CCII offers versatile performance, low-power consumption, and high bandwidth. The major disadvantage of CCII was single input voltage terminal. Due to this, differential mode signals could not be applied to CCII. Two or more CCIIs had to be used for applications requiring differential mode signals. The shortcoming of CCII was overcome by incorporating DDA block as input stage of CCII to form DDCC active block.

Generally, passive components are used along with active blocks to design filters. A similar circuit was implemented using a single DDCC active block, three passive resistors, and two capacitors to implement a universal filter [10]. In the proposed configuration, passive resistors are replaced by active resistors. These active resistors are implemented using NMOS transistors in triode region. The use of active resistors offers various advantages as shall be discussed in the results section.

Section 2 of this paper covers the proposed configuration for the design of universal filter using DDCC and active resistors. Section 3 covers the results obtained after performing AC, noise, and Total Harmonic Distortion analyses. Finally, Sect. 4 concludes the work done.

2 Methodology

The DDCC active block used is an eight-terminal active block as shown in the Fig. 1. Three voltage input terminals namely V_{Y1}, V_{Y2}, and V_{Y3} are present. One current input terminal is present namely V_X. There are four current output terminals namely

Fig. 1 Symbol of DDCC active block
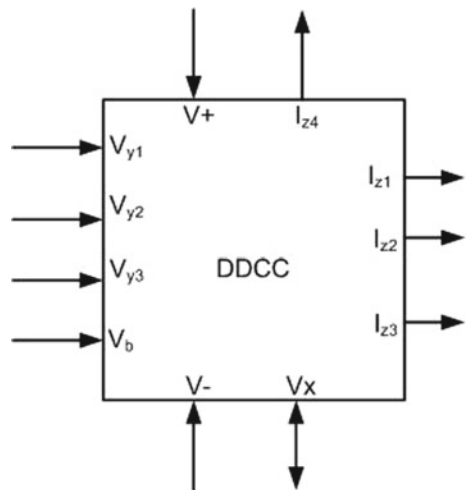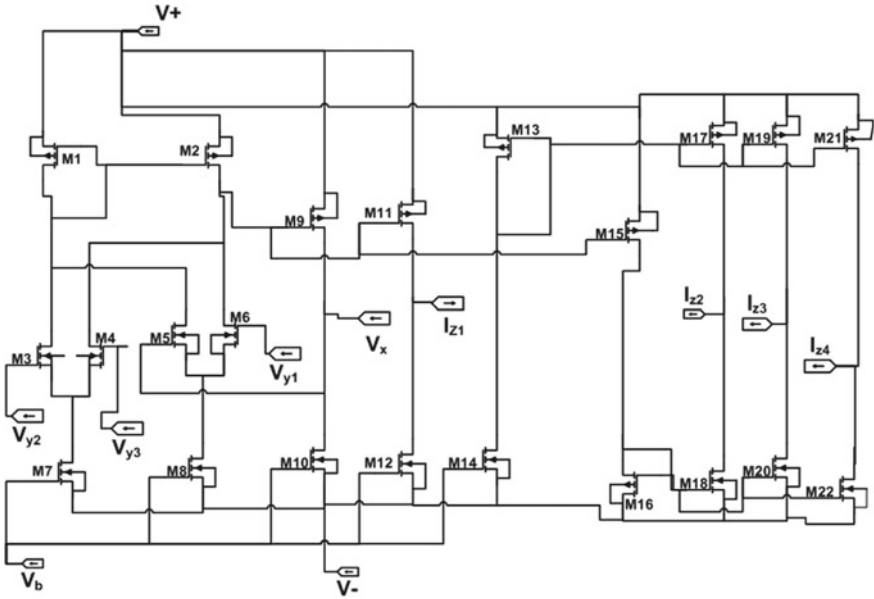
Fig. 2 CMOS realization of DDCC [10]

I_{Z1}, I_{Z2}, I_{Z3}, and I_{Z4}. Among these, I_{Z1} is non-inverting whereas I_{Z2}, I_{Z3}, and I_{Z4} are inverting terminals. All current output terminals follow the current through input current terminal V_X. For the non-inverting terminal, direction of current is same as that of input current terminal but for inverting terminals, direction of current is opposite to that of input current terminal (Fig. 2).

The input stage of DDCC is a DDA block. In this block, transistors M1 and M2 form the current mirror load and the transistors M3, M4 and M5, M6 form the differential pairs. Thus, M3, M4, M5, and M6 amplify the differential difference voltages. It is assumed that the transistors M1 and M2 are perfectly matched. This forces the current through these transistors to be equal. Applying KCL at source to drain nodes of M1 and M2, sum of currents through transistors M3 and M5 equals the sum of currents through transistors M4 and M6. Thus

$$I_1 = I_2 = I_3 + I_5 = I_4 + I_6; \tag{1}$$

This implies

$$I_5 = I_6 - I_3 + I_4; \tag{2}$$

Since the differential output currents of the differential pairs depend on gate voltages of respective transistors, the following expression is obtained:

$$V_X = V_{y1} - V_{y2} + V_{y3}; \tag{3}$$

The output stage of DDCC is current conveyor circuit. In the output stage, the current through terminal V_X is conveyed to the non-inverting terminal I_{Z1} using transistors M9, M11, and M12. Current mirrors (using transistors M13, and M17, M16, and M18) are used further to invert the direction of current from I_{Z1} to I_{Z2}, I_{Z3}, and I_{Z4}. Thus, both inverting and non-inverting current conveyor is obtained in the same circuit.

3 Proposed Configuration

DDCC active block can be used to design various circuits. A single DDCC active block was used to design universal filter [10]. Which is named as the first configuration. In this configuration, three passive resistors (1 kΩ each), two capacitors (10 pF each) along with a DDCC active block are used. It is observed that the transfer functions obtained at various output ports depend on external components (resistors and capacitors) used.

In the proposed configuration, passive resistors are replaced with active resistors. Active resistors are implemented using NMOS transistors working in triode (linear) region of operation. R_{on} represents the resistance offered by NMOS transistors (Fig. 3).
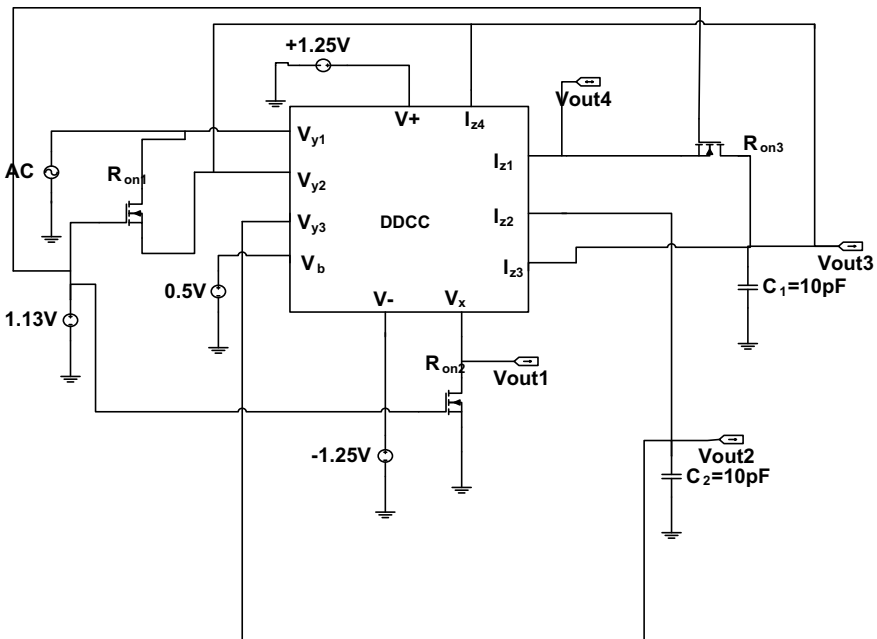
Fig. 3 Circuit diagram showing the second configuration for design of universal filter using DDCC

The transfer functions obtained at various output ports are given by the equations that follow.

$$\frac{Vout1}{Vin} = \frac{s^2C_1C_2}{s^2C_1C_2 + s[C_2(1/R_{on1} - 1/R_{on2}) + C_1 \cdot 1/R_{on2}] + 1/R_{on1} \cdot 1/R_{on2}} \quad (4)$$

$$\frac{Vout2}{Vin} = \frac{-sC_1 \cdot 1/R_{on2}}{s^2C_1C_2 + s[C_2 \cdot (1/R_{on1} - 1/R_{on2}) + C_1 \cdot 1/R_{on2}] + 1/R_{on1} \cdot 1/R_{on2}} \quad (5)$$

$$\frac{Vout3}{Vin} = \frac{s \cdot C_2(1/R_{on1} - 1/R_{on2}) + 1/R_{on1} \cdot 1/R_{on2}}{s^2 \cdot C_1 \cdot C_2 + s[C_2 \cdot (1/R_{on1} - 1/R_{on2}) + C_1 \cdot 1/R_{on2}] + 1/R_{on1} \cdot 1/R_{on2}} \quad (6)$$

$$\frac{Vout4}{Vin} = \frac{s^2C_1C_2 \cdot \frac{R_{on3}}{R_{on2}} + s \cdot C_2 \cdot (1/R_{on1} - 1/R_{on2}) + 1/R_{on1} \cdot 1/R_{on2}}{s^2 \cdot C_1 \cdot C_2 + s[C_2 \cdot (1/R_{on1} - 1/R_{on2}) + C_1 \cdot 1/R_{on2}] + 1/R_{on1} \cdot 1/R_{on2}} \quad (7)$$

Equations (4), (5), (6), and (7) give transfer function for High-Pass filter, Band-Pass filter, Low-Pass filter, and Notch filter when $R_{on1} = R_{on2}$ respectively.

The angular frequency and quality factor of the filters are given by the following equations:

$$\omega_o = \sqrt{1/R_{on1}R_{on2}C_1C_2} \quad (8)$$

$$Q = \frac{\sqrt{C_1C_2 \cdot 1/R_{on1} \cdot 1/R_{on2}}}{C_2.[1/R_{on1} - 1/R_{on2}] + C_1 \cdot 1/R_{on2}} \quad (9)$$

The proposed configuration realizes High-Pass, Band-Pass, Low-Pass, and Notch filters simultaneously using a single DDCC active block, three active resistors, and two grounded passive capacitors. As can be seen from Eqs. (4)–(7), transfer functions of various filters depend on external components, namely resistors and capacitors. With the use of active resistors, value of the resistance offered by it can be easily controlled. Thus, use of active resistor offers programmability of the filter.

4 Simulation Results

The circuit proposed by Chiu et al. [10] was simulated in cadence using United Microelectronics Corporation (UMC) 0.18 µm technology file. Aspect ratio for NMOS transistors is W/L = 1.8 µm/0.18 µm and that for PMOS transistors is W/L = 7.2 µm/0.18 µm.

The DDCC block is used in two configurations. In the first configuration proposed by Chiu et al. [10], three passive resistors, each of 1 kΩ, are used to implement the universal filter. In the second configuration, passive resistors are replaced by

NMOS transistors working as active resistors (Triode region). The W/L ratio is 0.96 µm/0.18 µm and biasing voltage is 1.13 V to bias NMOS in triode region.

In both configurations, the DC voltages are ± 1.25 V and the biasing voltage (V_b) is 0.5 V. Capacitors of value 10 pF are used in both the configurations.

In Fig. 4, output Vout1, Vout2, Vout3, and Vout4 gives High-Pass Filter (HPF), Band-Pass Filter, Low-Pass Filter (LPF), and Notch Filter response respectively. The center frequency for the universal filter is found to be 15.9 MHz which matches with the theoretical analysis.

Similarly, in proposed configuration, HPF, BPF, LPF, and Notch filter response is obtained at Vout1, Vout2, Vout3, and Vout4, respectively. The center frequency of the proposed configuration is also found to be 15.9 MHz shown in Fig. 5.

With the use of active resistors, the value of the resistance offered by it can be easily controlled by changing the applied gate voltage. Thus, use of active resistor offers reduction of noise and area and enhances programmability of the filter also. The following Fig. 6, Fig. 7, Fig. 8, and Fig. 9 depict the programmability of HPF, BPF, LPF, and Notch filter, respectively and Table 1 summarizes the results of universal filter at different cutoff voltages.
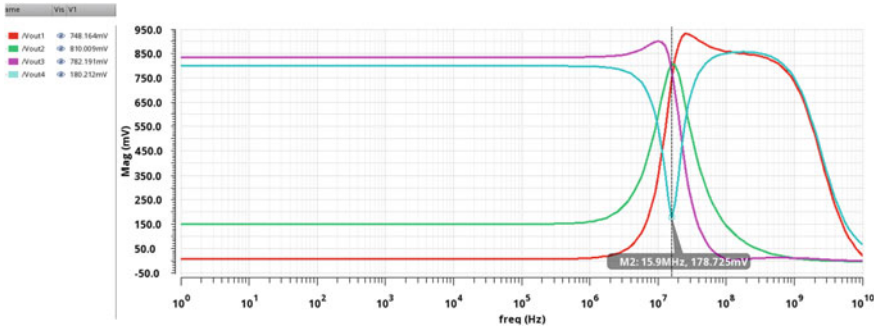
Fig. 4 Magnitude plot for the universal filter in the first configuration
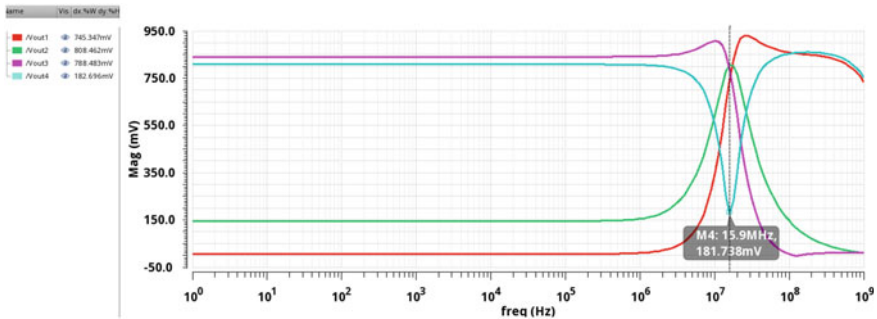
Fig. 5 Magnitude plot for the universal filter in the second configuration
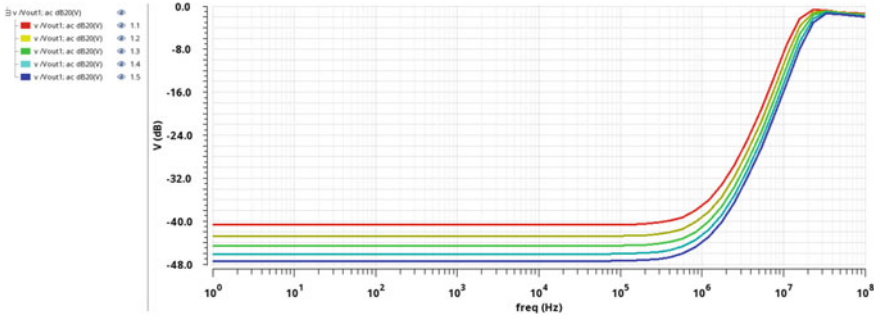
Fig. 6 Magnitude plot of HPF in the proposed configuration
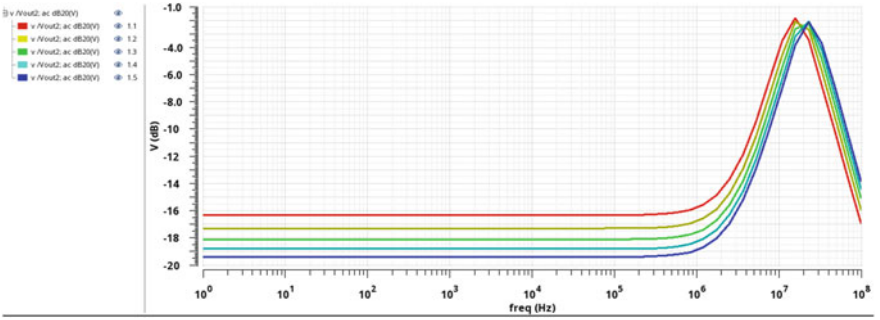
Fig. 7 Magnitude plot of BPF in the proposed configuration
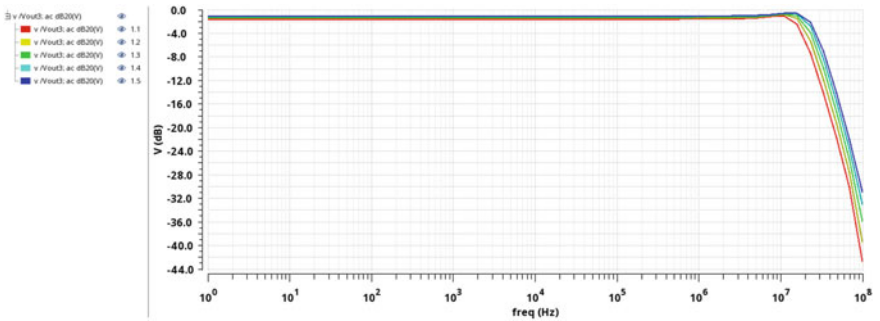
Fig. 8 Magnitude plot of LPF in the proposed configuration
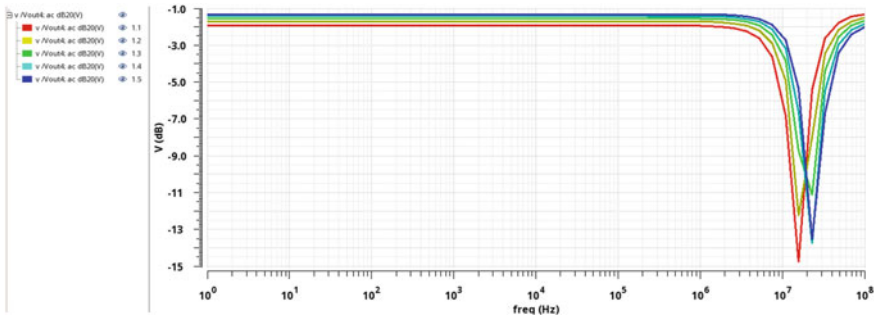
Fig. 9 Magnitude plot of Notch Filter in the proposed configuration

Table 1 Performance of programmable filter

Name of filter	Cutoff frequency (Vg = 1.1 V) (MHz)	Cutoff frequency (Vg = 1.2 V) (MHz)	Cutoff frequency (Vg = 1.3 V) (MHz)	Cutoff frequency (Vg = 1.4 V) (MHz)	Cutoff frequency (Vg = 1.5 V) (MHz)
High-pass filter	13.9	16.02	17.7	19.05	20.3
Band-pass filter	15.82	17.87	19.8	21.47	23
Low-pass filter	19.1	21.73	24.13	26.3	28.34
Notch filter	14.89	17.29	19.83	22.2	24.3

Noise analysis has been done on both configurations. Value of square of noise voltage is integrated over the frequency range 1 Hz–100 MHz. Table 2 shows the values for noise analysis.

For Vout1 (HPF), the noise value decreases by 19.11% with the use of active resistor as compared to passive resistor. Similarly, the noise values of the other outputs are also decreased viz Vout2 (BPF) by 10.76%, Vout3 (LPF) by 24.2%, and Vout4 (Notch) by 24.5%.

Table 2 Noise analysis comparison between passive resistor and active resistor

Output	Noise passive resistor (dB)	MOSFET resistor (dB)	Percentage decrease in noise (%)
VOUT1 (HPF)	2.93×10^{-7}	2.37×10^{-7}	19.11
VOUT2 (BPF)	9.5289×10^{-7}	8.503×10^{-7}	10.76
VOUT3 (LPF)	3.0567×10^{-7}	2.3159×10^{-7}	24.2
VOUT4 (Notch)	5.198×10^{-7}	3.92×10^{-7}	24.5

Third Harmonic Distortion (HD3) was also analyzed at 15.9 MHz and 360 mV input. The plots of HD3 for the passive resistor and active resistor configurations are shown in Fig. 10, Fig. 11, Fig. 12, and Fig. 13 for HPF, BPF, LPF, and Notch filter, respectively.

It can be conclusively observed in Table 3 that Harmonic Distortion has increased for each of the filter response with the use of active resistor. Power dissipation for both the configurations was found to be 3.048 mw.
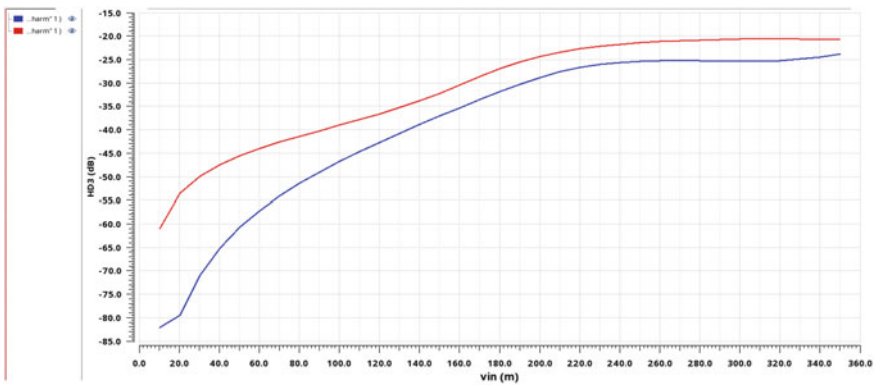
Fig. 10 HD3 plot for the HPF
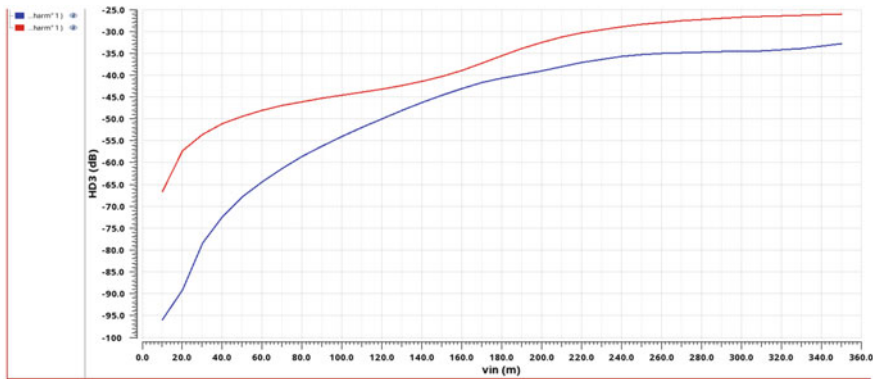
Fig. 11 HD3 plot for the BPF
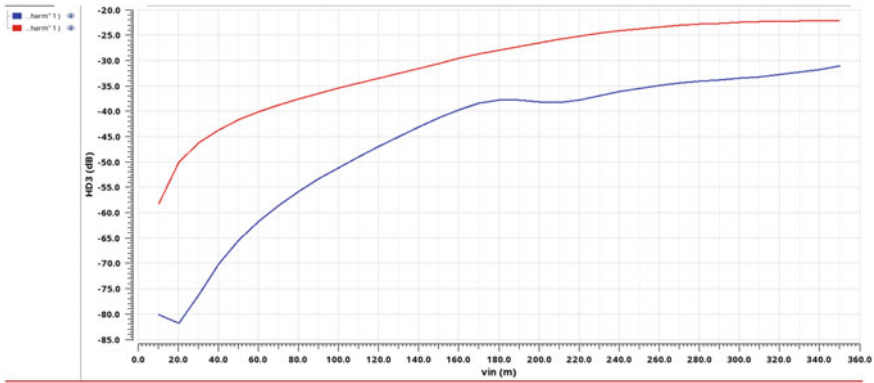
Fig. 12 HD3 plot for the LPF
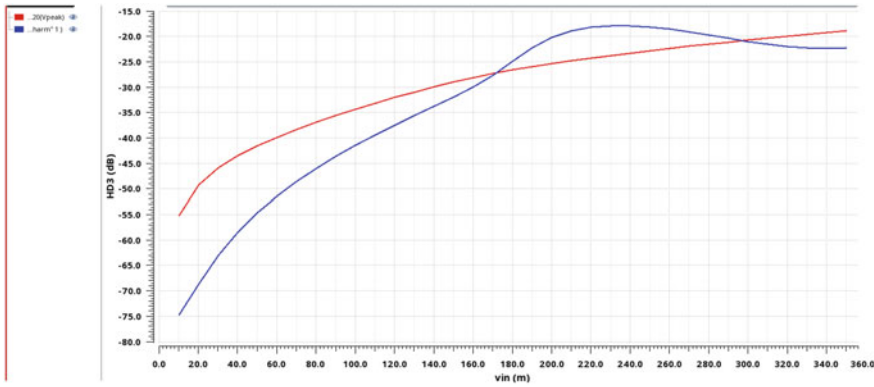
Fig. 13 HD3 plot for the Notch Filter

Table 3 HD3 analysis comparison between passive resistor and active resistor at 250 mV input voltage

Name of filter	HD3 (dB) for passive resistor at 250 mV (dB)	HD3 (dB) for active resistor at 250 mV (dB)	Percentage increase in HD3 of active resistor Vs passive resistor (%)
High-pass filter	−35	−28	20
Band-pass filter	−25	−21	16
Low-pass filter	−35	−24	31.4
Notch filter	−18	−23	−27.78

5 Conclusion

Frequency response of the filter depends on external components (resistors and capacitors) when DDCC active block is used in the circuit designed by Chiu et al. [10]. Therefore, the cutoff frequency can be controlled by changing external resistors. The circuit was implemented in two configurations. In the first configuration, passive resistors are used whereas in proposed configuration, active resistors are used. The active resistors, implemented using NMOS transistors, are operating in triode region. With the use of active resistors, value of resistance can easily be controlled to set desired value of cutoff and central frequencies as well as quality factor (Q) even after post-fabrication. Hence, the proposed configuration offers programmability of the filter.

It can be seen that the frequency response of both the configurations match when the value of active resistor is set to 1 kΩ (same as the value of passive resistor in first configuration). After performing noise analysis, it was observed that MOS as resistor offers more noise immunity. However, Total Harmonic Distortion slightly increases with the use of MOS as active resistor. The proposed configuration offers programmability more noise immunity at the cost of slight increase in Total Harmonic Distortion.

References

1. Kuntman, H.H., Uygur, A.: New possibilities and trends in circuit design for analog signal processing. In: 2012 International Conference on Applied Electronics (AE). IEEE (2012)
2. Chiu, W., et al.: CMOS differential difference current conveyors and their applications. IEE Proc. Circuits Devices Syst. **143**(2), 91–96 (1996)
3. Horng, J.W., Chiu, T.Y., Jhao, Z.Y.: Tunable versatile high input impedance voltage-mode universal biquadratic filter based on DDCCs. Radioengineering **21**(4), 1260–1268 (2012)
4. Prasad, D., Bhaskar, D.R., Srivastava, M.: Universal current-mode biquad filter using a VDTA. Circuits Syst. **4**(01) 29 (2013)
5. Chang, C.-M., Swamy, M.N.S., Soliman, A.M.: Analytical synthesis of voltage-mode even/odd-nth-order differential difference current conveyor and fully differential current conveyor II-grounded resistor and capacitor universal filter structures. Int. J. Circuit Theory Appl. **43**(10) 1263–1310 (2015)
6. Horng, J.-W.: Three-input-one-output current-mode universal biquadratic filter using one differential difference current conveyor. Indian J. Pure Appl. Phys. (IJPAP) **52**(8), 556–562 (2015)
7. Kushwaha, A.K., Paul, S.K.: Current mode universal filter using single current controlled differential difference current conveyor transconductance amplifier. Circuits Syst. **6**(10), 224 (2015)
8. Sackinger, E., Guggenbuhl, W.: A versatile building block: the CMOS differential difference amplifier. IEEE J. Solid-State Circuits **22**(2), 287–294 (1987)
9. Smith, K.C., Sedra, A.: The current conveyor—a new circuit building block. Proc. IEEE **56**(8), 1368–1369 (1968)
10. Chiu, W.-Y., Horng, J.-W.: Voltage-mode highpass, bandpass, lowpass and notch biquadratic filters using single DDCC. Radioengineering **21**(1), 297–303 (2012)

0.5 V Two-Stage Subthreshold Fully Differential Miller Compensated OTA Using Voltage Combiners

Sougata Ghosh and Vijaya Bhadauria

Abstract A simple high-performance architecture for low-voltage and power-efficient gate along with bulk-driven miller compensated Fully Differential Operational Transconductance Amplifier (FDOTA) for biomedical applications is presented in this paper. The proposed design is suitable for operation under sub-1 V single supply and consists of two gain stages. Voltage Combiner (VC) based pseudo-differential circuit has been used in the second stage in order to increase the DC gain. In the proposed design, all the MOSFETs are biased to operate in the subthreshold region for minimum power consumption. The OTA is implemented in the Cadence Virtuoso Environment using 180 nm standard CMOS technology under 0.5 V and consumes only 70nW power. DC gain, Unity Gain Bandwidth (UGB), and phase margin are found to be 68.0656 dB, 9.395 kHz, and 71.90425°, respectively at a capacitive load of 5 pF. The minimum input-referred noise at 10 Hz and 10 kHz frequencies is found to be 4.6161 µV/sqrtHz and 276.323 nV/sqrtHz, respectively. The OTA is simulated for different process corners and temperature variations also.

Keywords Operational Transconductance Amplifier (OTA) · Voltage Combiner (VC) · Unity Gain Bandwidth (UGB)

1 Introduction

Nowadays in our daily life, the demand of portable handheld devices such as notebook and laptops, cell phone, wireless sensor networks, and biomedical implantable devices [1] have become attractive. With the rapid growth of these devices, the design of low-power CMOS circuits is always preferable because it ensures compact bat-

S. Ghosh (✉) · V. Bhadauria
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Allahabad 211004, Uttar Pradesh, India
e-mail: sougata.vlsi@gmail.com

V. Bhadauria
e-mail: vijaya@mnnit.ac.in

tery size, lightweight, lesser costs, and enhanced battery backup. The reduction of supply voltage and reduced dimensions of the components are the main factors to achieve the minimum power consumption and improved performance. To meet the above requirements, subthreshold transistors biased with currents below 500 nA is recommended. Therefore, the low-frequency applications which require low-speed and bandwidth specifications within a few kHz can be better implemented using sub-threshold operated circuits. The subthreshold operation suits low-voltage design due to reduced drain-to-source voltage (V_{DS}) of around 78 mV as compared to 250 mV required in strong inversion [2–4]. With the downscaling of the feature size in CMOS technology, the supply voltage has decreased to avoid breakdown and increase the reliability of the device. Analog circuit designers face difficulties to maintain reliable performance as the supply voltage and threshold voltage are not scaled down pro-portionally. Hence the headroom of the analog operations is decreased. Also, shorter channel devices tend to have short channel effects which reduce its intrinsic gain. So, it is very difficult to obtain sufficient gain of single-stage amplifier. Cascode transistors cannot be used since they limit the output swing. To increase the gain, a self-cascode [3–6] transistor is often used as it gives high-resistance value with a single transistor output voltage compared to cascade transistors.

In recent times, several low-voltage CMOS circuits using bulk-driven transis-tors such as differential amplifiers, current mirrors, and voltage buffers have been designed to increase the ICMR range as well as output swing as the threshold volt-age is eliminated from the signal path. But in this proposed design, gate along with bulk-driven MOS transistors are used to provide a sufficient gain as compared to the present state-of-the-art.

In this paper, a low-voltage, low-power, subthreshold, two-stage, pseudo-class AB amplifier (comprised of DTMOS-based input pair) is designed which works satisfactorily with 0.5 V supply. To achieve sufficient DC gain, the proposed design employs two-stage architecture. The input stage of OTA works as pseudo inverter which increases the input transconductance (g_{mI}). A Voltage Combiner technique is used to enhance the output transconductance (g_{mII}). The proposed OTA has been designed with a load capacitance of 5 pF.

This paper has been organized into five sections. Section 2 deals with the circuit description. The detailed analysis of the derivations of required expressions is given in Sect. 3, followed by the simulated results in Sect. 4 as well as comparison of different amplifiers designed with 180 nm technologies in Sect. 5. Finally, the paper is concluded in Sect. 6.

2   Circuit Description

The circuit diagram of the proposed OTA is shown in Fig. 1. The first stage is pseudo inverter-based differential stage M_1–M_8, the second stage is a Voltage Combiner M_9–M_{12}. The inverting and non-inverting input terminals are denoted as V_{IN} and
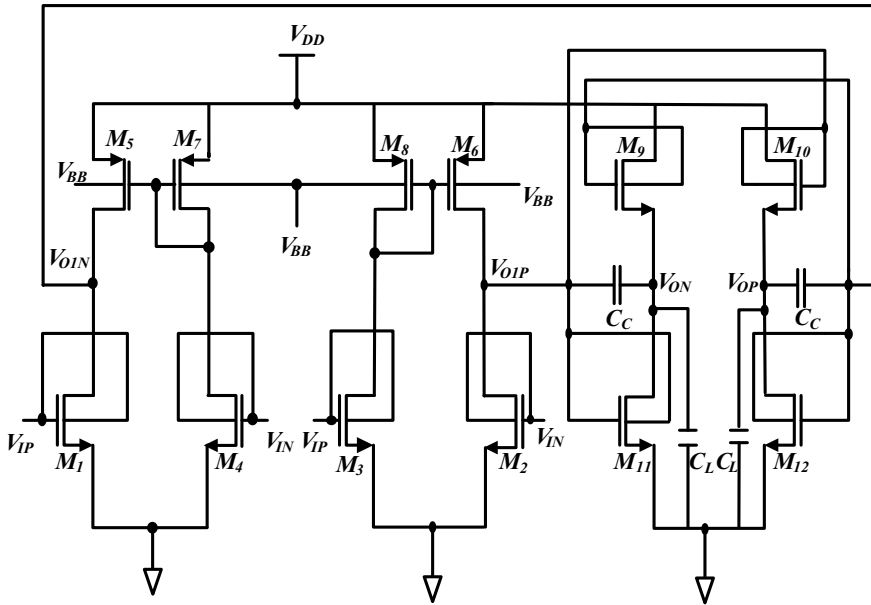
Fig. 1 Schematic of the proposed OTA

V_{IP}, similarly, V_{ON} and V_{OP} are the output terminals of the OTA. The first stage of this proposed OTA is built by dual current mirror load (M_5–M_7 and M_6–M_8) in two pseudo-differential pairs M_1–M_4 with M_2–M_3. As observed from the circuit diagram, the additional inputs are fed to the gates of M_5 and M_6 through a circuitry M_3–M_4 making the circuit as a pseudo inverter. Because of additional inputs, the configuration makes an improvement in input transconductance (g_{mI}). The threshold voltage of PMOS transistors M_5–M_8 is reduced due to the bulk voltage V_{BB}. The diode-connected MOS M_7 and M_8 set predefined output common-mode voltage which eliminates the need of additional common-mode feedback circuit in the first stage. Input is given to the gate and bulk terminal of input stage driver transistor M_1–M_4 which also increases g_{mI} but that requires an additional fabrication step. The voltage gain of the first stage is given by

$$A_{Vdiff1} = \left[(g_{m1} + g_{mb1}) + g_{m5} \left(\frac{g_{m4} + g_{mb4}}{g_{m7} + g_{ds7} + g_{ds4}} \right) \right] \left(\frac{1}{g_{ds1} + g_{ds5}} \right) \quad (1)$$

The circuit in Fig. 1 is designed to operate in a fully differential configuration. Identical pair of transistors are M_1 and M_2, M_3 and M_4, … and M_{11} and M_{12}. In Eq. (1) g_{mX}, g_{mbX}, and g_{dsX} are gate transconductance, bulk transconductance, and drain conductance of Xth transistor. The second stage configuration in this OTA is a Voltage Combiner [7] based pseudo-differential amplifier (M_9–M_{12}) with dynamic

body biasing. The primary aim of the second stage is to provide large transconductance in order to improve overall gain and also to keep the Right Half Plane (RHP) zero at a higher frequency for the improvement of phase margin. Voltage Combiner based pseudo-differential amplifier consists of NMOS transistors M_9 and M_{10} in CD (common drain) configuration and other NMOS transistors M_{11} and M_{12} in CS (common source) configuration. The open-loop gain expression of this second stage VC-based differential amplifier is expressed as

$$A_{Vdiff,2} = \frac{g_{m9} + g_{mb9} + g_{m11} + g_{mb11}}{g_{ds9} + g_{ds11} + g_{m9} + g_{mb9}} \tag{2}$$

The schematic of the proposed two-stage pseudo fully differential OTA is shown in Fig. 1.

3   Analysis and Design

This section describes the characteristics of the proposed OTA along with the required design expressions with derivations.

3.1   Differential Gain and Common-Mode Gain for First Stage

From Fig. 1, it is obvious that the first stage is symmetric, so differential gain as well as common-mode gain can be found using half circuit method.

To determine the differential gain, a small-signal differential voltage ($V_{id}/2$) is applied at both the input terminals of the half circuit keeping the common-mode voltage V_{CM} zero as shown the Fig. 2. In order to calculate the differential gain, first V_A is to be determined.

The corresponding small-signal equivalent of the half circuit is shown in Fig. 3a. Similarly, V_B is determined from the small-signal equivalent circuit of the left side of the half circuit. This is shown in Fig. 4.

From this Fig. 3a, the following relation is obtained:

$$V_A(g_{ds7} + g_{m7} + g_{ds4}) = (g_{m4} + g_{mb4})\frac{V_{id}}{2} \text{ or } V_A = \frac{(g_{m4} + g_{mb4})\frac{V_{id}}{2}}{g_{ds7} + g_{m7} + g_{ds4}} \tag{3}$$

From Fig. 4a, the following relation is obtained:

$$V_B(g_{ds5} + g_{ds1}) = -(g_{m1} + g_{mb1})\frac{V_{id}}{2} - g_{m5}V_A \tag{4}$$
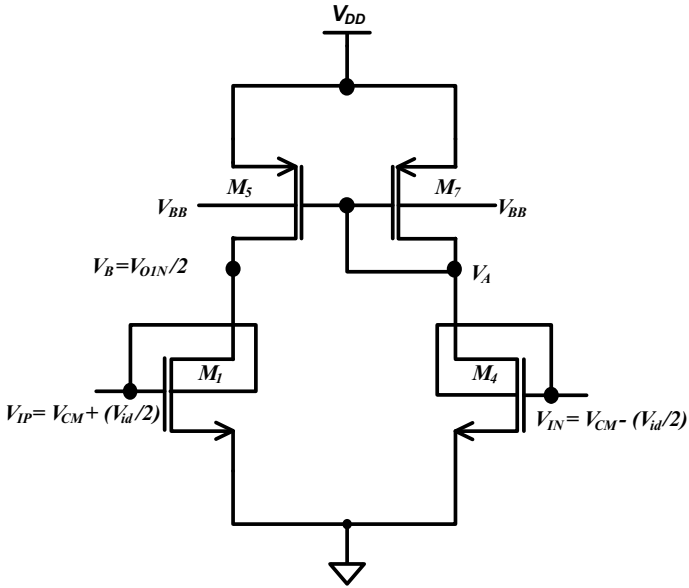
Fig. 2 Half circuit of first stage pseudo-differential amplifier
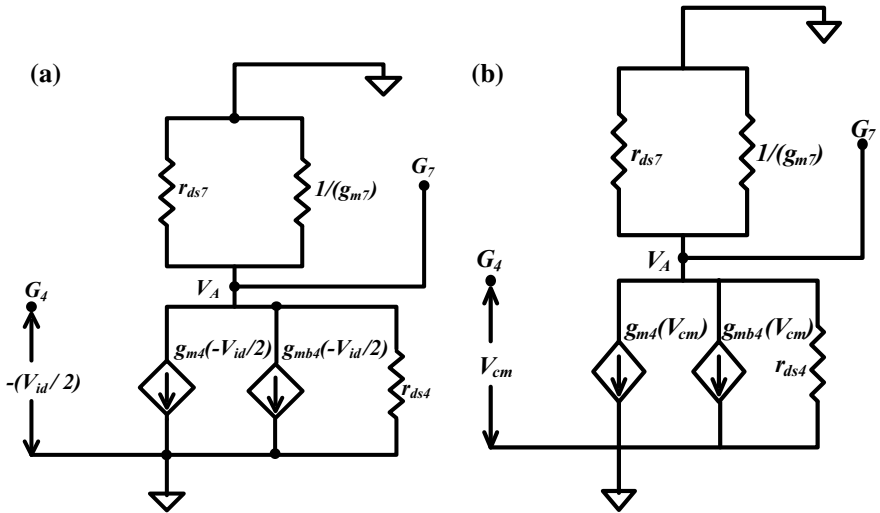
Fig. 3 Low-frequency small-signal equivalent of right side of the half circuit for finding V_A: differential signal (**a**), common-mode signal (**b**)
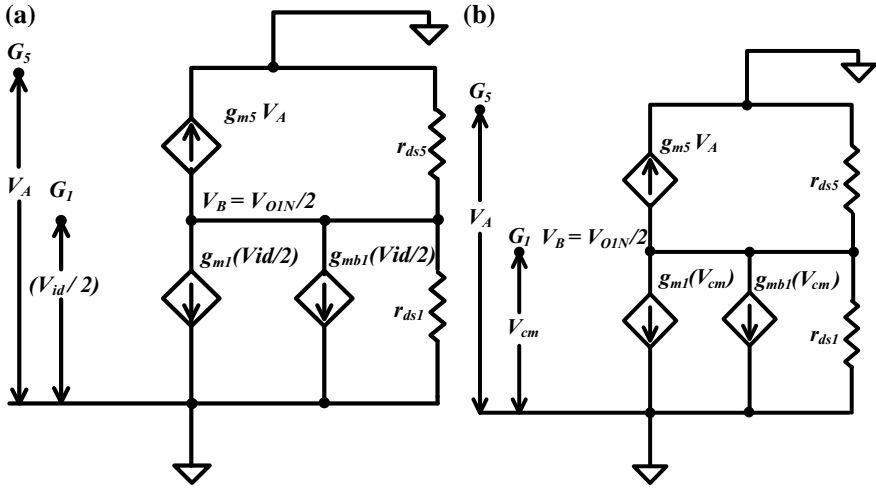
Fig. 4 Low-frequency small-signal equivalent of left side of the half circuit for finding V_B: differential signal (**a**), common-mode signal (**b**)

Substituting the value of V_A from Eq. (3) into Eq. (4)

$$V_B = \left[-(g_{m1} + g_{mb1}) \frac{V_{id}}{2} - \frac{g_{m5}(g_{m4} + g_{mb4})(\frac{V_{id}}{2})}{g_{ds7} + g_{m7} + g_{ds4}} \right] \left(\frac{1}{g_{ds5} + g_{ds1}} \right)$$

$$\frac{V_B}{-\frac{V_{id}}{2}} = A_{Vdiff1} = \left[g_{m1} + g_{mb1} + \frac{g_{m5}(g_{m4} + g_{mb4})}{g_{ds7} + g_{m7} + g_{ds4}} \right] \left(\frac{1}{g_{ds5} + g_{ds1}} \right)$$

$$A_{Vdiff1} = \left[(g_{m1} + g_{mb1}) + g_{m5}(\frac{g_{m4} + g_{mb4}}{g_{m7} + g_{ds7} + g_{ds4}}) \right] \left(\frac{1}{g_{ds5} + g_{ds1}} \right) \tag{5}$$

Common-mode gain is calculated with V_{id} set to zero. In Fig. 3b, the final expression for V_A is given by Eq. (6).

$$V_A(g_{m7} + g_{ds7}) = -\frac{V_A}{r_{ds4}} - (g_{m4} + g_{mb4})V_{CM} \text{ or, } V_A = \frac{-(g_{m4} + g_{mb4})V_{CM}}{g_{m7} + g_{ds7} + g_{ds4}} \tag{6}$$

In a similar way, in Fig. 4b, the expression of V_B (for common-mode gain) is given by Eq. (7).

$$\frac{V_B}{r_{ds5}} + g_{m5}V_A + \frac{V_B}{r_{ds1}} + (g_{m1} + g_{mb1})V_{CM} = 0 \tag{7}$$

Substituting the value of V_A from Eq. (6) into Eq. (7)

$$g_{m5}\frac{-(g_{m4}+g_{mb4})V_{CM}}{g_{m7}+g_{ds7}+g_{ds4}}+(g_{m1}+g_{mb1})V_{CM}=-V_B\big[g_{ds1}+g_{ds5}\big]$$

$$\frac{-V_B}{V_{CM}}=A_{Vcm,1}=\left[(g_{m1}+g_{mb1})-g_{m5}\frac{(g_{m4}+g_{mb4})}{g_{m7}+g_{ds4}+g_{ds7}}\right]\left(\frac{1}{g_{ds5}+g_{ds1}}\right) \quad (8)$$

The final expression for common-mode gain is obtained as depicted in Eq. (8).

3.2 Differential Gain and Common-Mode Gain for Second Stage

As the second stage is Pseudo-differential amplifier; CMRR $= 1$, i.e., differential gain and common-mode gain are equal and from the small-signal analysis of the Voltage Combiner, the open-loop gain (DC gain) and the common-mode gain may be expressed as

$$A_{Vdiff2}=A_{Vcm,2}=\frac{g_{m9}+g_{mb9}+g_{m11}+g_{mb11}}{g_{ds9}+g_{ds11}+g_{m9}+g_{mb9}} \quad (9)$$

3.3 Total Differential Gain and Common-Mode Gain of the Proposed OTA

The total gain from two stages can be obtained by multiplying individual gain expressions, which is given by

$$A_{V,DM}=A_{Vdiff1}\cdot A_{Vdiff2}=\left(\frac{g_{mI}}{g_{ds1}+g_{ds5}}\right)\cdot\left(\frac{g_{m9}+g_{mb9}+g_{m11}+g_{mb11}}{g_{ds9}+g_{ds11}+g_{m9}+g_{mb9}}\right) \quad (10)$$

where

$$g_{mI}=(g_{m1}+g_{mb1})+g_{m5}\left(\frac{g_{m4}+g_{mb4}}{g_{m7}+g_{ds7}+g_{ds4}}\right)$$

$$A_{VCM}=A_{Vcm,1}\cdot A_{Vcm,2}$$

$$A_{V,CM}=\left[\left[(g_{m1}+g_{mb1})-g_{m5}\frac{(g_{m4}+g_{mb4})}{g_{m7}+g_{ds4}+g_{ds7}}\right]\left(\frac{1}{g_{ds1}+g_{ds5}}\right)\right]\left(\frac{g_{m9}+g_{mb9}+g_{m11}+g_{mb11}}{g_{ds9}+g_{ds11}+g_{m9}+g_{mb9}}\right)$$
$$(11)$$

Combining Eqs. (10) and (11)

$$\text{CMRR} = \frac{A_{V,DM}}{A_{V,CM}} = \frac{(g_{m1} + g_{mb1}) + g_{m5}\left(\frac{g_{m4}+g_{mb4}}{g_{m7}+g_{ds7}+g_{ds4}}\right)}{(g_{m1} + g_{mb1}) - g_{m5}\left(\frac{g_{m4}+g_{mb4}}{g_{m7}+g_{ds7}+g_{ds4}}\right)} \tag{12}$$

3.4 Frequency Response

In this paper, Miller compensation technique has been used to compensate the proposed two-stage OTA in order to make it stable. Here, a compensation capacitance is used between two high-resistance nodes for splitting the poles. The OTA is designed to have a higher transconductance (g_{mIIC}) in the second stage, to have sufficient phase margin by keeping zero at a higher frequency.

The dominant, nondominant pole, and Right Half Plane (RHP) zero is located at

$$P_{dominant} = \frac{1}{2\pi R_1 g_{mII} R_2 C_c} \tag{13}$$

$$P_{non-dominant} = -\frac{g_{mII}}{2\pi C_L} \tag{14}$$

$$P_Z = \frac{g_{mI}}{2\pi C_c} \tag{15}$$

where R_1, R_2 denote output impedance at input and output stages.

The Unity Gain Bandwidth of this OTA in Fig. 1 is expressed by

$$\text{Unity gain bandwidth (UGB)} = \frac{g_{mI}}{2\pi C_c} \tag{16}$$

4 Simulation Results

The proposed OTA was implemented and simulated using Cadence Virtuoso Tool with 180 nm standard CMOS technology. The supply voltage is taken as 0.5 V. In this design, the aspect ratio of all transistors have been chosen to optimize the gain and input-referred noise.

1 V magnitude of an AC signal is applied to the input with a load capacitance of 5 pF. The simulated open-loop gain and phase response with load capacitance (C_L) of 5 pF are shown in Fig. 5. A gain of 68.0656 dB, a UGB of 9.395 kHz with a phase margin of 71.90425° have been observed. It is observed from the simulation results that UGB appears to be low since all the transistors are operated in weak inversion.
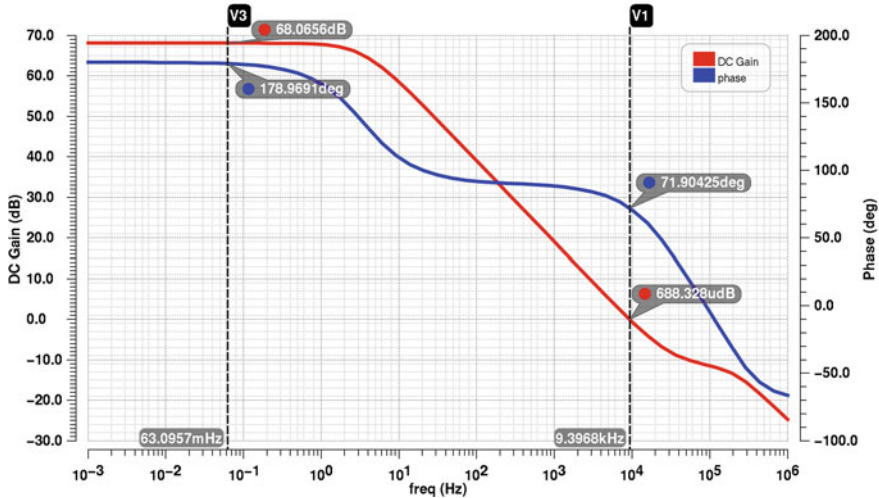
Fig. 5 Open-loop frequency response for gain and phase of the proposed OTA

Figure 6 gives the response of common-mode and power supply rejection of OTA. The differential gain and common-mode gain is found to be 68.07 dB and −8.774 dB, respectively. So from the plot, CMRR is 76.84 dB and the PSRR is 71 dB at 1 mHz approximately. The second stage of the OTA results in a slightly larger common-mode gain at the output node which reduces the CMRR.

The input-referred noise of the proposed OTA is presented in Fig. 7. Due to the large input stage transconductance which is almost two times larger as compared to conventional differential amplifiers [8], the noise is found to be very less, i.e., 588.105 nV/sqrtHz at 1 kHz of frequency.

To examine the robustness of gain and phase response of the OTA, a Monte Carlo analysis for 1000 samples has been done which is shown in Fig. 8. The Gain-Bandwidth value slightly deviates from the actual result due to the process variations.

The proposed OTA is designed to operate well under different device corners and process variations. Five device corners (nominal, SS, FF, SF, FS) have been performed to examine the effect of process corners over the performance of proposed OTA. The gain in typical corner is 68.06 dB. The gain is found to be minimum (65.13 dB) in FF corner and maximum (69.89 dB) in SS corner. Also, all these corners are analyzed for two temperatures 27 and 75 °C to establish the operation of OTA in wide temperature ranges. The variations of gain and phase for all the five corners are shown in Fig. 9.

(a)

(b)
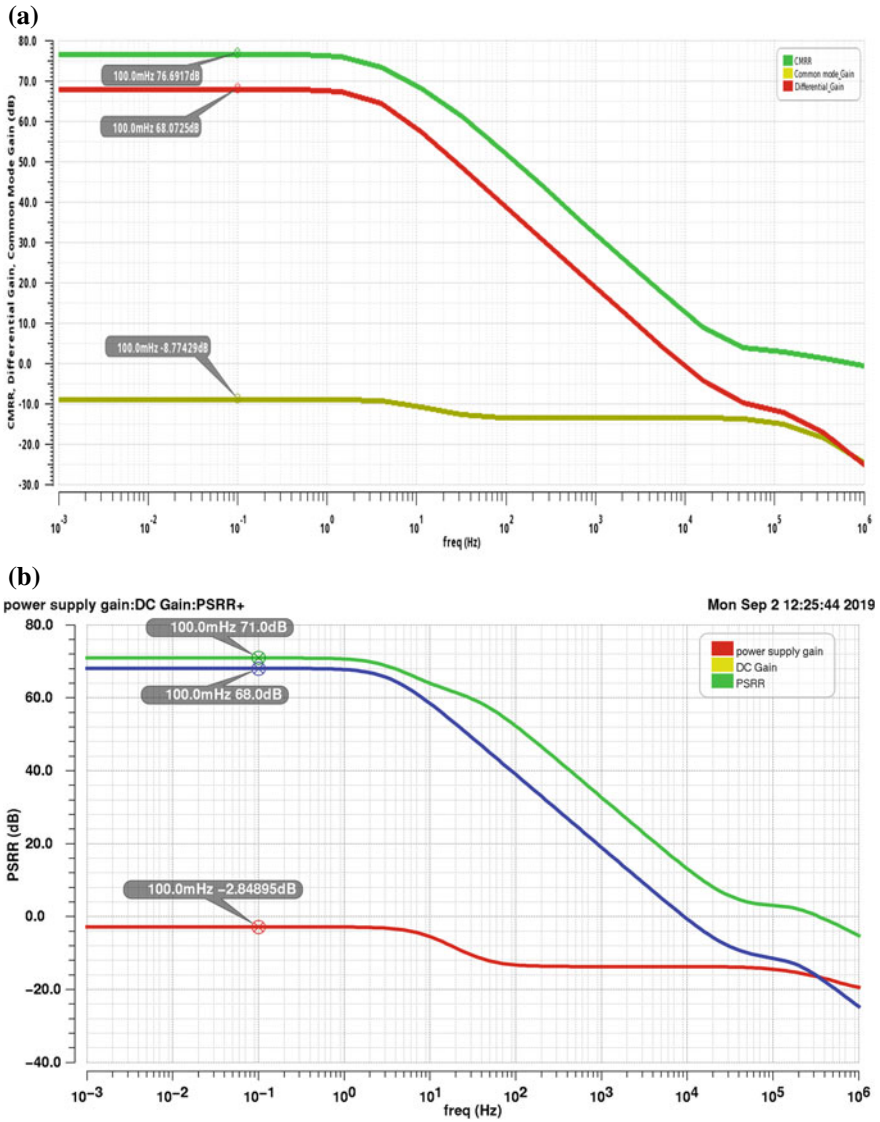
Fig. 6 Response of the proposed OTA at V_{CM} of 0.3 V: CMRR (**a**) and PSRR (**b**)

Input Referred Noise
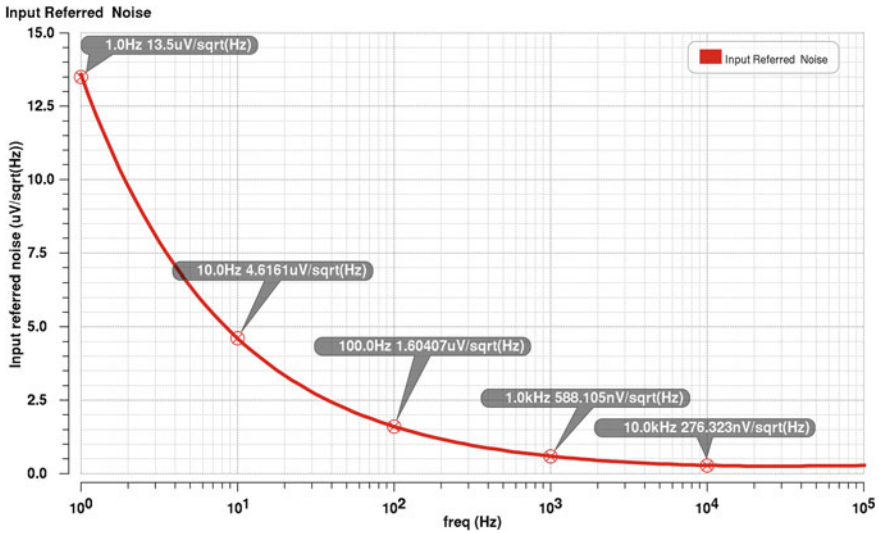
Fig. 7 Input-referred noise characteristics of the proposed OTA

5 Comparison

Performance comparison of proposed OTA with other OTAs designed in 180 nm technology is given in Table 1.

In the context of performance comparison with other OTAs designed in 180 nm technology, the proposed design shows overall better results in open-loop gain, power consumption, phase margin, CMRR. As the proposed design is intended for ultra-low-voltage and low-power applications especially in biomedical narrowband applications for frequencies up to a few KHz. It is very much suitable for narrowband applications. However, the bandwidth can be increased at the cost of power consumption and phase margin. The proposed design consumes less power and it is only 0.25%, 0.28%, 7%, 0.09% of reference paper reported [1, 5, 8, 9] respectively. The simulated results of the proposed design slightly vary with GBW, phase margin, and slew rate at the capacitive load of 10 and 20 pF. In our design, settling time values are 35 μs and 140 μs for 0.1% and 0.01% error, respectively. However, the applications like wireless sensor networks and biomedical applications where speed is not a constraint, the proposed design demands more attention.

Table 1 Performance comparison of proposed OTA, for $C_L = 5$, 10, 20 pF with other OTAs designed in 180 nm technology

Parameters	[5]	[7]	[9]	[1]	This work		
					$C_L = 5$ pF	$C_L = 10$ pF	$C_L = 20$ pF
Power Supply (V)	0.5	0.7	0.8	0.5	0.5	0.5	0.5
CMOS technology (nm)	180	180	180	180	180	180	180
DC gain (dB)	63	57.5	51	62	68.065	68.066	68.065
GBW (MHz)	0.55	3	0.04	10	0.00939	0.00845	0.00718
PM (°)	50	60	65	60	71.904	72.77	73.91
CMRR (dB)	86	19	65	–	76.69	76.70	76.70
PSRR (dB)	81	52	–	–	71	71	71
Load capacitance (pF)	20	20	10	20	5	10	20
Slew rate (V/μs)	0.23	2.8	0.12	2	0.085/0.057	0.072/0.050	0.065/0.047
Power consumption (μW)	28	25	1	75	0.07	0.07	0.07

(a)

(b)
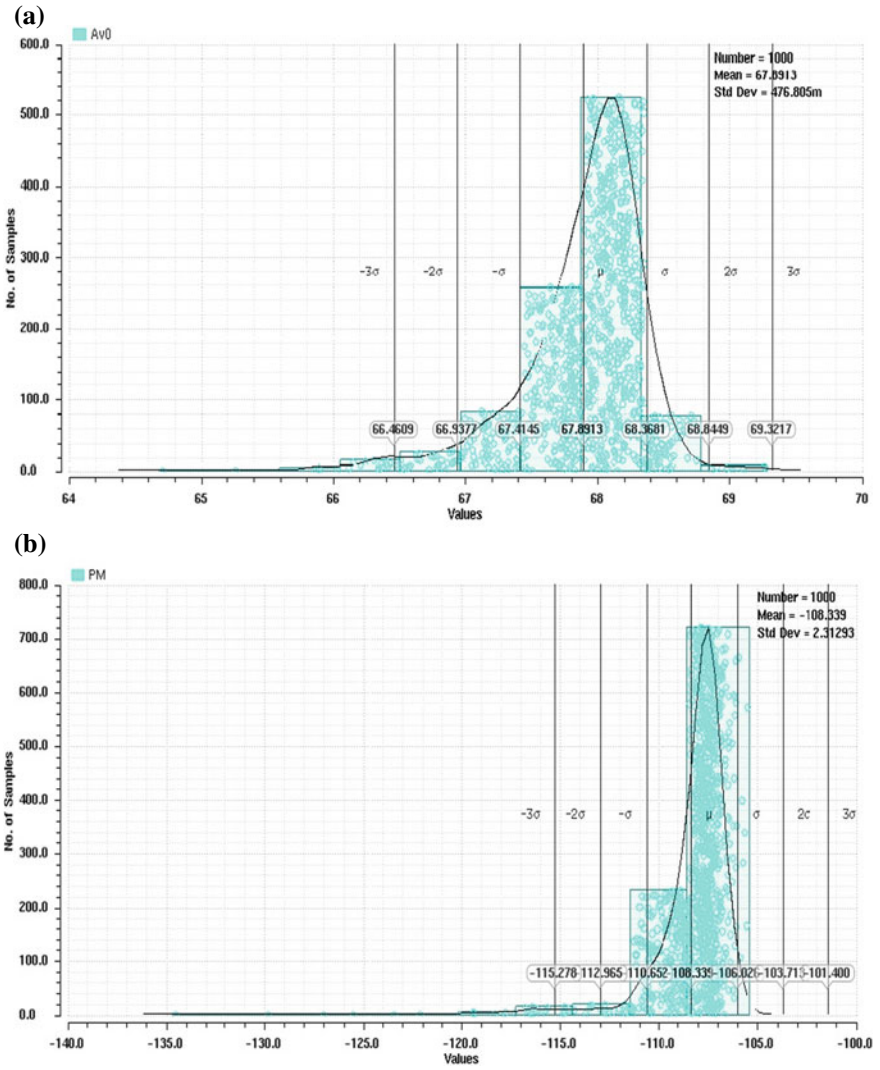
Fig. 8 Monte Carlo simulation results for gain (**a**), phase (**b**), and GBW (**c**)

(c)
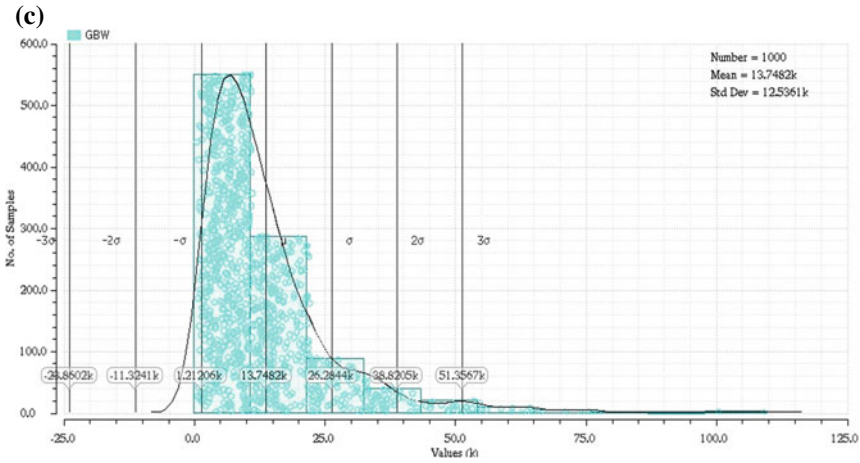
Fig. 8 (continued)

6 Conclusion

This paper presents the approach of low-voltage, low-power, two-stage, fully differential OTA design using Voltage Combiners in the second stage to improve the overall gain. Compared to contemporary reports of OTAs, the proposed design shows better performances with low-voltage, low-power supply. The power consumption is only 70 nW at 0.5 V supply; therefore, it is suitable for biomedical narrow band filter applications. The input-referred noise at 1 kHz is found to be 588.11 nV/sqrtHz and is useful in the biomedical field.

(a)

(b)
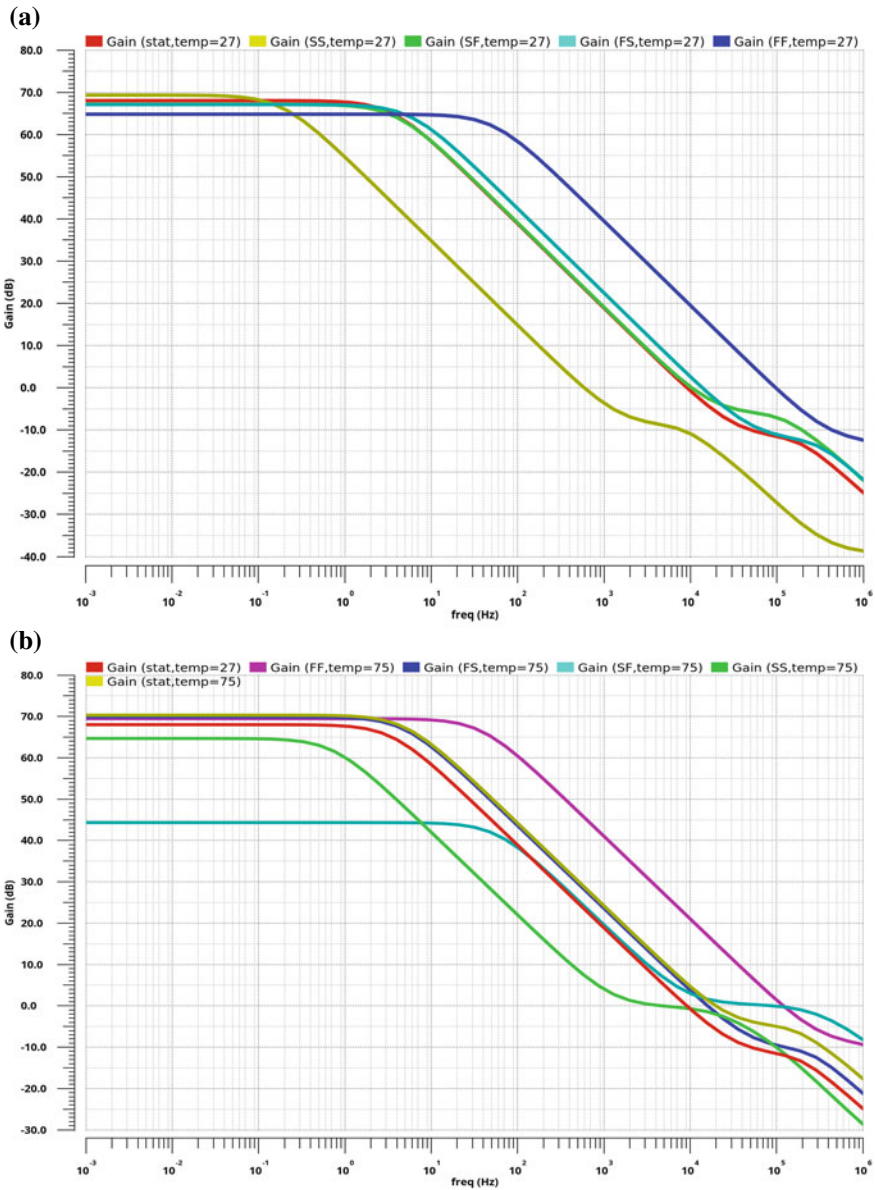
Fig. 9 The effect of process corners on gain at 27 °C (**a**), gain at 75 °C (**b**), phase at 27 °C (**c**), phase at 75 °C (**d**)

(c)

(d)
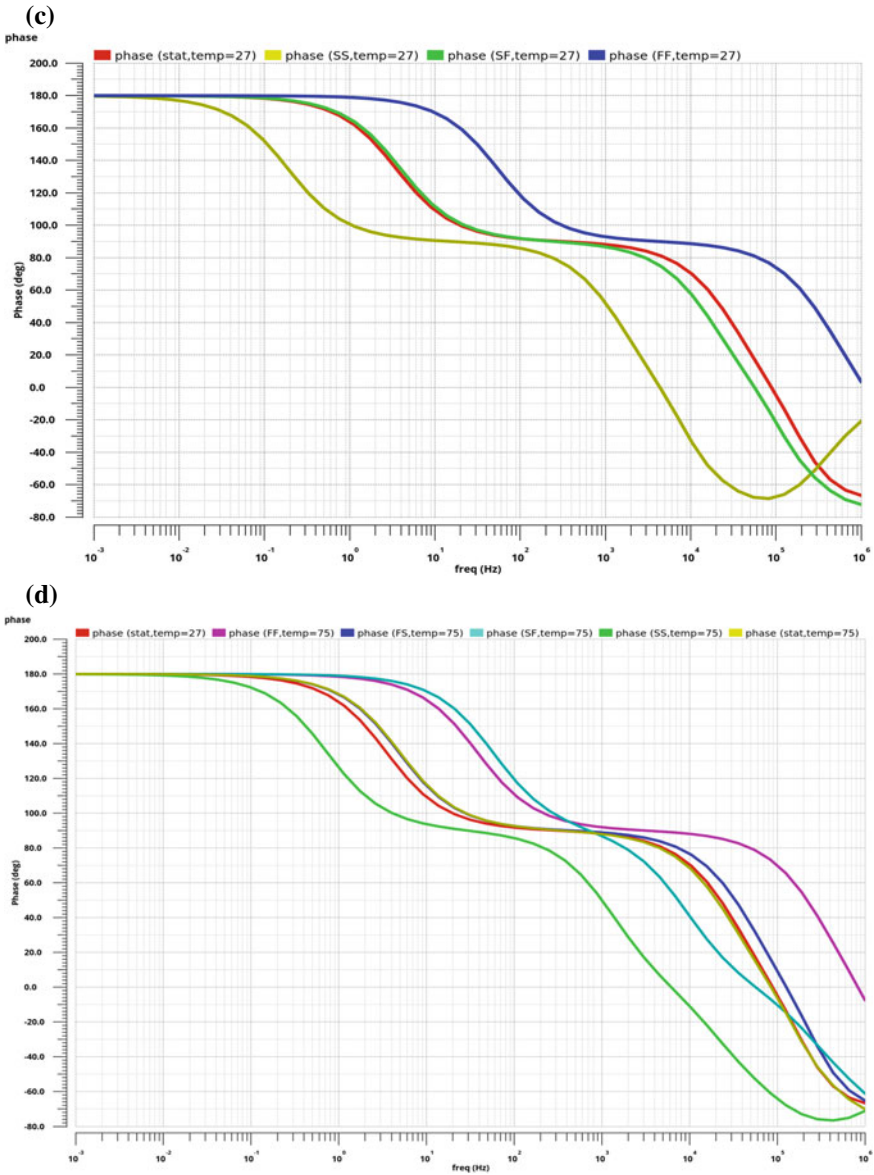
Fig. 9 (continued)

References

1. Chatterjee, S., Tsividis, Y., Kinget, P.: 0.5-V analog circuit techniques and their applications in OTA and filter design. IEEE J. Solid-State Circuits **40**(12), 2372–2387 (2005)
2. Sharan, T., Bhadauria, V.: Fully differential, bulk driven, class AB, sub-threshold OTA with enhanced slew rates and gain. J. Circuits Syst. Comput. **26**(1), 1750001 (2017)
3. Ferreira, L.H.C., Pimenta, T.C., Moreno, R.L.: An ultra- low-voltage ultra-low-power weak inversion composite MOS transistor concept and applications. IEICE Trans. Electron. **91**(4), 662–665 (2008)
4. Ferreira, L.H.C., Pimenta, T.C., Moreno, R.L.: An ultra-low-voltage ultra-low-power CMOS miller OTA with rail-to-rail input/output swing. IEEE Trans. Circuits Syst. II: Express Briefs **54**(10), 843–847 (2007)
5. Trakimas, M., Sonkusale, S.: A 0.5 V bulk-input OTA with improved common-mode feedback for low-frequency filtering applications. Analog Integr. Circuits Signal Process. **59**(1), 83–89 (2009)
6. Ferreira, L.H.C., Sonkusale, S.R.: A 60-dB gain OTA operating at 0.25-V power supply in 130-nm digital CMOS process. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1881–1884. IEEE (2014)
7. Cabrera-Bernal, E., Pennisi, S., Grasso, A.D., Torralba, A., Carvajal, R.G.: 0.7-V three-stage class-AB CMOS operational transconductance amplifier. IEEE Trans. Circuits Syst. I: Regul. Pap. **63**(11), 1807–1815 (2016)
8. Achigui, H.F., Fayomi, C.J.B., Sawan, M.: 1-V DTMOS-based class-AB operational amplifier implementation and experimental results. IEEE J. Solid-State Circuits **41**(11), 2440–2448
9. Valero, M., Celma, S., Medrano, N., Calvo, B., Azcona, C.: An ultra low-power low-voltage class AB CMOS fully differential OpAmp. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1967–1970 (2012)
10. Panigrahi, A., Parhi, A.: Design of 0.5 áV voltage-combiner based OTA with 60 ádB gain 250 ákHz UGB in CMOS. Analog Integr. Circuits Signal Process. **92**(1), 159–165 (2017)

Current Feedback Operational Amplifier-Based Biquadratic Filter

Tripurari Sharan, Khoirom Johnson Singh and Anil Kumar Gautam

Abstract This paper presents a current feedback operational amplifier (CFOA) based on the topology of the second-generation positive current conveyor (CCII+) and an output buffer. The CCII+ is comprised of n-input and p-input balanced current mirror load OTAs with negative feedback from X node to inverting inputs of the differential pairs. Non-inverting inputs of differential pair make Y input node. The dual-input pair with balanced load structure ensures well-matched circuit characteristic and very low-output offset. This CFOA is biased in the strong inversion region using a dual power supply of ± 0.6 V with a bias current of 10 µA. It dissipates the total power of 377 µW and satisfies good CFOA characteristics up to 70 MHz frequency. The CFOA cells have been utilized to design a single input multi output (SIMO) function voltage-mode universal filter which dissipates the total power of 1.79 mW. These circuits have been simulated using cadence simulator tool in 180 nm standard n-tub bulk-CMOS process in its UMC environment.

Keywords Symmetric OTA · Current conveyor II+ · Current feedback operational amplifier · Biquadratic voltage-mode filter

1 Introduction

The CMOS analog design methodology has proved itself better than bipolar technology in low-area, low-voltage, low-power aspects which have enabled highly compact low-cost systems having higher design simplicity. Various analog cells are being

T. Sharan (✉) · A. K. Gautam
NERIST Deemed to be University, Nirjuli, Arunachal Pradesh 791109, India
e-mail: tsh1962mzp@gmail.com

A. K. Gautam
e-mail: anilgautam19@gmail.com

K. J. Singh
National Institute of Technology, Chumukedima, Dimapur, Nagaland 791103, India
e-mail: khoirom@rediffmail.com

481

used as voltage and current amplifiers, current integrators and differentiators, capacitance multipliers, impedance simulators, analog-to-digital and digital-to-analog converters, instrumentation amplifiers, oscillators, and waveform generators, etc. [1–4]. Though the CMOS devices suffer from threshold voltage mismatch between nMOS and pMOS devices, body bias effect and reduced transconductance g_m value than bipolar-based circuits, yet its low-cost fabrication, especially in standard n-tub CMOS technology process, has enhanced its popularity, success, and utility over bipolar technology [3]. The various analog design cells, such as an op-amp, OTA, OTRA, CCII+, and CFOA are being used for analog system design. They offer their own merits and demerits over one another and selection of a cell for a given analog system is based on types of input and output signal as well as input and output impedances at input and output ports [4–6]. Nowadays, current-mode circuit approach has become more popular than its voltage-mode circuits counterparts. To assure full integration within the SoC chip, the OTA and OTA-C filters have gained more popularity than op-ampbased RC-filter design [7–11]. The second-generation positive and negative current conveyer II has also been used for many analog systems designs in both voltage- and current-mode aspects but voltage-mode cell, the op-amp has slew rate and gain-bandwidth limitations over current feedback operational amplifier-based circuits [12, 13]. So, the CCII+ and CFOA have proved itself more popular for LV-LP design having enhanced slew rate with constant gain independent of its bandwidth. The CCII+ and CFOA cells operate well at low-supply voltage and reduced power dissipation and are suitable design cells for both voltage- and current-mode design approaches [14].

The design of mixed-signal system utilizes some of the analog cells such as op-amps, operational transconductance amplifiers (OTAs), second-generation positive current conveyors (CCII+), and CFOAs [15, 16]. Among these cells, the CFOA which came into existence in around 1985, has a constant bandwidth (up to 100 MHz) independent of its closed-loop gain. It offers a very high slew rate (typically 2 kV/µs) suitable for very high-frequency applications [17]. All the analog building block made of op-amps could also be realized using CFOAs [18]. The CFOA is particularly well suited for applications requiring variable closed-loop gains with constant bandwidth, such as in automatic gain control applications [14]. The CCII+ cum output buffer-based topology is most popular among other versions of CFOA topologies addressed in the literature [19]. The input stage of the CFOA is a CCII+ , whereas the CCII+ contains an OTA and some current buffers [14]. So, the overall performance of CFOA is dependent on its OTA as well as its current and voltage buffers. The low-voltage CFOA requires the supply voltage of at least 2 $(V_T + V_{OV})$ where V_T is the threshold voltage of the MOSFET and V_{OV} is its overdrive voltage $(V_{GS} - V_T)$. This 2 $(V_T + V_{OV})$ is nearly 1.5 V in 180 nm CMOS process and can further be reduced to 1.2 V with a rail-to-rail dual-input pair-based design approach as addressed in Ref. [14].

AD844 is a commercially available current feedback op-amp (CFOA) with an accessible Z-terminal. It has the option that its compensation pin (number 5) is externally accessible while still maintaining pin capability with op-amps [14]. The prominent nonidealities of the CFOAs include a finite nonzero input resistance Rx

(a)

(b)
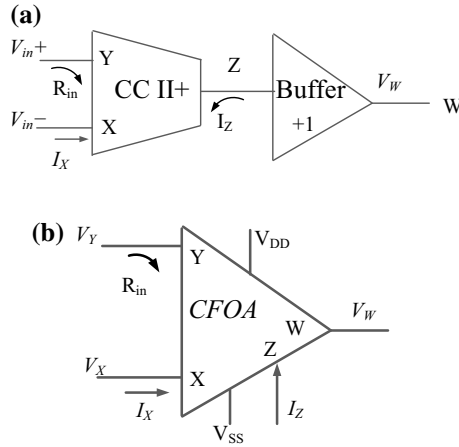
Fig. 1 **a** Used CFOA topology and **b** circuit symbol of CFOA
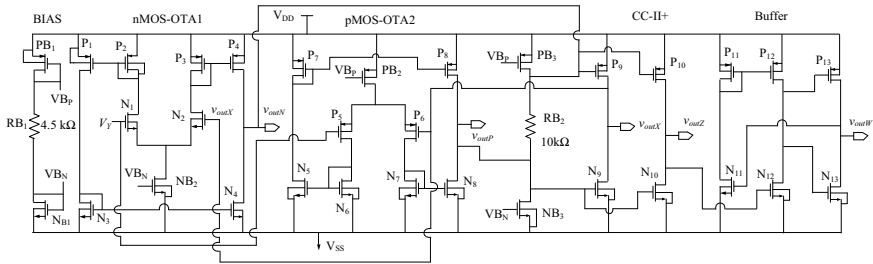
Fig. 2 Schematic of low-voltage CMOS CFOA

at port-X (typically around 50 Ω), the Z-port parasitic impedance consisting of a parasitic resistance Rp (typically 3 MΩ) in parallel with a parasitic capacitance Cp (typically, 5 pF) and the Y-port parasitic impedance consisting of a parasitic resistance Ry (typically 2 MΩ) in parallel with a parasitic capacitance Cy (typically, 2 pF) [14, 20, 21].

In this paper, a low-voltage CFOA cell, based on CCII+ followed by an output buffer has been described in Sect. 2. The CFOA cells have been utilized to design a biquadratic single input multiple output functions (SIMO) type voltage-mode universal filter in Sect. 3. Section 4 presents the comparison of this work with some of the referred research papers and Sect. 5 concludes the various results.

2 CFOA Circuit Schematics

Figure 1a, b show the used topology and circuit symbol of CFOA whereas Fig. 2 shows the overall circuit schematic of CMOS CFOA designed in bulk n-tub standard CMOS process technology in strong inversion region with gate-driven input approach to ensure high-frequency applications up to 100 MHz frequency. The design of this CFOA is based on CCII+ followed by an output buffer approach [22].

The CCII+ has utilized two complementary differential pairs (DPs). The first one is n-input DP and the latter one is p-input DP. These OTAs have used balanced current mirror load structures as its input core. Non-inverting inputs of dual DPs are tied as Y input whereas the inverting inputs of both the OTAs have been tied as one node for X terminal of CCII+. The output nodes of DPs are combined using a combiner network and current mirrored nodes which comprise the final X input node as well as final output, i.e., the Z node of the CCII+. The final X node is achieved at the common drain point of the CMOS devices P_9 and N_9. The 100% negative feedback from the final X node to the inverting inputs of OTAs ensures unity gain voltage conversion within Y and X input nodes as well as very low-input impedance at the X terminal. The final X terminal is mirrored to get Z output node, comprised of CMOS transistors P_{10} and N_{10} and is achieved at their common drain node [15]. This circuit has utilized two-pMOS and two-nMOS current sources, one each as a tail current source of nMOS and pMOS DPs and one each in the combiner network. This circuit uses one diode connected pMOS (PB_1) and one diode connected nMOS (NB_1) with a bias resistor of RB_1 across the dual power supply (V_{DD} and V_{SS}) of values ± 0.6 V, to generate gate bias voltages of pMOS and nMOS current tail transistors, VB_P and VB_N, respectively, (see Fig. 2). The bias current is dependent on the power supplies, so well-regulated power supplies are desired. In the case of battery-operated systems, the dual power supply should be derived using some regulator sub-circuit. The complementary DPs in the input core of CCII+ ensure its rail-to-rail and low-voltage (LV) operation. In the output section of the circuit, there is a class AB unity gain buffer comprised of the CMOS devices P_{11}–P_{13} and N_{11}–N_{13} and which has been adapted from Ref. [13]. The common drain node of CMOS devices P_{13} and N_{13} configures the buffered output node (W) of the CFOA. The X, Z, and W nodes of this CFOA utilized the load resistor of 10 kΩ, not shown in Fig. 2.

2.1 The Basic Principle of CFOA

The input as well as output voltage and current variables of 4-terminal CFOA are V_Y, V_X, V_Z, V_W and I_Y, I_X, I_Z, I_W. The Y and Z nodes are associated with very high input/output impedance whereas X and W nodes offer a very low-input and output impedance [17]. It senses the output signal in the form of voltage at its buffered output terminal W which is low-impedance node. Ideally, voltage follower action between Y and X nodes ensures $V_Y = V_X$ and the current mirror action between X

and Z terminals ensures $I_Z = I_X$ whereas the voltage buffering action in between Z and W nodes ensures $V_W = V_Z$. The ideal relation between the input and output variable of CFOA is given by

$$
\begin{bmatrix} I_Y \\ V_X \\ I_Z \\ V_W \end{bmatrix} = \begin{bmatrix} 0\ 0\ 0\ 0 \\ 1\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0 \\ 0\ 0\ 1\ 0 \end{bmatrix} \begin{bmatrix} V_Y \\ I_X \\ V_Z \\ I_W \end{bmatrix}
\tag{1}
$$

which states that $I_Y = 0$, $V_X = V_Y$, $I_Z = I_X$, and $V_W = V_Z$. The impedances at Y input as well as Z output terminals is ∞ whereas the very low impedances are associated with node X and node W. The voltage gain (α), i.e., $V_X/V_Y = 1$, $V_W/V_Z = 1$, and the current gain (β), i.e., $I_Z/I_X = 1$. Under the influence of device mismatch in the fabrication process and threshold voltage differences even among various nMOS and pMOS transistors, the value of α and β gains may slightly be less than unity, say 0.998 [22–24]. The current gain β is given by

$$
\beta = \left(g_{mP_{10}} + g_{mN_{10}}\right)/\left(g_{mP_9} + g_{mN_9}\right)
\tag{2}
$$

The output impedance R_Z is given by

$$
R_Z = 1/\left(g_{dsP_{10}} + g_{dsN_{10}}\right)
\tag{3}
$$

whereas its output impedance at W terminal is given by

$$
R_W \approx 1/(g_{mN13} + g_{mP13})
\tag{4}
$$

The input impedance at X terminal is also very low owing to 100% shunt negative feedback at this terminal. The CFOA supports very high slew rates and bandwidth which is independent of its closed-loop voltage gain, so it is a suitable cell to design various low-voltage analog building blocks for high-frequency applications at reasonable power dissipation [22]. Table 1 shows the aspect ratios of the CMOS devices used in the circuit schematic this CFOA.

Table 1 Width (W) of CMOS devices used in CFOA schematic. Channel length each is 1 μm

nMOS	W (μm)	pMOS	W (μm)
N_1, N_2	20	P_1–P_4	10
N_5–N_8	4	P_5–P_6	40
N_9–N_{10}	25	P_7, P_8	10
N_{11}, N_{12}, N_{13}	5 and 25	P_9–P_{12}	50
NB_1–NB_3, N_3, N_4	5	PB_1–PB_3, P_{13}	10, 100

2.2   Simulation Results

The DC sweep result has been shown in Fig. 3. The input DC voltage, V_Y applied at Y node, has been swept in ±0.6 V range and resulting outputs at X, Z, and W nodes have shown a very good rail-to-rail operation (see Fig. 3). However, the response of W nodes shows slight deviation above 320 mV for the case of large input signal.

Figure 4 shows DC sweep response with input voltage applied to Y node and simulated output currents indicated at X, Z, and W nodes. It depicts very good current matching of I_X and I_W for input voltage range of -0.6 V to 420 mV. However, the current at W node shows good matching from -0.6 V to 260 mV. It depicts matched output current range of $-34\ \mu$A to 58 μA associated with W node and the matched current range of $-48\ \mu$A to 58 μA for currents associated with X and Z nodes.
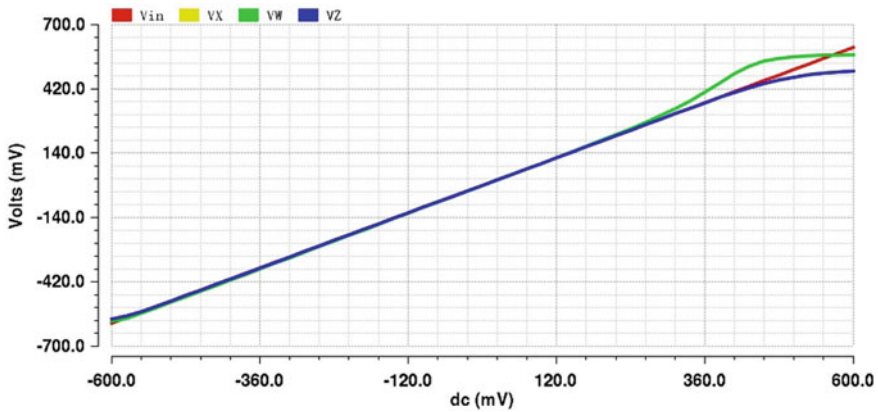
Fig. 3 DC voltage sweep with applied input to Y node and outputs at X, W, and Z nodes
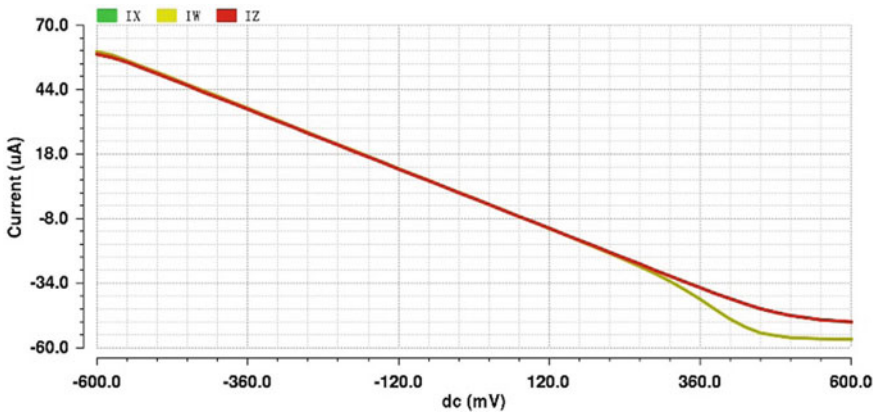
Fig. 4 DC sweep with input voltage to Y node and resulting currents at X, W, and Z nodes

The Fig. 5 shows AC response and voltage gains associated with X, Z, and W nodes when input V_{AC} has been applied to Y node and all X, Z, and W nodes are equally loaded by three resistors each of 10 kΩ value. It showed very good gain matching between X and W nodes for frequency up to 3 MHz. However, the gain associated with W node differs slightly by 65 mdB as compared to voltage gains concerned to X and Z nodes.

Table 2 shows some of the performance characteristics of the CFOA. It also shows the slew rates, 3-dB bandwidth, and total harmonic distortion (THD) associated with X, W, and Z nodes.
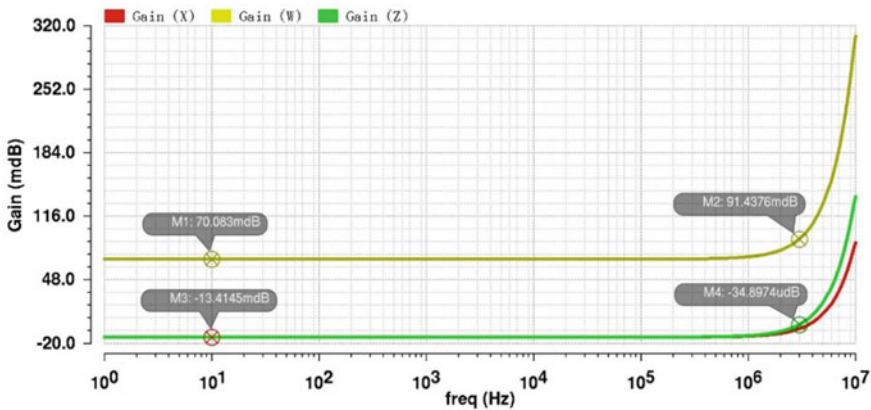
Fig. 5 AC response with V_{AC} to Y node and resulting voltage gains for X, W, and Z nodes

Table 2 Main characteristics (simulated results) of the CFOA shown in Fig. 2	Data	Value
	Voltage Supply	±0.6 V
	Total power consumption	377 μW
	3 dB bandwidth, at X, Z, W (MHz)	100.63, 102.56, 155.22
	Biasing current	10 μA
	Voltage gain (α), i.e., V_X/V_Y	0.998
	Voltage gain (V_W/V_Z)	0.99
	Current gain (β), i.e., I_Z/I_X	1.000571 and 0.9994
	X and Y node resistances	239 Ω, 3.74 MΩ
	Z and W node impedances	102.9 kΩ, 45.14 Ω
	Output offsets at X, Z, and W	145 μV, 145 μV, 1.38 mV
	Average slew rates at X, Z, and W nodes (V/μs)	55.3, 64.0, 498.0
	THD of X, Z, W nodes (%)	0.24, 0.24, 0.29

The simulated results listed in this table depicts that this CFOA has offered slew rate and 3-dB bandwidth of 498 V/μs and 155 MHz, respectively at its output (W) terminal. It dissipates a total power of 377 μW and offers voltage and current gain ratios very close to unity. The THD offered at X, Z, and W terminals are 0.24%, 0.24%, and 0.29%, respectively with a 100 mV amplitude of 1 kHz sine wave input at its Y terminal. Its input impedance at Y is very high, typically, Tera Ω order at DC frequency. Its impedance at X and W nodes are found to be 239 Ω and 45 Ω, respectively.

3   Biquadratic Universal Filter Using CFOA Cells

These low-voltage CFOA cells have been utilized to design a second-order voltage-mode SIMO type filters. A popular application of CFOA is to realize voltage-mode (VM) and current-mode (CM) biquadratic filters [14]. The circuit shown in Fig. 6 realizes all the five generic filter functions, namely: low-pass (LP), band-pass (BP), high-pass (HP), notch and all-pass (AP). This circuit generates all second-order voltage-mode filter functions based on SIMO approach.

The circuit has been adapted from the book referred in [14]. It has utilized two integrators made of CFOA$_2$ and CFOA$_3$ with two-grounded capacitors C$_1$, C$_2$ and two current summers comprised of CFOA$_1$ and CFOA$_2$ (see Fig. 6). Assuming ideal CFOAs, characterized as $I_y = 0$, $I_z = I_x$, $V_x = V_y$, and $V_w = V_z$ and using routine analysis, their transfer functions could be obtained as listed in Table 3 [14].

The low-pass, band-pass, high-pass, and all-pass functions are realized by the circuit at the corresponding nodes indicated as V$_{LP}$, V$_{BP}$, V$_{HP}$, and V$_{AP}$, respectively [14]. However, to realize the band-reject (BR) response, the switch shown in Fig. 6 must be set to BR position. The equal component design has been considered for simplicity which enabled to use all resistors of value 10 kΩ. However, the capacitors
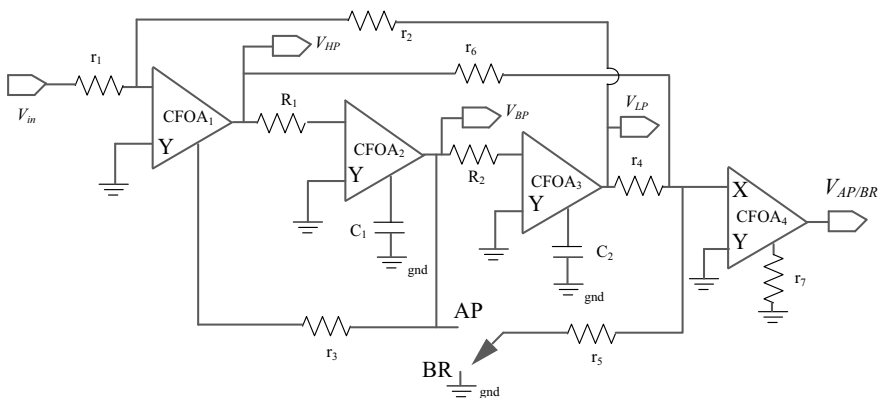
Fig. 6   Biquadratic universal SIMO type voltage-mode filters as adapted from [14]

Table 3 Transfer functions of voltage-mode SIMO type filter responses adapted from [14]

Filter type	Transfer function
Low-pass	$\dfrac{V_{01}}{V_{IN}} = \dfrac{H_0 \omega_0^2}{D(s)} = \dfrac{-\left(\frac{r_2}{r_1}\right)\frac{r_3}{r_2 R_1 C_1 R_2 C_2}}{s^2 + \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}}$
Band-pass	$\dfrac{V_{02}}{V_{IN}} = \dfrac{H_0 \frac{\omega_0}{Q_0} s}{D(s)} = \dfrac{\left(\frac{r_3}{r_1}\right)\frac{1}{R_1 C_1}s}{s^2 + \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}}$
High-pass	$\dfrac{V_{03}}{V_{IN}} = \dfrac{H_0 s^2}{D(s)} = \dfrac{-\left(\frac{r_3}{r_1}\right)s^2}{s^2 + \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}}$
All-pass	$\dfrac{V_{04}}{V_{IN}} = \dfrac{H_0\left(s^2 - \left(\frac{\omega_0}{Q_0}\right)s + \omega_0^2\right)}{D(s)} =$ $\dfrac{\left(\frac{r_3}{r_1}\right)\left(s^2 - \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}\right)}{s^2 + \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}}$
Band-reject	$\dfrac{V_{04}}{V_{IN}} = \dfrac{H_0\left(s^2 + \omega_0^2\right)}{D(s)} = \dfrac{\left(\frac{r_3}{r_1}\right)\left(s^2 + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}\right)}{s^2 + \frac{1}{R_1 C_1}s + \frac{r_3}{r_2 R_1 C_1 R_2 C_2}}$

$C_1 = C_2 = C$ has been swept in the range of 10 pF to 10 nF, which showed central and notch frequency range of 1.6 MHz to 1.6 kHz. This CFOA cell and filter circuit can also be utilized to design current-mode single input and single output (SISO) type filter [14], not considered due to space limitation in this work. For equal component design, the quality factor (Q) of band-pass and band-reject response is unity. The typical notations of lower cutoff frequency (f_L), higher cutoff frequency (f_H), central frequency (f_O), and notch frequency (f_N) are used in further description of this filter.

Figures 7 and 8 show all five filter responses when all resistors of 10 kΩ and two equal capacitors of 1nF have been used. The generated notch and mid-frequency are very closely matched to $f_O = f_N = 1/2\pi RC$ which is 15.9 kHz. Table 4 lists simulation results of four types of function, basically their cutoff frequencies, mid/notch
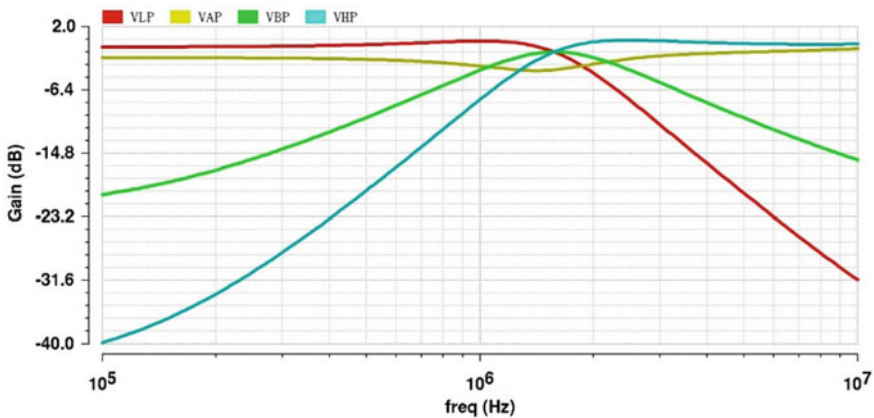
Fig. 7 Low-pass, high-pass, all-pass, and band-pass gain vs frequency plots with C = 1nF
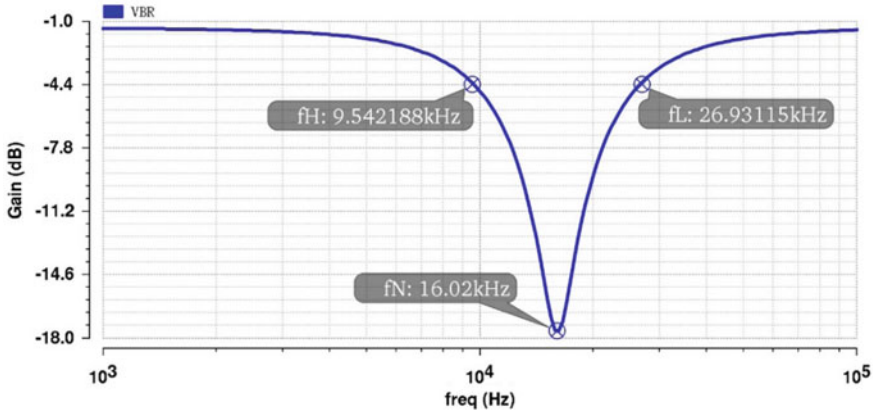
Fig. 8 SIMO, VM notch filter's response (f_N = 16 kHz) with all C of value 1nF

frequencies when capacitor C is considered of value 1nF, 10 nF, and 10 pF. The Q of notch filter, Q = f_O/BW = 16020/16289 = 0.98. This table also depicts that filter's corner, central, and notch frequencies are inversely proportional to capacitors used in this filter, as desired.

Figure 9 shows the transient response associated with X, Z, and W terminals when a 1 kHz sine wave of 100 mV amplitude is applied to Y node. This figure depicts that the voltage level at X, Z, and W terminals very closely matches with the Y input. The THD at X, Z, and W terminals has been found to be 0.24%, 0.24%, and 0.29%, respectively.

Figure 10 shows the pulse response concerned with output terminal (W) when input pulse of frequency 20 MHz, amplitude 0.6 V having rise and fall time of 1 ps has been applied to the Y input node. The pulse response has ensured the positive and negative slew rates of 545 V/μs and 451 V/μs, respectively. These slew rates are very high as compared to voltage-mode operational amplifiers.

4 Performance Comparison

The performances of this CFOA have been compared by research work referred in [22–24] as shown in Table 5.

The figure of merit (FoM) called power to bandwidth efficiency is used for the performance comparison as shown in (5). It dictates the ratio of total power dissipation to 3 dB bandwidth (concerned to W terminal) in the unit of pW/Hz.

$$FoM(\eta) = Power\ dissiption/Band\ Width\ (pW/Hz) \tag{5}$$

Table 4 Lower, higher cutoff frequencies, central and notch values of biquadratic filter

Type of Filter response	For $C_1 = C_2 = 1$ nF			For $C_1 = C_2 = 10$ nF			For $C_1 = C_2 = 10$ pF		
	f_L (kHz)	f_H (kHz)	f_O (kHz)	f_L (kHz)	f_H (kHz)	f_O (kHz)	f_L (kHz)	f_H (kHz)	f_O (MHz)
Low-pass	–	19.58	–	–	1.958	–	–	1935	–
High-pass	13.15	–	–	1.315	–	–	1325	–	–
Band-pass	9.505	27.05	16.02	1.067	2.413	1.602	938.214	2673	1.58
Band-reject	26.93	9.54	16.02	2.67	0.962	1.602	2630	953.037	1.58
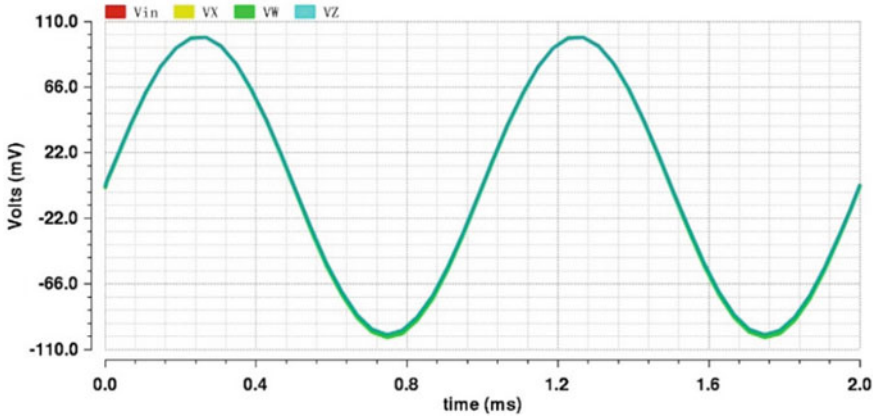
Fig. 9 Transient response with 1 kHz pulse of 100 mV amplitude
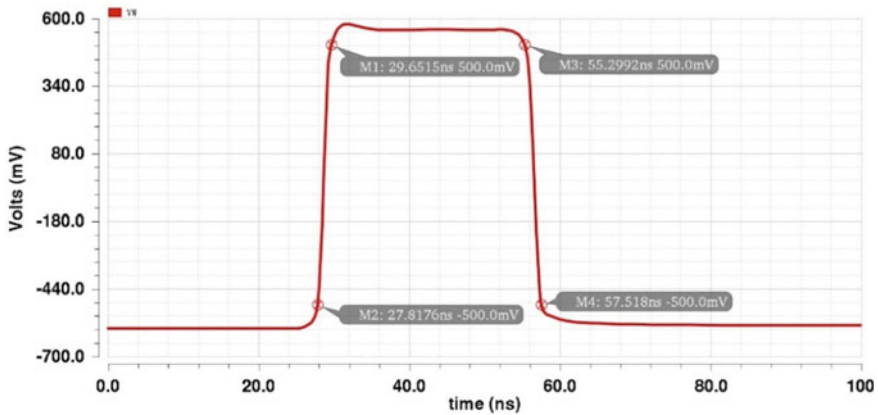
Fig. 10 Pulse response of W node with input pulse of 20 MHz and 0.6 V amplitude to Y

It can be concluded that this CFOA has provided better power to bandwidth ratio as compared to [23, 24] and circuit-3 of [22], whereas this parameter is slightly worse than circuit-1 and circuit-2, addressed in Ref. [22]. It has offered the best average R_Z of 1.03 MΩ (for sweep range of 1 Hz to 10 MHz) as compared to others. Though its power dissipation is more than [23], its power to frequency conversion ratio (pW/Hz) is better than CFOAs addressed in Ref. [23]. Thus, this CFOA cell is a suitable choice to design filters, oscillators, instrumentation amplifiers (IAs), etc., requiring moderate power consumption.

Table 5 Performance comparison of this CFOA with some referred work [22–24]

Parameters	This work	[22] Circuit-1	[22] Circuit-2	[22] Circuit-3	[23] CFOA	[23] DVCFOA	[24] CFOA
Power Supply (V)	±0.6	±0.75	±0.75	±0.75	1.5	1.5	±1.25
BW(V_X/V_Y), MHz	100.6	267.7	134.6	70.4	13.8	9.5	592
BW (I_Z/I_X), MHz	102.56	806.8	149.9	140.8	28.4	29.3	335
BW (V_W/V_Z), MHz	155.22	335.3	291.1	104.1	27	22	574
I_{Bias} (μA)	10	50	50	50	10	10	NA
R_X (Ohm)	239	25.4	39.4	113.4	51.5	103	4
R_Y (MΩ)	∞	∞	∞	∞	∞	∞	∞
R_Z (MΩ)	1.029	0.193	0.183	0.085	1.0	1.0	0.054
R_W (Ω)	45.14	25.4	39.4	113.4	51.5	103	NA
Gain (V_X/V_Y)	0.998	0.9948	1.018	0.956	0.99	0.99	0.989
Gain (V_W/V_Y)	0.99	0.9948	1.027	0.961	1	1	0.989
Gain (I_Z/I_X)	1	1	1.073	1.074	1	1	1.024
THD of V_W (%)	0.29	1.1	10	21.5	0.65	1	NA
THD of V_X (%)	0.24	0.6	4.4	2.4	0.4	0.7	NA
Power (μW)	377	552	544	756	110	140	1330
FoM(η) (pW/Hz)	2.432	1.6463	1.8687	7.2622	4.07	6.36	2.317

5 Conclusion

In this paper, dual symmetric (balanced) current mirror load OTAs based CCII+ which is followed by an output buffer is utilized to configure a low-voltage CMOS CFOA. This simple structure provides very close rail-to-rail input swing and low-output offset errors. It is concluded from the simulation results that the proposed amplifier provides very good CFOA features up to 100 MHz frequency as well as very high slew rate of 350 V/μs associated with output node W. The designed voltage-mode SIMO filter provides varying central frequencies ranging in 1.6 kHz up to 1.6 MHz for its filter capacitors ranging in 10 nF to 10 pF, respectively. It contains gate-driven input pair-based CMOS OTAs and works well with reduced dual power supply of ±0.6 V as compared to other gate-driven CFOA circuits which utilized

± 0.75 V supply. This CFOA cell is a suitable choice to design voltage-mode and current-mode filters, rectifiers, frequency multipliers, oscillators, etc.

References

1. Sedra, A.S., Smith, K.C.: A second generation current conveyor and its applications. IEEE Trans. Circuit Theory 17, 132–134 (1970)
2. Toumazou, C., Lidgey, F., Haigh, D.: Analogue IC Design: The Current-Mode Approach, 1990th edn. IEE Press, Edition (1990)
3. Sedra, A.S., Roberts, G.W., Gohh, F.: The current conveyor. History, progress and new results. IEE Proc, Part G: Electron. Circuits Syst. **137**, 78–87 (1990)
4. Franco, S.: Analytical foundations of current-feedback amplifiers. IEEE Int. Symp. Circuits Syst. **2**, 1050–1053 (1993)
5. Huelsman, L.P.: Active and Passive Analog Filter Design. McGraw-Hill. Inc, New York (1993)
6. Chang, C.M., Hwang, C.S., Tu, S.H.: Voltage-mode notch, low pass and band pass filter using current-feedback amplifiers. Electron. Lett. **30**, 2022–2023 (1994)
7. Senani, R., Gupta, S.S.: Universal voltage-mode/current-mode biquad filter realised with current feedback op-amps. Frequenz **51**, 203–208 (1997)
8. Abuelma'atti, M.T., Al-Zaher, H.A.: New universal filter with one input and five outputs using current-feedback amplifiers. Analog Integr. Circuits Signal Process. **16**, 239–244 (1998)
9. Senani, R.: Realization of a class of analog signal processing/signal generation circuits: novel configurations using current feedback op-amps. Frequenz **52**, 196–206 (1998)
10. Hou, C.L., Huang, C.C., Lan, Y.S., Shaw, J.J., Chang, C.M.: Current-mode and voltage-mode universal biquads using a single current-feedback amplifier. Int. J. Electron **86**, 929–932 (1999)
11. Mahmoud, S.A., Elwan, H.O., Soliman, A.M.: Low voltage rail to rail CMOS current feedback operational amplifier and its applications for analog VLSI. Analog Integr. Circ. Sig. Process **25**, 47–57 (2000)
12. Ferri, G., Guerrini, N.: High-valued passive element simulation using low-voltage low-power current conveyors for fully integrated applications. IEEE Trans. Circ. Syst. II Analog Digital Signal Process. **48**, 405–409 (2001)
13. Maundy, B.J., Finvers I.G., Aronhime, P.: Alternative realizations of CMOS current-feedback amplifiers for low-voltage applications. Anal. Int. Circ. Signal Process. **32**, 157–168 (2002)
14. Senani, R., Bhaskar, D.R., Singh, A.K., Singh, V.K.: Current Mode Operational Amplifier and Their Applications. Springer, ACSP series, New York (2013)
15. Ferri, G., Guerrini, N.C.: Low-Voltage Low-Power CMOS Current conveyors. Kluwer Academic Publishers Boston, pp. 59–93 (2003)
16. Sharma, R.K., Senani, R.: Universal current mode biquad using a single CFOA. Int. J. Electron. **91**, 175–183 (2004)
17. Singh, A.K., Senani, R.: CFOA-based state-variable biquad and its high-frequency compensation. IEICE Electron Express **2**, 232–238 (2005)
18. Maundy, B.J., Sarkar, A., Gift, S.J.G.: A new design topology for low-voltage CMOS current feedback amplifiers. IEEE Trans. Circuits Syst. II Express Briefs **53**, 34–38 (2006)
19. Singh, V.K., Singh, A.K., Bhaskar, D.R., Senani, R.: New universal biquads employing CFOAs. IEEE Trans. Circ. Syst II **53**, 1299–1303 (2006)
20. Bhaskar, D.R., Prasad, D.: New current mode biquad filter using CFOAs. J Active Passive Electron Devices **2**, 292–298 (2007)

21. Johns, D., Martin, K.: Analog Integrated Circuit Design, pp. 531–573. Wiley (1997)
22. Rakios, G., Costas, P.: Low-voltage current feedback operational amplifiers. Circuits Syst. Signal Process **28**, 377–388 (2009)
23. Samiotis, P., Psychalinos, C.: Design of filters with only grounded passive elements using differential voltage current feedback differential operational amplifiers. J. Circuits Syst. Comput. **19**, 573–580 (2010)
24. Yuce, E., Shahram, S., Ibrahim, M.A.: A novel full-wave rectifier/sinusoidal frequency doubler topology based on CFOAs. Analog Integr. Circ. Signal Process. (2017). https://doi.org/10.1007/s10470-017-1033-0

Modeling and FEM-Based Simulations of Composite Membrane Based Circular Capacitive Pressure Sensor

Rishabh Bhooshan Mishra, S. Santosh Kumar and Ravindra Mukhiya

Abstract In Micro-electro-mechanical Systems (MEMS) based pressure sensors and acoustic devices, deflection of a membrane is utilized for pressure or sound measurements. Due to advantages of capacitive pressure sensor over piezoresistive pressure sensors (low power consumption, less sensitive to temperature drift, higher dynamic range, and high sensitivity), capacitive pressure sensors are the second largest useable MEMS-based sensors after piezoresistive pressure sensors. We present a normal capacitive pressure sensor, for continuous sensing of normal and abnormal Intraocular Pressure (IOP). The composite membrane of the sensor is made of three materials, i.e., Si, SiO_2, and Si_3N_4. The membrane deflection, capacitance variation, mechanical sensitivity, capacitive sensitivity, and nonlinearity are discussed in this work. Mathematical modeling is performed for analytical simulation, which is also compared with Finite Element Method (FEM) simulations. MATLAB® is used for analytical simulations and CoventorWare® is used for FEM simulations. The variation in the analytical result of deflection in membrane w.r.t. FEM result is about 7.19%, and for capacitance, the variation is about 2.7% at maximum pressure of 8 kPa. The nonlinearity is about 4.2492% for the proposed sensor for fabrication using surface micro-machining process.

Keywords Circular composite membrane · Capacitive pressure sensor · Mathematical modeling · FEM simulations

1 Introduction

Micro-machined sensors are integrated with an electrical interface to make electromechanical systems. MEMS sensors are systems which interact with measurands (like displacement, acceleration, flow, pressure, temperature, etc.) and then convert it to an electrical signal, which is used to analyze the measurand so that further con-

R. B. Mishra · S. Santosh Kumar (✉) · R. Mukhiya
Smart Sensor Area, CSIR-Central Electronics Engineering Research Institute (CEERI), Pilani 333031, Rajasthan, India
e-mail: santoshkumar.ceeri@gmail.com

trolling, monitoring, and/or alarming actions can take place. ICs are used to perform signal conditioning so that the obtained signal can be used for further decision-making/transmission/communication [1]. The market for MEMS-based devices is growing at a fast rate from several years. In the sensors domain, MEMS-based inertial sensors, pressure sensors, and micro-actuators have large applications not only for consumer electronics, defense systems, and automobile application but also for biomedical applications. In some biomedical applications, the devices can be implanted into living-beings, and then biomedical signal can be transmitted using wireless communication technology, i.e., telemetry [2–4].

MEMS-based pressure sensors have replaced conventionally available pressure measuring devices like bourdon tubes, bellows, diaphragms, capsules, and various vacuum measuring devices (such as Pirani gauge and McLeod gauge.). In conventional pressure measuring devices, the pressure causes mechanical movement that rotates the pointers/dial. However, MEMS devices directly convert input pressure into corresponding electrical signal. MEMS pressure sensors are fabricated on silicon wafers (SOI, double SOI or single crystal silicon wafers) using surface, bulk or a combination of these two micro-machining techniques. Utilizing the abovementioned technologies, the diaphragms are released which acts as a sensing element [5].

Increase or fluctuation in IOP can be the cause of glaucoma. According to the World Health Organization (WHO), glaucoma is one of the major causes of blindness. Glaucoma causes irreversible eye disease, which damages the optical nerves. The normal range of IOP is 1.6–2.8 kPa. Therefore, the accurate measurement in early stage can save the eye from permanent blindness [2, 3].

The normal mode capacitive pressure sensors have a fixed plate and a movable (usually conductive) plate/membrane, which are separated by a medium, i.e., vacuum, air or dielectric materials. The capacitance variation can be obtained by following three techniques (usually the first technique is utilized for designing MEMS capacitive pressure sensor):

- Changing the separation gap between parallel plates.
- Changing the overlapping area between parallel plates.
- Movement changing in dielectric materials which is filled between plates.

This paper presents mathematical/theoretical modeling and FEM simulation of composite membrane-based normal mode capacitive pressure sensor for IOP measurement (0–8 kPa). The obtained results are also compared with one of our previous works, in which the modeling and FEM simulation of normal mode capacitive pressure sensor is carried out for the same application. In that work, only the silicon material of Young modulus of elasticity 169 GPa and Poisson ratio of 0.066 is used as diaphragm material. For different diaphragm thicknesses after optimization, comparison of theoretical and simulation results is presented [3]. In this presented work, the composite membrane is made of silicon (with same material properties), silicon dioxide (Young's modulus of elasticity 70 GPa and Poisson ratio 0.17), and silicon nitride (Young's modulus of elasticity 222 GPa and Poisson ratio 0.27).

2   Mathematical Modeling of Sensor

2.1   Membrane Deflection and Capacitance Variation

After application of pressure P on thin, clamped, and flat circular diaphragm of radius
a, thickness *t* which is made of homogeneous, isotropic, and elastic material with
Young's modulus of rigidity E and Poisson ratio *v*, the diaphragm deflection can be
given by [3]

$$w(r) = \frac{Pa^4}{64\,D}\left[1 - \left(\frac{r}{a}\right)^2\right]^2 \tag{1}$$

Here, D (flexural rigidity) can be given by

$$D = \frac{Et^3}{12(1-v^2)} \tag{2}$$

The base capacitance of the sensor can be given by

$$C_{base} = \frac{\varepsilon A}{d} \tag{3}$$

Here, A, d, and ε are the overlapping areas between plates, separation gap, and
permittivity of medium, respectively.

To find the capacitance of parallel plate capacitor after pressure application, cut
an annulus at radius r of width dr from the deflected clamped circular diaphragm, as
shown in Fig. 1. The capacitance due to this annulus (a small element) can be given
by
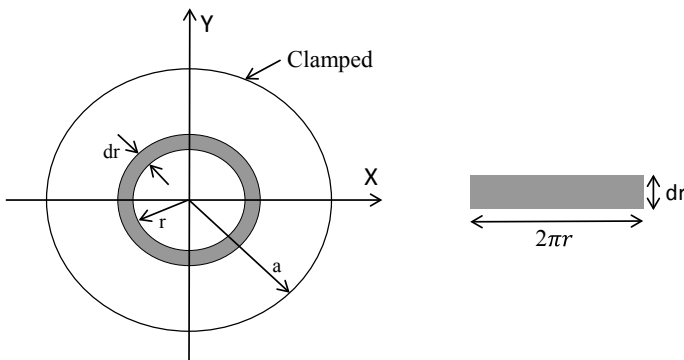
Fig. 1 Circular clamped diaphragm depicting an annulus is taken from composite membrane

$$\partial C_d = \frac{\varepsilon(2\pi r dr)}{d - w(r)} \tag{4}$$

After performing the integration over the whole area of the sensor, the capacitance can be given by

$$C_d = \int_0^a \frac{\varepsilon(2\pi r dr)}{d - \frac{Pa^4}{64\,D}\left[1 - \left(\frac{r}{a}\right)^2\right]^2} \tag{5}$$

After solving the above equation [3]:

$$C_w = 4\pi\varepsilon\sqrt{\frac{D}{Pd}}\ln\left|\frac{a^2\sqrt{P} + 8\sqrt{dD}}{a^2\sqrt{P} - 8\sqrt{dD}}\right| \tag{6}$$

If the diaphragm material is made of three different materials silicon, silicon dioxide, and silicon nitride [$Si_3N_4(E_1, v_1, h_1)/SiO_2(E_2, v_2, h_2)/Si(E_3, v_3, h_3)$],as shown in Fig. 2, then flexural rigidity of the composite membrane can be given by [5]

$$D_c = \frac{E_1\left[(h-e)^3 - (h-e-h_1)^3\right]}{3(1-v_1{}^2)} + \frac{E_2\left[(h-e-h_1)^3 - (h_3-e)^3\right]}{3(1-v_2{}^2)} + \frac{E_3\left[(h_3-e)^3 + e^3\right]}{3(1-v_3{}^2)} \tag{7}$$

Here, e is the neutral plane (the plane within the composite plate/membrane which is not in tension, compression or stress when there is application of pressure on the plate/membrane), this can be given by [5]

$$e = \frac{\frac{E_1}{1-v_1}h_1(h_1 + 2h_2 + 2h_3) + \frac{E_2}{1-v_2}h_2(h_2 + 2h_3) + \frac{E_2}{1-v_3}h_3{}^2}{2\left[\frac{E_1}{1-v_1}h_1 + \frac{E_2}{1-v_2}h_2 + \frac{E_2}{1-v_3}h_3\right]} \tag{8}$$

Then deflection in the composite membrane, of radius L, flexural rigidity, and thickness can be given by

$$W_{\text{composite}}(r) = \frac{Pa^4}{64\,D_c}\left[1 - \left(\frac{r}{a}\right)^2\right]^2 \tag{9}$$

Fig. 2 Composite membrane made of three different materials
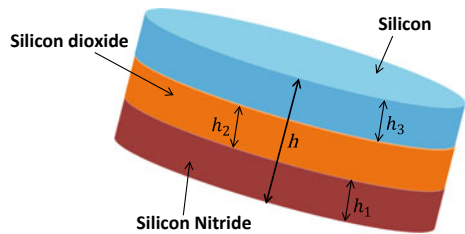
The capacitance variation due to pressure application is obtained by modifying Eq. (4) and can be given by

$$C_{w\_composite} = 4\pi\varepsilon\sqrt{\frac{D_c}{Pd}}\ln\left|\frac{a^2\sqrt{P}+8\sqrt{dD_c}}{a^2\sqrt{P}-8\sqrt{dD_c}}\right| \tag{10}$$

2.2 Sensitivity and Nonlinearity

The mechanical sensitivity is a useful parameter, if maximum membrane deflection varies in different designs. The mechanical sensitivity is the slope of maximum deflection versus the pressure range curve. The mechanical sensitivity of the composite membrane can be given by [2, 3]

$$S_{mech,composite} = \frac{a^4}{64\,D_c} \tag{11}$$

The capacitive sensitivity of composite membrane-based sensor is obtained by the ratio of change in capacitance and the applied pressure range [2–4]:

$$S_{cap,composite} = \frac{C_{max} - C_{min}}{P_{max} - P_{min}} \tag{12}$$

Mechanical and capacitive sensitivity, both are terms which define the performance and specifications of sensors.

The nonlinearity of the sensor, at a particular point, can be defined by [6]

$$NL_i(\%) = \frac{C_{d\_composite} - C_{d\_composite} \times \frac{P_i}{P_m}}{C_{d\_composite}} \times 100 \tag{13}$$

Here, P_i is applied pressure at any point on the calibrated curve, P_m is maximum pressure, $C_{d\_composite}(P_i)$ is capacitance at a particular point on the calibrated curve and $C_{d\_composite}(P_m)$ is the capacitance at maximum pressure.

3 Results and Discussion

The sensor must have good sensitivity, minimum nonlinearity, low power consumption, robustness, and small size. Before fabrication of the pressure sensor, several steps need to be carried out like designing, mathematical formulation, verification of mathematical modeling, analytical simulations, and comparison of analytical simula-

tion with FEM simulation. According to the mathematical formulation and analytical simulation, the optimized design of the sensor can be obtained. However, in modeling and analytical simulation, several assumptions need to be taken. Therefore, the FEM simulation is performed to analyze the behavior of the sensor in a practical situation. The optimization of device using fabrication runs is complex, costly, time-taking, and needs more effort. Hence, it is highly desirable to have accurate and precise design and modeling. Considering the practical scenario, in this work, we have modeled composite membrane and verified the modeling with FEM simulations.

In a capacitive pressure sensor, base capacitance must be in pF range and change must be in fF range for the capacitance measurement to be performed effectively and efficiently. Therefore, we should try to increase base capacitance and change in capacitance by choosing appropriate and optimized design parameters for sensors. If the overlapping area is increased for a particular thickness of the membrane, then deflection of membrane increases, so that mechanical sensitivity as well as capacitive sensitivity increase. If thickness of membrane is increased by keeping overlapping area between plates constant/same, then deflection in diaphragm decreases, so that both mechanical and capacitive sensitivity decrease.

3.1 Analytical Design for Membrane Deflection

The comparison of diaphragm deflection in 6 μm thick silicon diaphragm and composite diaphragm of the same thickness is performed. In all the designs, composite membranes have a thickness of 6 μm (t). Silicon thickness is kept 5.3 μm thick and thickness of SiO_2 and Si_3N_4 are varied as shown in Table 1. The deflection in various composite diaphragms and in the silicon diaphragm with the same overall thickness is shown in Fig. 3. The radius of the diaphragm is 360 μm and the applied pressure is 8 kPa. The flexural rigidity and maximum deflection in the membrane of these four membranes are indicated in Table 1.

According to Table 1 and Fig. 3, if flexural rigidity of the membrane is larger, deflection is lesser; therefore, the mechanical sensitivity is less. Since the flexural rigidity of Model No. 4 is less and deflection is maximum, it is chosen for FEM simulation. The radius is modified according to separation so that sensor can be fabricated for IOP measurements using appropriate process flow.

Table 1 Flexural rigidity and maximum deflection for different models

Model no.	Composite diaphragm thicknesses specification (μm)	Flexural rigidity (Pa cm^3)	Maximum deflection (μm)
1	t = 6	3.0553	0.68717
2	$h_1 = 0.2, h_2 = 0.5, h_3 = 5.3$	2.7556	0.76192
3	$h_1 = 0.15, h_2 = 0.55, h_3 = 5.3$	2.6948	0.77910
4	$h_1 = 0.1, h_2 = 0.6, h_3 = 5.3$	2.6312	0.79793

Fig. 3 Membrane deflection of different thicknesses
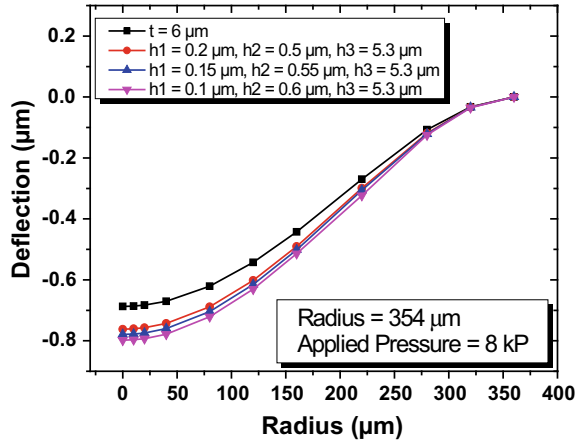
3.2 FEM Simulations and Comparison with Analytical Results

While performing FEM simulations using CoventorWare®, composite membranes are merged. Then membrane edges are clamped and the bottom plate is fixed. The pressure is applied to the top of the membrane. The "Tetrahedron" meshing type of parabolic element order is used to mesh the model of sensor with element size = 5.

The MemMech module and "Mechanical" type of physics are utilized for mechanical analysis, i.e., for obtaining deflection in membrane.

The FEM result of composite membrane deflection of thicknesses $h_1 = 0.1\,\mu m$, $h_2 = 0.6\,\mu m$, $h_3 = 5.3\,\mu m$ and radius $354\,\mu m$ is shown in Fig. 4.
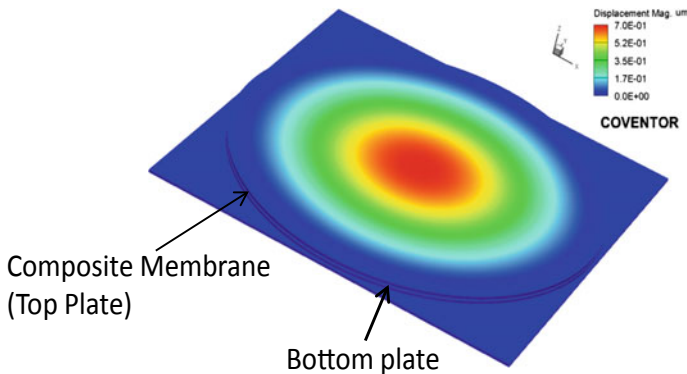
Fig. 4 FEM result of deflection in composite membrane-based circular-shaped capacitive pressure sensor at 8 kPa pressure using CoventorWare®

Fig. 5 Analytical and FEM results of maximum deflection w.r.t. pressure variation
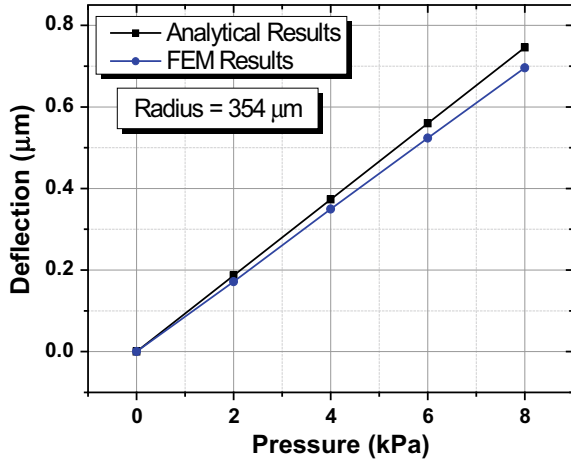
The comparison of analytical and FEM results are close to each other which is shown in Fig. 5. The variation in analytical and FEM results for membrane deflection increases as applied pressure increases.

The MemElectro module and "Electrostatics" type of physics is utilized for obtaining base capacitance. The bottom plate is grounded and 1 V voltage is applied on the top conductor. To obtain the capacitance at a particular applied pressure, the CoSolveEM module is utilized in which Surface_BCs and DC_ConductorBCs are used to define the boundary conditions. In CoSolveEM module, mechanical as well as electrical analysis can be performed at a time.

The base capacitance of sensor according to analytical simulation is 1.1619 and 1.206062 pF according to FEM simulation. The analytical and FEM result of capacitance variation w.r.t. applied pressure for membrane with a radius of 354 μm is shown in Fig. 6.

The mechanical sensitivity of the sensor according to analytical and FEM simulation is 93.25625 nm/kPa and 86.999075 nm/kPa, respectively. The capacitive sensitivity according to analytical and FEM simulations is 14.2375 fF/kPa and 13.1535 fF/kPa, respectively.

4 Conclusion

The deflection, capacitance variation, and sensitivity (mechanical and capacitive both) of composite membrane-based capacitive pressure sensor is discussed with mathematical modeling and simulations. The membrane of the proposed design is made of three different materials namely silicon, silicon dioxide, and silicon nitride with thicknesses of 5.3 μm, 0.6 μm, and 0.1 μm, respectively. The deflection curve w.r.t. applied pressure is obtained linear which validates the Hook's Law. The ana-

Fig. 6 Analytical and FEM results of capacitance variation w.r.t. pressure variation
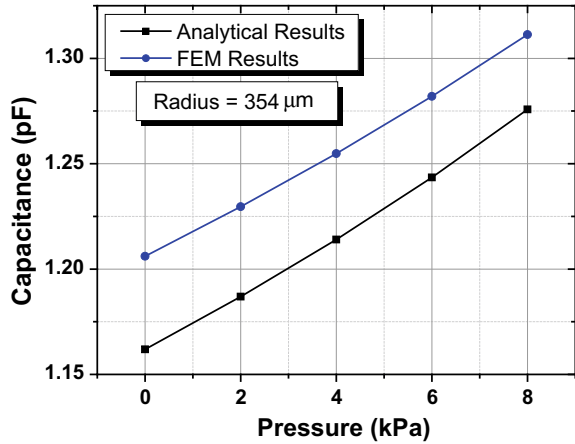
lytical and FEM simulation result of capacitance variation w.r.t. applied pressure is also determined. The sensor has 4.2492% nonlinearity. The variation in FEM results with analytical result in membrane deflation curve is due to Kirchhoff's assumptions which have been considered in mathematical modeling/formulation. In mathematical formulations and analytical simulations of base capacitance and capacitance variation, the fringing field, parasitic effects plus dielectric constant of silicon oxide and silicon nitride (i.e., 3.9 and 8, respectively) are not considered which are leading to the variation in analytical and FEM results of capacitance variation w.r.t. applied pressure.

References

1. Fischer, A.C., Forsberg, F., Lapisa, M., Bleiker, S.J., Stemme, G., Roxhed, N., Niklaus, F.: Integrating MEMS and ICs. Microsyst. Nanoeng. Nat. (2015)
2. Mishra, R.B., Kumar, S.S., Mukhiya, R.: Design and simulation of capacitive pressure sensor for blood pressure sensing application. In: IC3E. University of Allahabad, India (2018)
3. Mishra, R.B., Kumar, S.S., Mukhiya, R.: Analytical modelling and FEM simulation of capacitive pressure sensor for intraocular pressure sensing. In: ICCRME, vol. 404, pp. 012026. IOP Conference Series: Material Science and Engineering, India (2018). https://doi.org/10.1088/1757-899x/404/1/012026
4. Mishra, R.B., Kumar, S.S.: Pre-stressed diaphragm based capacitive pressure sensor for blood pressure sensing application. In: IAC3T. University of Allahabad, India (2018)

5. Nie, M., Bao, H.Q.: A theoretical model and analysis of composite membrane of a piezoresistive pressure sensor. AIP Adv. (2016)
6. Kumar, S.S., Pant, B.D.: Polysilicon thin film piezoresistive pressure microsensor: design, fabrication and characterization. Microsyst. Technol. (Springer) **21**(9), 1949–1958 (2015). https://doi.org/10.1007/s00542-014-2318-1

Comparative Study on Structural and Electrical Characteristics of TiO$_2$ Film Deposited by Plasma-Enhanced Atomic Layer Deposition and RF Sputtering

Rajesh Kumar Jha ⬤**, Prashant Singh** ⬤**, Manish Goswami and B. R. Singh**

Abstract Titanium dioxide film has been deposited on the p-type (100) silicon substrate using RF magnetron sputtering and Plasma-Enhanced Atomic Layer Deposition (PEALD) techniques. The effect of deposition techniques and the post-deposition annealing on the structural and electrical properties of TiO$_2$ film have been investigated. Multiple angle ellipsometry, X-Ray diffraction (XRD) analysis, Capacitance–Voltage (C-V), Current density–Voltage (J-V), and breakdown voltage investigations has been carried out to obtain the structural and electrical characteristics of the deposited films. XRD data illustrates the amorphous nature of the deposited film for the annealing temperature from 375 to 500 °C. Multiple angle ellipsometry results show that the PEALD and sputtered films refractive index variation from 2.0463 to 2.1348 and 2.21 to 2.75, respectively, for annealing temperature 375 to 500 °C. Electrical characteristics show the leakage current density from 10^{-5} to 10^{-6} A/cm^2 and flatband voltage shift in the positive side of 2.4 V for PEALD and 0.2 V for sputter-deposited films.

Keywords PEALD · RF sputtering · Multiple angle ellipsometry · X-Ray Diffraction (XRD) analysis · CV · IV · Breakdown

1 Introduction

Titanium dioxide is a high-k dielectric material that possesses excellent optical [1] and electrical properties [2], is deposited by different techniques such as pulsed laser deposition [3], MOCVD [4], sputtering [5], Sol-Gel [6], etc., and has been extensively studied in the past few decades. Since the beginning of semiconductor technology, well-established process technology and the thermodynamically stable electrical as well as structural properties of SiO$_2$ have served as the most preferred

R. K. Jha · P. Singh (✉) · M. Goswami · B. R. Singh
Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Allahabad 211015, Uttar Pradesh, India
e-mail: psingh3688@gmail.com

507

gate dielectric. Since 1965, following Moore's law, the continuous scaling in CMOS-based memory technologies has brought the SiO_2 to its verge. As the scaling trend forces the thickness of gate oxide (SiO_2) down to ~2 nm, the direct tunneling leakage current increases to 1 A/cm^2 even at a gate voltage of 1 V. Such a large leakage current raises the concern of excessive standby power consumption in the integrated circuits. The limitation of SiO_2 physical thickness in the VLSI technology leads to the discovery of alternative dielectric materials called high-k as the possible substitute for SiO_2 [7]. Various high dielectric constant materials such as SiNx (7), Al_2O_3 (9–11), HfO_2 (25), ZrO_2 (25), and La_2O_3 (30) have been actively investigated in the last few decades as a gate dielectric. Previously, TiO_2 deposited by MOCVD has been studied as the gate dielectric in the MOSFET by S. A. Campbell et al. 1997 [8] and as the buffer layer in the Ferroelectric FET by Dan Xie et al. 2010 [9]. With the development in deposition techniques such as Atomic Layer Deposition, it is now possible to achieve high-quality and thermodynamically stable TiO_2–Si interface, an essential requirement of the semiconductor and ferroelectric FET devices [10]. In the ferroelectric FETs, it is difficult to form high-quality thermodynamically stable ferroelectric–silicon interface, which results in large gate leakage and ferroelectric–silicon interdiffusion. To overcome these problems, FeFET with an additional high-k dielectric layer incorporated between the ferroelectric layer and silicon substrate was proposed. This high-k layer named as buffer layer inhibits the interdiffusion and increases the data retention capacity significantly [11]. The important requirements for the buffer layer are high dielectric constant, high bandgap and band offset, and good thermal stability with silicon [12].

To facilitate the comparison of the gate insulator deposited by sputtering and PEALD techniques, the structural and electrical characteristics of the TiO_2 film at different process parameters have been analyzed. Quality of dielectric films was evaluated on the basis of refractive index, XRD pattern, capacitance, leakage current density, and breakdown voltage.

2 Experimental Procedure

The silicon wafer used in this work was p-type <100> with 275 μm thickness and having resistivity 1–10 Ω-cm. A standard cleaning process called Radio Corporation of America (RCA-1 and RCA-2) followed by the buffered Hf dip and nitrogen drying was carried out to get the impurity-free substrate. The deposition parameters of RF sputtering and PEALD are compiled in Table 1.

The annealing process was done after the TiO_2 film deposition at different temperatures in the nitrogen gas ambient for 30 min in a quartz tube equipped furnace. After the annealing, ellipsometric analysis was carried out using multiple angle ellipsometer having HeNe laser of wavelength 632.8 nm, procured from M/s SENTECH Instruments GmbH to obtain the thickness and the refractive index of the deposited film. Film crystal orientation was investigated using 1.54 Å Cu-Kα laser source. MIS

Table 1 Process parameters of TiO$_2$ deposited by PEALD and sputtering

Deposition technique	Plasma-enhanced atomic layer deposition	RF magnetron sputtering
Process material	98+% pure Titanium (IV) i-propoxide precursor	99.99% pure TiO$_2$ target
Base pressure	3.3 Pa (3.3 $\times$ 10^{-2} mbar)	5 $\times$ 10^{-5} mbar
Working pressure	27.13 Pa (27.13 $\times$ 10^{-2} mbar)	1.1 $\times$ 10^{-2} mbar
Argon gas flow rate	–	19.6 sccm
Oxygen gas flow rate	75 sccm	–
RF power (W)	200	100
Substrate temperature (°C)	200	200

structures were fabricated with the top circular metal electrode of area 3.14×10^{-4} cm^2, deposited by thermally evaporating aluminum wire with the purity of 99% using a metal mask with a circular opening. Devices were electrically characterized using semiconductor characterization system, procured from M/s Keithley Instruments, USA.

3   Results and Discussion

Figure 1 shows the XRD pattern of TiO$_2$ film deposited by PEALD and sputtering annealed at different temperatures.

XRD results show the only peak at $2\theta \approx 51°$ representing the silicon substrate. No other significant intensity peaks were observed in the films, attributed to the amorphous film structure throughout the selected annealing temperature range. Amorphous film serves as an excellent buffer layer in field-effect transistor and the passivation layer in the solar cells. The annealed films were characterized to obtain the refractive index summarized in Table 2.

Variation in annealing temperature from 375 to 500 °C shows the refractive index changing from 2.0463 to 2.1348 for PEALD and from 2.21 to 2.75 for sputter-deposited TiO$_2$ films under the conditions summarized in Table I. Film deposited by sputtering shows higher refractive index indicating higher density as compared to the PEALD films [13]. Perhaps in sputtered TiO$_2$, high-energy neutral ion bombardment during deposition provides local rearrangement of atoms allowing them to relax into lower energy sites which results in high density but also causes residual damage to the film. This residual damage can be further minimized by annealing [11]. Film thickness and deposition rates in PEALD and sputtering are directly dependent on the process parameters, as shown in Fig. 2.

The deposition rate of TiO$_2$ was 0.7143 nm per cycle in the PEALD process at the substrate temperature of 200 °C and RF power of 200 W. The deposition rate was found to be 3 nm per minute at the RF power density of 1.234 W/cm^2 in the
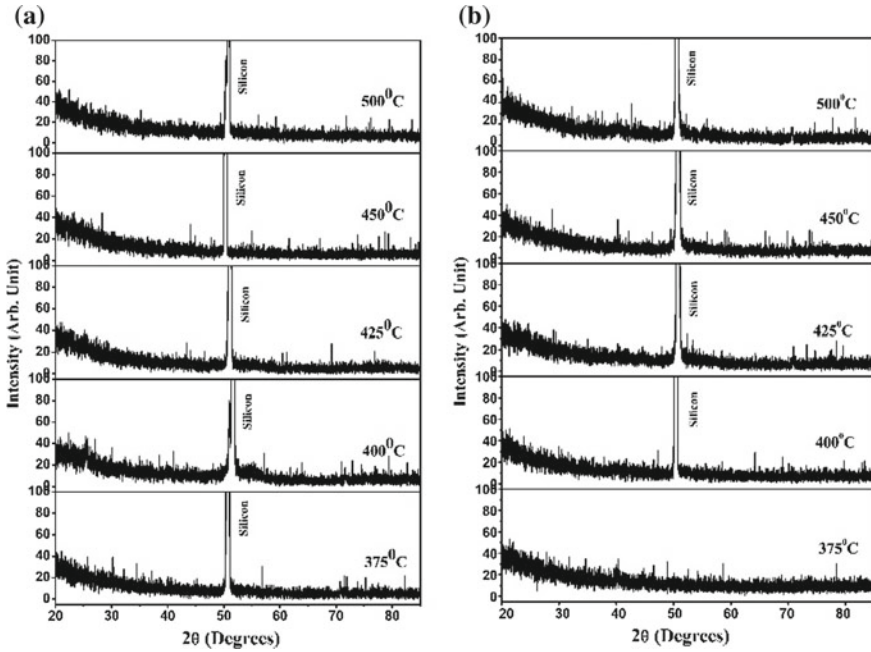
Fig. 1 XRD pattern of TiO$_2$ film annealed at 375, 400, 425, 450, and 500 °C, **a** PEALD deposited, and **b** sputtering deposited

Temperature (°C)	PEALD	Sputtering
375	2.0635	2.65
400	2.0595	2.72
425	2.0463	2.21
450	2.1348	2.75
500	2.0959	2.56

Table 2 Refractive Index of PEALD and sputtered-deposited TiO$_2$ and annealed at different temperatures

sputtering process. The fabricated MIS structures were tested from accumulation to inversion in different voltage range and the obtained capacitance–voltage (C-V) results are shown in Fig. 3.

The C-V curves clearly indicate the effect of post-deposited annealing temperature on the film capacitance and flatband voltages. For the PEALD-deposited films, the flatband voltage shifts toward positive with the annealing temperature up to 425 °C and shifts negative with further increase in the temperature. The maximum flatband voltage of 2.4 and 0.2 V was observed for PEALD and sputtering, respectively. Higher positive shift in the flatband voltage has been observed in the PEALD-deposited TiO$_2$ film; this is due to the higher negative charges at the TiO$_2$–Si interface (-1.37×10^{12})
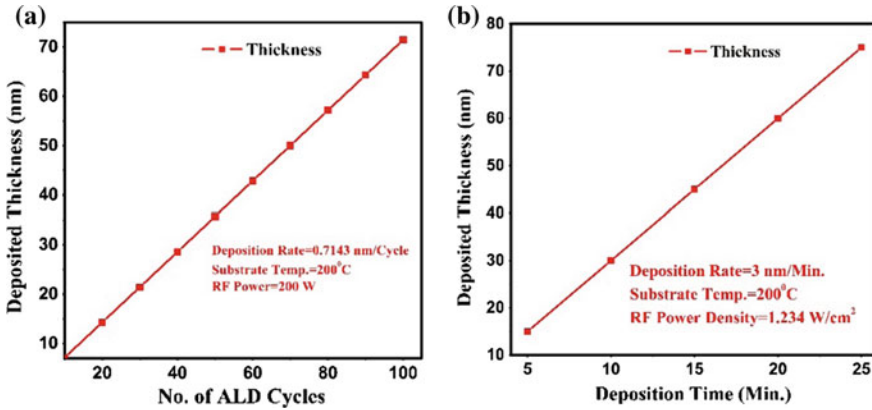
Fig. 2 TiO$_2$ deposition rate and deposition parameters for **a** PEALD, and **b** sputtering processes
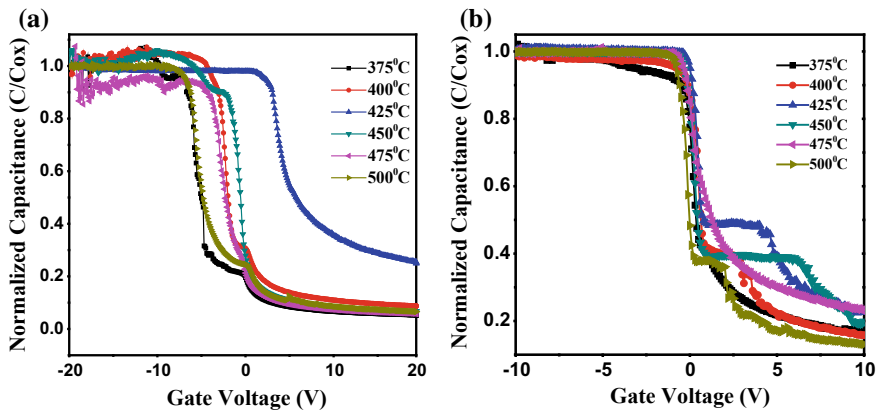
Fig. 3 C-V characteristic of the MIS capacitor with TiO$_2$ film annealed at different temperature and deposited by, **a** PEALD, and **b** sputtering

in PEALD as compared to the negative charges (-182.98×10^9) in the sputtered film. The film quality and the interface charges have a direct impact on the device leakage current density as shown in Fig. 4.

Minimum leakage current density was observed at the annealing temperature of 425 °C for the PEALD-deposited and sputter-deposited TiO$_2$ film. The leakage current density was found one order lower in the PEALD-deposited TiO$_2$ as compared to the sputtered film. Low leakage current density is due to the high-quality and defect-free amorphous film deposited by PEALD, which is in agreement with the C-V characteristics observed. Hopping current mechanism dominates at the low electric field for the films deposited by both techniques. Annealing temperature dependence on the device leakage current was seen in both the cases, where the leakage current density increases exponentially for the gate voltage up to 3 volts indicating thermionic
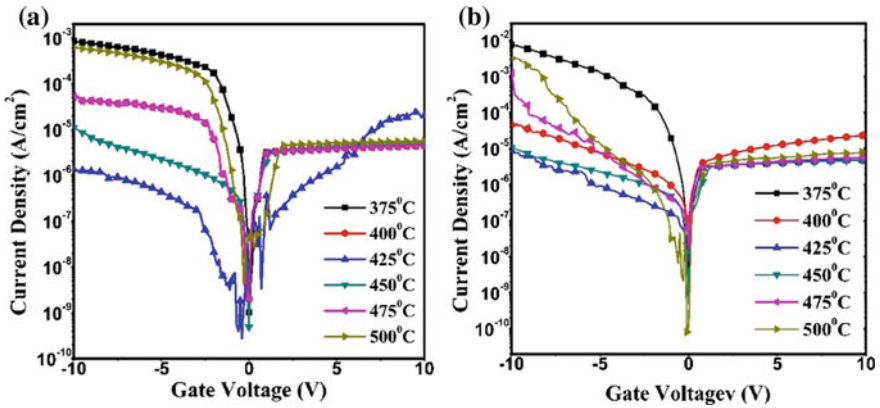
Fig. 4 J-V characteristic of the MIS capacitor with TiO$_2$ film annealed at different temperatures, **a** PEALD, and **b** sputtering

emission in the fabricated structures. High leakage current density in the reverse bias as observed from the J-V characteristic is due to the substrate injection whereas low leakage in the positive bias is attributed to the Schottky contact at the metal–insulator junction. Current density increases with applied gate voltage/electric field in forward and reverse bias. The current almost saturates for the voltage of magnitude above 5 V, indicating the direct tunneling which become constant after a certain high applied voltage/electric field. For the fabricated MIS structures annealed at 425 °C, stress voltage was applied and the stress current was measured to obtain its breakdown voltage, as shown in Fig. 5.

Fig. 5 Breakdown voltage of PEALD and sputtered TiO$_2$ film
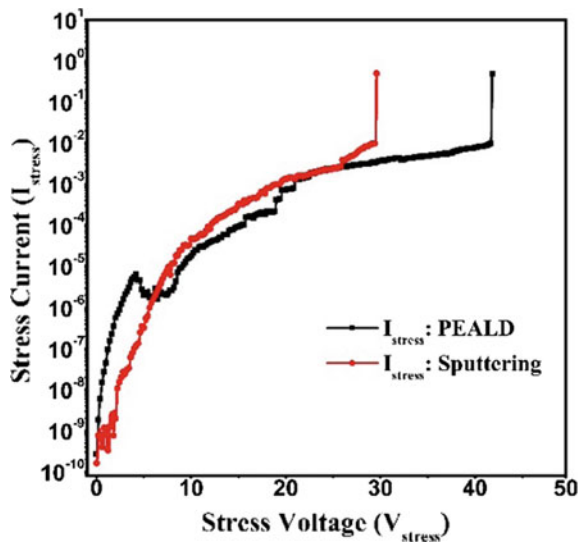
Presence of interfacial defects and traps lead to the breakdown of sputtered TiO_2 film at 29.7 V whereas the PEALD TiO_2 film shows the breakdown voltage of 41.8 V.

4 Conclusion

We have deposited TiO_2 film as a gate insulator by PEALD and RF sputtering. Amorphous nature of the deposited film was confirmed from the XRD results. Electrical characteristics show the positive flatband voltage shift and minimum leakage current density of the order 10^{-6} at the annealing temperature 425 °C. Obtained results show that the deposited film can be implemented effectively as the high-k dielectric (due to the amorphous nature), passivation layer (due to large negative charge, deposited by PEALD), and anti-reflecting coating (due to high refractive index).

References

1. Richards, B.S.: Comparison of TiO_2 and other dielectric coatings for buried-contact solar cells: a review. Prog. Photovoltaics Res. Appl. **12**, 253 (2004)
2. Kim, S.K., Kim, K.M., Jeong, D.S., Jeon, W., Yoon, K.J., Hwang, C.S.: Titanium dioxide thin films for next-generation memory devices. J. Mater. Res. **28**, 313 (2013)
3. Paily, R., DasGupta, A., DasGupta, N., Bhattacharya, P., Misra, P., Ganguli, T., Kukreja, L.M., Balamurugan, A.K., Rajagopalan, S., Tyagi, A.K.: Pulsed laser deposition of TiO_2 for MOS gate dielectric. Appl. Surf. Sci. **187**, 297 (2002)
4. Rausch, N., Burte, E.P.: Thin high-dielectric TiO_2 films prepared by low pressure MOCVD. Microelectron. Eng. **19**, 725 (1992)
5. Okimura, K., Maeda, N., Shibata, A.: Characteristics of rutile TiO_2 films prepared by rf magnetron sputtering at a low temperature. Thin Solid Films **281–282**, 427 (1996)
6. Hassan, A.K., Chaure, N.B., Ray, A.K., Nabok, A.V., Habesch, S.: Structural and electrical studies on sol–gel derived spun TiO_2 thin films. J. Phys. D Appl. Phys. **36**, 1120 (2003)
7. Yeo, Y.C., King, T.J., Hu, C.: Direct tunneling leakage current and scalability of alternative gate dielectrics. Appl. Phys. Lett. **81**, 2091 (2002)
8. Campbell, S.A., Gilmer, D.C., Wang, X.C., Hsieh, M.T., Kim, H.S., Gladfelter, W.L., Yan, J., Trans, I.E.E.E.: MOSFET transistors fabricated with high permitivity TiO/sub2/dielectrics. Electron Devices **44**, 104 (1997)
9. Xie, D., Han, X., Li, R., Ren, T., Liu, L., Zhao, Y.: Characteristics of Pt/BiFeO$_3$/TiO$_2$/Si capacitors with TiO_2 layer formed by liquid-delivery metal organic chemical vapor deposition. Appl. Phys. Lett. **97**, 172901 (2010)
10. Campbell, S.A., Kim, H.-S., Gilmer, D.C., He, B., Ma, T., Gladfelter, W.L.: Titanium dioxide (TiO2)-based gate insulators. IBM J. Res. Dev. **43**, 383 (1999)
11. Singh, P., et al.: On the structural and electrical properties of metal–ferroelectric–high k dielectric–silicon structure for non-volatile memory applications. Bull. Mater. Sci. **41**(4), 101 (2018)
12. Lu, X.: High-k Dielectrics in Ferroelectric Gate Field Effect Transistors for Nonvolatile Memory Applications. High-kGate Dielectr.C. Technol. (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany), pp. 471–499 (2012)
13. Mudavakkat, V.H., Atuchin, V.V., Kruchinin, V.N., Kayani, A., Ramana, C.V.: Structure, morphology and optical properties of nanocrystalline yttrium oxide (Y_2O_3) thin films. Opt. Mater. (Amst). **34**, 893 (2012)

Impact of HfO$_2$ as Passivation Layer in the Simulation of PERC-Type Solar Cell

Rajesh Kumar Jha [ID]**, Prashant Singh** [ID]**, Manish Goswami and B. R. Singh**

Abstract We report the simulation of high-efficiency c-silicon Passivated Emitter Rear Contact (PERC) type solar cell structure with rear side passivated with HfO$_2$ as a passivating material. Variation in the half-length of pyramid has been carried out to investigate its effect on the solar cell electrical characteristics such as Fill Factor (FF), Open-Circuit Voltage (Voc), and efficiency. Aluminum Back Surface Field (Al-BSF) and PERC-type solar cell with Al$_2$O$_3$ passivation layer structures were also modeled for comparison. Effect of variation in passivation layer (HfO$_2$), thickness (10 and 15 nm), and permittivity (k = 14 and 25) on the solar cell electrical characteristics has been investigated. The result shows the efficiency improvement in the PERC solar cell with HfO$_2$ passivation layer by 0.5941 and 0.983% as compared to the Al-BSF and PERC with Al$_2$O$_3$ passivation layer at 8 μm pyramid half-length. Increased series resistance and reduced FF has been observed with the incorporation of passivation layer at the solar cell structure. Negligible effect of passivation layer thickness has been observed on the solar cell electrical parameters whereas the permittivity value does have significant effect.

1 Introduction

The demand for energy in industrial, household, and agricultural applications are increasing day by day and projected to be doubled by 2040 [1]. The nonrenewable energy resources (fossil fuels) are being harvested at a very high rate to meet these energy requirements and contribute a total 80% of the world's energy demand [2]. Continuously depleting nonrenewable energy resources and increase in the greenhouse gasses resulting in global warming causes a great threat to the human existence [3]. To mitigate the cause of global warming and continuously increasing air, water, and soil pollution, development of renewable energy sources has been promoted

R. K. Jha (✉) · P. Singh · M. Goswami · B. R. Singh
Department of Electronics and Communication Engineering, Indian Institute of Information Technology Allahabad, Allahabad 211015, Uttar Pradesh, India
e-mail: rajajha25@gmail.com

worldwide [4]. In the present scenario, India in spite of being the third largest electricity producer country in the world is not able to provide electricity to around 237 million people and around 81 million households due to the geographical challenges [5]. By setting up the solar power generation unit in the rural areas will also generate the economic opportunities and improvement in education, health, and social welfare of that region [6]. As per the Ministry of New and Renewable Energy (MNRE) annual report 2015–16, India has the total solar power potential of 750 GW and the government is targeted to achieve the solar power capacity to 100 GW by the year 2022 [7].

Among different available renewable energy resources (Wind, Hydro, Tidal, Solar, etc.), solar energy shows the maximum potential to fulfill the world energy requirement. About 60% of the incoming solar energy reaches the Earth's surface, which can be effectively transformed into electrical energy using high-efficiency photovoltaic cells. Due to the high availability of solar energy, only 0.1% conversion into electrical energy at the efficiency of 10% would be four times larger than the total world's electricity generating capacity, i.e., about 5000 GW [8].

Solar cells or photovoltaic cells are required for the conversion of solar radiations into electrical energy. The solar cell, basically, is a *p–n* junction device that converts the solar energy into electric current by generating electron-hole pairs [9]. Some semiconductor materials exhibit photoelectric effect that causes them to absorb the photons of light and release electron-hole pairs resulting in nonzero net current [10]. Solar energy is a clean and renewable energy source, which will never drain out. Solar energy production costs are huge and storage is a problem as well. Photovoltaics is the fastest growing market with compound annual growth rate of PV installation of 40% [11]. Different types of solar cell technologies have been developed in order to achieve maximum efficiencies in the best possible cost-effective ways. Silicon wafer (monocrystalline and multi-crystalline) based solar cell is accounted for a total share of over 90% in terms of cumulative installed capacity [12]. Development of Silicon solar cell technology experiences high fabrication cost. The three important interrelated factors affecting the price of solar cell are mass production, improved cell efficiency, and reduced material cost [13]. To reduce the cost of the silicon-based solar cells, thickness of the silicon wafer has to be reduced. The reduced thickness of the silicon wafer results in the increase of the recombination losses in the solar cells due to the increase of the surface/volume ratio [14, 15]. Hence, to maintain the high efficiency of solar cell for reduced thickness, these recombination losses should be minimized. More than 50% of the recombination losses occur at the surface of the silicon wafer. Therefore, surface passivation, a technique to minimize these surface recombination losses has been carried out in this work with HfO_2 as passivation layer. HfO2 has been selected as the passivation layer due to its high dielectric constant, high energy bandgap, and thermodynamic stability with the silicon surfaces. In addition, HfO_2 is a novel candidate for c-silicon solar cell surface passivation [16, 17].

2   Solar Cell Simulation

Silvaco TCAD software has been used in this study for the designing and simulating different solar cell structures. This software package provides tools for Electronic Design Automation (EDA) and TCAD process and device simulation. Different simulators have been used like Athena, Atlas, DeckBuild, and DevEdit. Athena process simulator has been used to create the solar cell structure and to determine the fundamental deposition and etch processes. Atlas device simulator has been used to simulate the devices physically and to calculate the electrical characteristics with input from Athena and DevEdit. DevEdit tool has been used to define the device structure and generate new mesh. Tonyplot is a visualization tool and used to visualize the TCAD structures in 1D, 2D, and 3D.

P-type <100> silicon wafer of thickness 275 μm and resistivity 1 Ω-cm has been taken as the substrate material. The front surface of silicon wafer has been etched using geometrical silicon etching. Phosphorus doping has been used for n+ emitter region formation with a sheet resistance of 90 Ω-cm. Si$_3$N$_4$ film of thickness 70 nm has been used as antireflection layer on the top of the designed cells. The top aluminum metal of thickness 1 μm has been used as the contact electrode. Textured silicon surface has been achieved by etching and phosphorus diffusion has been done with the dopant dose of 5×10^{17} for 30 min at 900 °C. HfO$_2$ passivation layer of thickness 15 μm has been deposited on the rear side of the solar cell. The passivation layer has been etched out to achieve the metallization fraction of 10% and the full coating of aluminum has been done for the rear contact. For the electrical characterization of modeled structure, Newton numerical method with maximum iteration limit set to 50 has been used to calculate the current, Voc, and fill factor. Electrical parameters comparison has been done for the solar cell with passivation layer and Al-BSF (structure without passivation layer). Various parameters such as material, contact and interface parameters have been specified for the solar cells. Mobility, Shockley–Read–Hall (SRH), and Auger recombination models have been used to simulate the mobility and recombination of charge carriers in the silicon solar cells. Luminous optoelectronic simulator, which calculates the optical intensity profiles of the semiconductor and converts it into the photo-generation rate has been used for solar cell structure simulation under this work.

3   Results and Discussion

3.1   Al-BSF and PERC Structure

Figure 1 shows the top view of the Al-BSF and PERC solar cell structures modeled in Silvaco TCAD.

On the top of the designed Al-BSF and PERC structures, no passivation layer has been deposited in order to keep the design simple for comparison purpose. The full
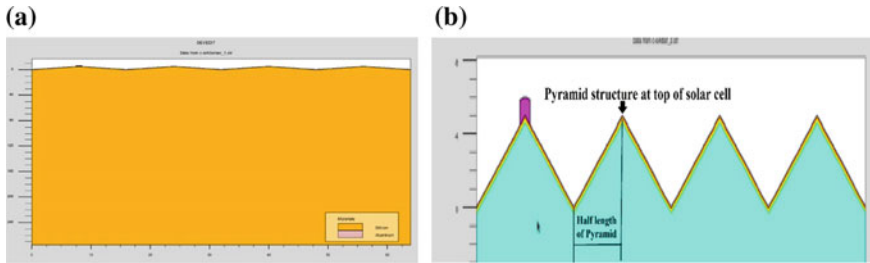
Fig. 1 Al-BSF and PERC solar cell structure, **a** top view, and **b** top pyramid
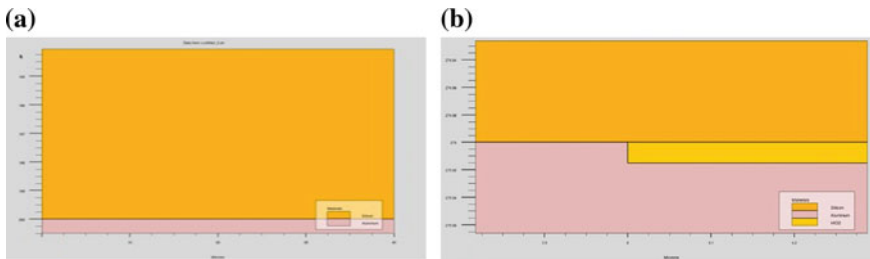
Fig. 2 Bottom view of modeled **a** Al-BSF, and **b** PERC solar cells

rear metal coating has been done for AL-BSF structure whereas HfO_2 passivation layer has been introduced at the rear side of the PERC solar cell between the silicon and metal contact, shown in Fig. 2.

Al-BSF solar cell structure has been simulated to obtain the Voc, FF, and efficiency. The obtained AL-BSF parameters have been used as the reference to compare the improvement in PERC structure characteristics. The PERC structure has been simulated with variation in permittivity (k = 14 and 25) and thickness (10 and 15 nm) of the HfO_2 passivation layer. The effect of difference in thickness and permittivity value on the AL-BSF and PERC solar cell electrical characteristics and efficiency has been compared.

3.2   Al-BSF and PERC Structure with HfO_2 Passivation Layer

3.2.1   Solar Cell Electrical Characteristics for k = 14

Different solar cell structures (Al-BSF and PERC) has been simulated with variation in the different lengths of half pyramid. For the constant permittivity value (k = 14) of the HfO_2 passivation layer, different solar cell structures have been simulated and electrical characteristics are plotted in Fig. 3.
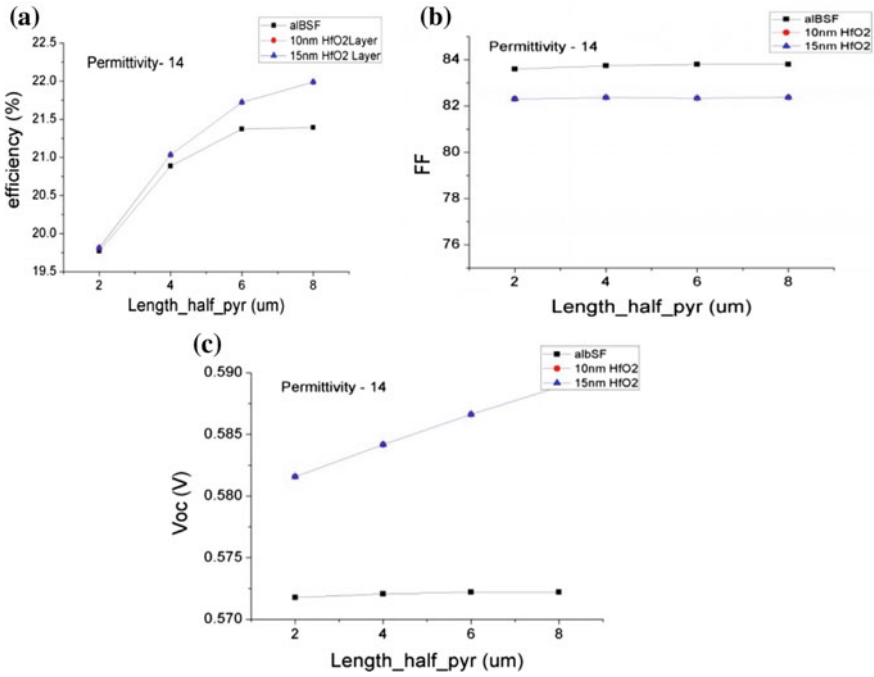
Fig. 3 Al-BSF and PERC solar cell characteristics, **a** efficiency versus length of half pyramid, **b** FF versus length of half pyramid, and **c** Voc versus length of half pyramid

The variation in the length of half pyramid has been kept constant for the range 2–8 μm for the AL-BSF and PERC (with 10 and 15 nm HfO₂ layer thickness) solar cells structures while investigating the efficiency, FF, and Voc. For the Al-BSF structure, it was found that the efficiency increases from 19.75 to 21.3919% with the increase in the length of half pyramid. Similarly, the improvement in the efficiency of PERC structure with 10 and 15 nm passivation layer thickness has been observed. The efficiency improved from 19.80 to 21.986% for the increase in the length of half pyramid for 10 and 15 nm passivation layer thickness. It has been observed that the passivation layer thickness does not influence the PERC solar cell efficiency for HfO₂ permittivity value of 14. Comparatively, the efficiency of the PERC solar cell structure improves from AL-BSF structure by 0.5941%. Efficiency improvement observed in the PERC structure is due to the reduced recombination rate of carriers and due to HfO₂ passivation layer as comparison to Al-BSF structure. Carrier recombination reduced in PERC solar cell due to the shielding of carriers to reach the metal contacts. Carrier shielding is provided because of the fixed negative interface charges of Si/HfO₂ interface [18].

The FF of Al-BSF and PERC structures has been found almost constant with varying length of half pyramid. A decrement of 1.4319 in the FF has been observed

in the PERC structures with 10 and 15 nm passivation layer as compared to the Al-BSF structure. The decrement on the FF is attributed to the increased series resistance in the PERC structure.

The Voc of modeled Al-BSF structure has been found 0.57221 and does not vary significantly with the increasing length of half pyramid. Voc of the PERC structure increases from 0.584 to 0.5885 V for the increase in the length of half pyramid from 2 to 8 μm. An improvement of 1.66% in the Voc for the PERC structure has been observed as compared to the Al-BSF structure.

3.2.2   Solar Cell Electrical Characteristics for k = 25

Figure 4 shows the efficiency, FF, and Voc characteristics of Al-BSF and PERC structures with respect to the length of half pyramid.

The efficiency of the PERC structure with the HfO_2 passivation layer of permittivity 25 and thickness 10 and 15 nm is found improved by 0.5951% as compared to the Al-BSF solar cell. Improved efficiency is due to the negative charges at the HfO_2/Si interface. The FF for Al-BSF and PERC structure is observed to be 83.7969 and 82.365, respectively. The FF remains almost constant for the varying length of half pyramid. Voc of Al-BSF is found 0.57221 and does not change significantly
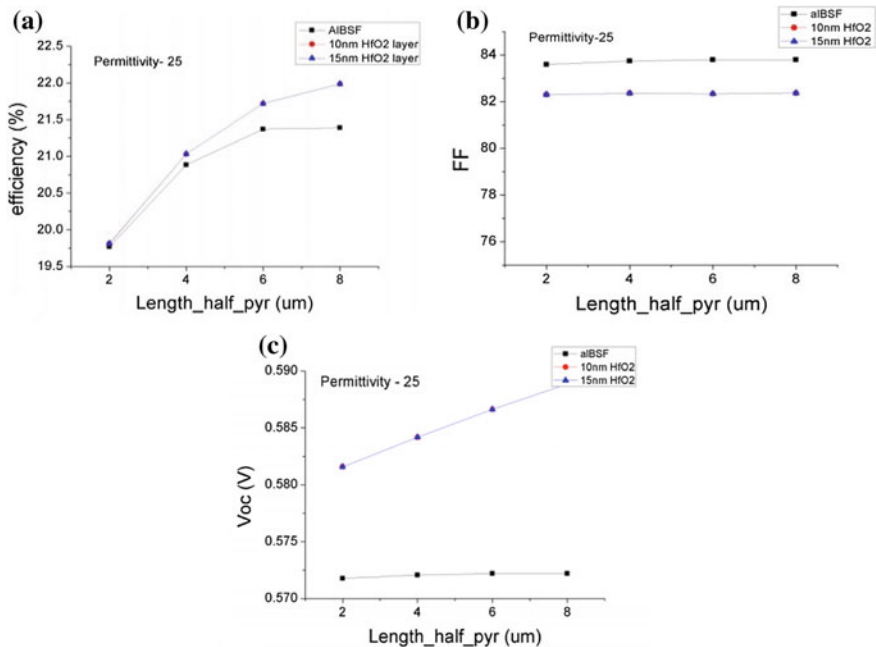
Fig. 4 Al-BSF and PERC solar cell, **a** efficiency versus length of half pyramid, **b** FF versus length of half pyramid, and **c** Voc versus length of half pyramid

with the varying half-length of pyramid. For the PERC structure, the Voc increases from 0.582 to 0.58885 with the length of half pyramid from 2 to 8 nm. The overall improvement of 1.664% has been observed in Voc for the PERC structure as compared to the Al-BSF structure.

3.3   PERC Structure with Al₂O₃ and HfO₂ Passivation Layer

PERC solar cell structure with Al_2O_3 and HfO_2 buffer layer has been simulated to obtain the Voc, FF, and efficiency. The analysis has been made for identical thickness (10 and 15 nm) of Al_2O_3 and HfO_2 passivation layers and summarized in Tables 1 and 2.

Results indicate that the efficiency of PERC structure with HfO_2 passivation layer is 0.983% higher than the cell with Al_2O_3 layer. Increase in the solar cell efficiency for HfO_2 passivation layer is the result of effective surface passivation as compared to the Al_2O_3 passivation layer. From the comparison table, it is clear that the passivation layer thickness does not influence the solar cell efficiency but the Voc. For 10 nm passivation layer thickness, the Voc for PERC (HfO_2) has been increased by 0.8% and FF is decreased by 0.161 as compared to the PERC (Al_2O_3). For 15 nm passivation layer thickness, the Voc and FF for PERC (HfO_2) have been decreased by 2.2% and 0.162, respectively, as compared to the PERC (Al_2O_3).

Table 1 PERC solar cell characteristics for HfO_2 and Al_2O_3 passivation layer of 10 nm

Solar cell type	Voc (V)	FF	Efficiency (%)
PERC_HfO₂ (10 nm)	0.588	82.365	21.987
PERC_Alumina (10 nm)	0.580	82.204	21.004
Overall improvement	0.008 (0.8% increment)	0.161 (decrement)	0.983

Table 2 PERC solar cell characteristics for HfO_2 and Al_2O_3 passivation layer of 15 nm

Solar cell type	Voc (V)	FF	Efficiency (%)
PERC_HfO₂ (15 nm)	0.588	82.365	21.986
PERC_Alumina (15 nm)	0.610	82.203	21.004
Overall improvement	0.022 (2.2% decrement)	0.162 (decrement)	0.983

4 Conclusion

Different electrical parameters (Voc, FF, and efficiency) of the PERC solar cell with HfO_2 passivation layer have been studied. Passivation layer thickness, permittivity, and length of half pyramid have been varied to observe its effect on the different solar cell electrical parameters. The obtained results were compared with the Al-BSF and Al_2O_3 passivated PERC structure. Significant efficiency improvement has been observed in the HfO_2 passivated PERC structure than the Al-BSF and Al_2O_3 passivated PERC structures. The optimized value of half pyramid length (8 μm) has been obtained for which best efficiency can be achieved in case of passivation and without passivation.

References

1. U.S. Energy Information Administration: International Energy Outlook 2016, vol. 0484 (2016)
2. World Bank: Fossil Fuel Energy Consumption (2014)
3. Warneck, P., Hansen, J.: Peter warneck. J. Atmos. Chem. **30**(3), 409–412 (1998)
4. Council of the European Union: European Council Brussels, 22 and 23 March 2005. Presidency Conclusions, vol. 2005 (2005)
5. Manju, S., Sagar, N.: Progressing towards the development of sustainable energy: a critical review on the current status, applications, developmental barriers and prospects of solar photovoltaic systems in India. Renew. Sustain. Energy Rev. **70**, 298–313, 01 Apr 2017 (Pergamon) (2017)
6. Jain, S., Jain, N.K., Vaughn, W.J.: Challenges in meeting all of India's electricity from solar: an energetic approach. Renew. Sustain. Energy Rev. **82**, 1006–1013 (2018)
7. Financial Relations: Annual report. Available http://mnre.gov.in/file-manager/annual-report/2015-2016/EN/Chapter1/chapter_1.htm. Accessed 17 Jan 2018 (Online)
8. World Energy Council: World Energy Resources: 2013 Survey (2013)
9. Goetzberger, A., Knobloch, J., Voß, B.: The p-n Junction. Cryst. Silicon Sol. Cells, pp. 49–65 (2014)
10. Goetzberger, A., Knobloch, J., Voß, B.: The principles of photovoltaics. In: Crystalline Silicon Solar Cells. Wiley, Chichester, UK, pp. 9–48 (2014)
11. Philipps, S., Ise, F., Warmuth, W.: Fraunhofer ISE: Photovoltaics Report (2017)
12. Solar Cell Market Size, Share, Industry Report, 2022: Available https://www.grandviewresearch.com/industry-analysis/solar-cell-market. Accessed 18 Jan 2018 (2016) (Online)
13. Solanki, C.: Solar Photovoltaics Fundamentals, Technologies and Applications (2009)
14. Hoex, B., Schmidt, J., Bock, R., Altermatt, P.P., van de Sanden, M.C.M., Kessels, W.M.M.: Excellent passivation of highly doped p-type Si surfaces by the negative-charge-dielectric Al2O3. Appl. Phys. Lett. **91**(11), 112107 (2007)
15. Hoex, B., Schmidt, J., Pohl, P., van de Sanden, M.C.M., Kessels, W.M.M.: Silicon surface passivation by atomic layer deposited Al_2O_3. J. Appl. Phys. **104**(4), 44903 (2008)
16. Wang, J., Farrokh Baroughi, M., Shanmugam, M., Samadzadeh-Tarighat, R., Sivoththaman, S., Paul, S.: Passivation of silicon surfaces using atomic layer deposited metal oxides. MRS Proc. **1153**, pp. 1153-A07-17 (2011)

17. Wang, J.W.J., Baroughi, M.F., Bills, B., Galipeau, D., Samadzadeh, R., Sivoththaman, S.: Substrate dependence of surface passivation using atomic layer deposited dielectrics. In: Proceedings of the 34th IEEE Photovoltaic Specialists Conference (PVSC), 2009, pp. 1988–1991 (2009)
18. Lin, F., et al.: Low-temperature surface passivation of moderately doped crystalline silicon by atomic-layer-deposited hafnium oxide films service low-temperature surface passivation of moderately doped crystalline silicon by atomic-layer-deposited hafnium oxide films. ECS J. Solid State Sci. Technol. **2**(1), 1–5 (2013)

Effect of Micro Lever Width on the Mechanical Sensitivity of a MEMS Capacitive Accelerometer

Apoorva Dwivedi, Prateek Asthana and Gargi Khanna

Abstract The paper presents a MEMS capacitive accelerometer using microlevers as a suspension system. The microlevers serve to amplify the output displacement of the accelerometer and hence increase the mechanical sensitivity of the device. A novel accelerometer design is presented in the paper to be used as a transducer for a totally implantable hearing application. The design considerations for surgical implantation of the accelerometer on the middle ear bone umbo are followed in selecting the dimensions of the accelerometer. The impact of the width of a micro lever on the displacement of the accelerometer is considered, and an analytical model including a correction factor is developed. The proposed accelerometer is designed and simulated in COMSOL MULTIPHYSICS 4.2. The developed analytical model is validated by comparing with the COMSOL simulation results. The mechanical sensitivity of the accelerometer with and without considering the effect of microlever width is compared. The enhanced mechanical sensitivity of 2.60 nm/g is obtained.

Keywords MEMS · Capacitive accelerometer · Microlevers · Mechanical amplification

1 Introduction

A MEMS Capacitive accelerometer is a device which is used to detect mechanical vibrations and convert them into capacitance changes which can be converted into readable electrical forms. The wide usability of such devices in a host of applications like biomedical sensing, tilt measurement, inertial navigation, earthquake detection, surveillance applications, microgravity and space applications, oil explorations, etc., necessitates high performance and hence enhanced sensitivity.

Some of the reported methods to increase the sensitivity of MEMS capacitive accelerometers include increasing the proof mass and nominal capacitance, low

A. Dwivedi (✉) · P. Asthana · G. Khanna
National Institute of Technology Hamirpur, Hamirpur 177005, HP, India
e-mail: apoorva.dwivedi07@gmail.com

damping and decreasing the stiffness constant along the sense axis [1–4]. Geometrical optimisation of the various geometry parameters of the device and the use of genetic algorithms have also been reported in the literature [5, 6]. Also, the use of advanced electronics interface circuitry is prevalent in research circles to accomplish the goal of enhanced sensitivity [7–9].

One method to enhance the sensitivity in capacitive accelerometers is by amplifying the displacement of the sensing element. This is achieved by using microlevers. The microlever connects the proof mass with the capacitive sense mechanism. The displacement output at the capacitive sensing stage is amplified by the property of the lever. The levers can be classified into three types based upon the position of the pivot, input effort and the output load as shown in Fig. 1. The type 1 levers amplify both the force and displacement. However, the effort and load move in opposite directions. In type 2 lever, both the load and effort move in the same direction. However, it can only amplify force. The type 3 lever is the ideal solution for increasing the total displacement and hence sensitivity as the effort and the load move in the same direction, and it can only amplify the displacement. The accelerometer proposed here utilised the type 3 lever as the effort and load moving in the same direction contribute to the increase in the overall displacement of the sensing element and hence increased sensitivity. The use of microlevers in microsystems has been reported in the literature [1, 4, 10]. However, the microsystems have been assumed to be wide and thus the effect of their width.

The paper studies the effect of variation in microlever width on the displacement sensitivity of a MEMS capacitive accelerometer which could be used in any application. The article is divided into four sections. The first section introduces the paper. The second section explains the design and working principle of the accelerometer

Fig. 1 The three types of micro levers
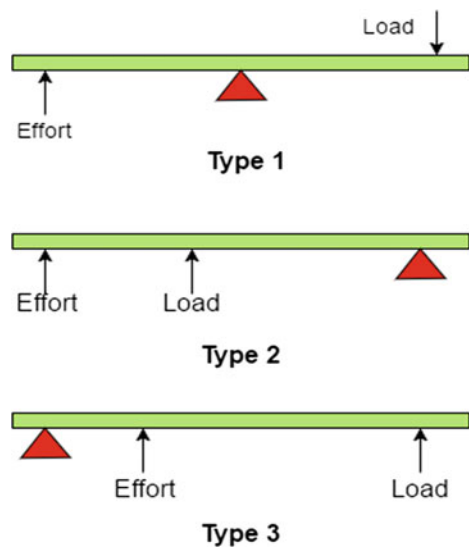
with the microlever. The results and discussions from the research work carried out are explained in the third section. The final section concludes the paper.

2 Dynamics of the Device

2.1 Device Structure and Working

The accelerometer consists of a proof mass connected to the capacitive sense mechanism on both sides by a system of four microlevers as shown in Fig. 2. The microlevers serve as the suspension element as well along with amplifying the displacement of the proof mass. The accelerometer is compactly designed such that it fulfils the requirements to be used as a fully implantable microphone surgically mounted on middle ear bone umbo [11]. The total number of sensing fingers is 276 on both the sides of the proof mass. The proof mass undergoes displacement under external acceleration which is then transferred to the capacitive sense mechanism. The change in capacitance can then be measured using an electronics interface circuitry.
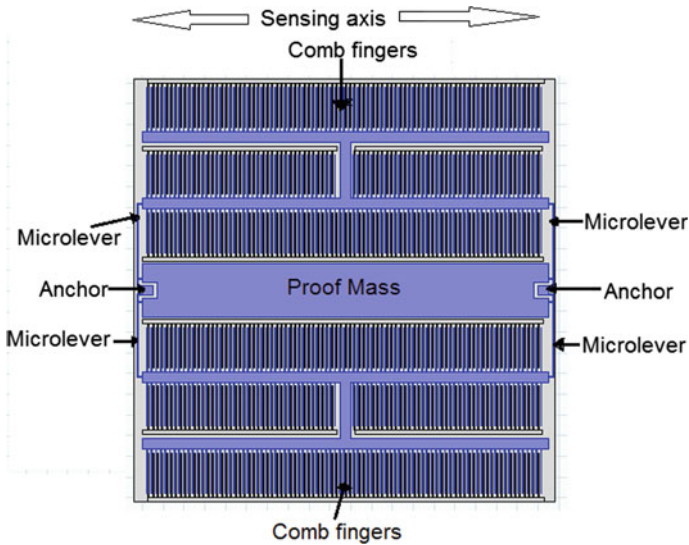
Fig. 2 The proposed novel accelerometer design
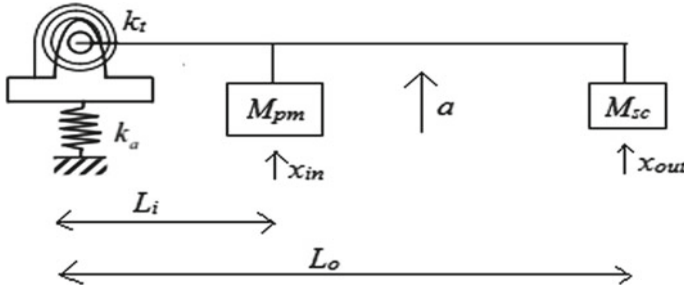
Fig. 3 The accelerometer lumped model with a microlever

2.2 Analytical Modelling

The microlever is the crucial feature of the proposed accelerometer design. Figure 3 shows the lumped model of the accelerometer with proof mass and capacitive sense mechanism. The pivot of the microlever is modelled as a hinged element representing the translation stiffness constituent and the rotational stiffness constituent k_{ai} and k_{ai}, respectively. The stiffness constituents are given as

$$k_{ai} = \frac{E_a}{L_p} \tag{1}$$

$$k_{ti} = \frac{EI}{L_p} \tag{2}$$

where E is Young's modulus of elasticity of material, $I = \frac{tw_b^3}{12}$ is the moment of inertia, L_p is the length of the pivot, $A = tw_b$ is the area of the cross section of the pivot, t represents the thickness of the lever, w_b denotes the width of the lever, L_0 is the length of the lever and L_i is the length of the effort arm. The stiffness constants at the output arm k_{ao} and k_{to} have the same values as k_{ai} and k_{ti}, respectively. Hence, the effective stiffness constants given by k_a and k_t are

$$k_a = k_{ai} + k_{ao} = 2k_{ai} \tag{3}$$

$$k_t = k_{ti} + k_{to} = 2k_{ti} \tag{4}$$

The external acceleration a causes the proof mass (m_{pm}) and the capacitive sense mechanism (m_{sc}) to displace under an applied force. The output displacement x_{out} of the capacitive sense stage [12] is evaluated as

$$x_{out} = \frac{(m_{pm} + m_{sc})a}{k_a} + \frac{L_i L_o (m_{pm} + \frac{L_o}{L_i} m_{sc})a}{k_t} \tag{5}$$
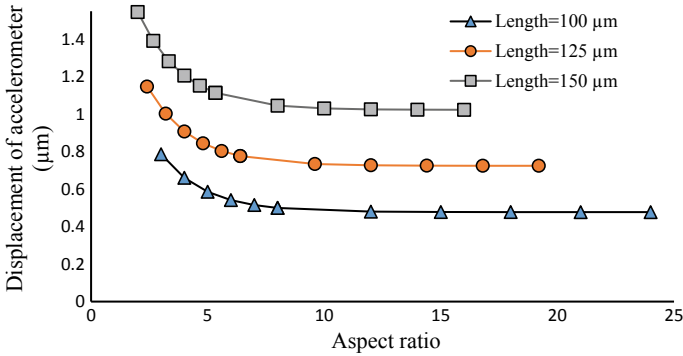
Fig. 4 Deviation in the displacement of the accelerometer with respect to the aspect ratio

In the above formula, the width of the lever is assumed to be wide and neglected. However, when the lever width is narrowed, it is found that the displacement sensitivity of the accelerometer is affected by the variation in the width of the lever. The simulations executed in COMSOL MULTIPHYSICS have offered substantial information on the effect of the width of microlever on the displacement of the accelerometer as shown in Fig. 4. The variation in accelerometer output displacement is evaluated with respect to the ratio of lever width to the lever length (aspect ratio). With a decrease in the microlever width, the displacement x_{out} is increased; thus, enhancing the sensitivity of the accelerometer as shown in Fig. 4. It is observed that an increase in the lever width causes the output displacement to stagnate and become independent of the lever width as already demonstrated by Eq. (5). The paper presents a correction factor β to include the displacement that is caused by the deformation of the lever arm with the aspect ratio (W_L/L_0).

$$\beta = L_0 \times 10^{-6} \times (2.31)^{100 \frac{W_L}{L_0}} \tag{6}$$

Including the correction factor β in Eq. (5), the output displacement becomes

$$x_{out} = \frac{(m_{pm} + m_{sc})a}{k_a} + \frac{L_i L_o (m_{pm} + \frac{L_o}{L_i} m_{sc})a}{k_t} - \left[L_0 \times 10^{-6} \times (2.31)^{100 \frac{W_L}{L_0}} \right] \tag{7}$$

The numbers 100 and 2.31 shown in Eq. (7) are acquired from an empirical analysis of the simulation findings. The mechanical/displacement sensitivity of the accelerometer is defined as x_{out}/a.

3 Results and Discussion

The proposed structure is designed in COMSOL MULTIPHYSICS 4.2. The material chosen for accelerometer design is silicon as it is the established technology in the fabrication of MEMS structures. The Young's modulus of elasticity of silicon is 131 Gpa and its density is 2330 kg/m^3. The dielectric material used is air as it offers very little damping. The three-dimensional geometry of the accelerometer structure is shown in Fig. 5.

The number of sensing capacitive fingers on both sides of the proof mass on the capacitive sense mechanism is 276. The thickness of the device layer is 25 μm.

Table 1 shows the geometrical parameters of the accelerometer. The range of the applied acceleration is from 0 to 1g. The geometry and input acceleration range for the accelerometer is selected such that it can be used in fully implantable hearing aids
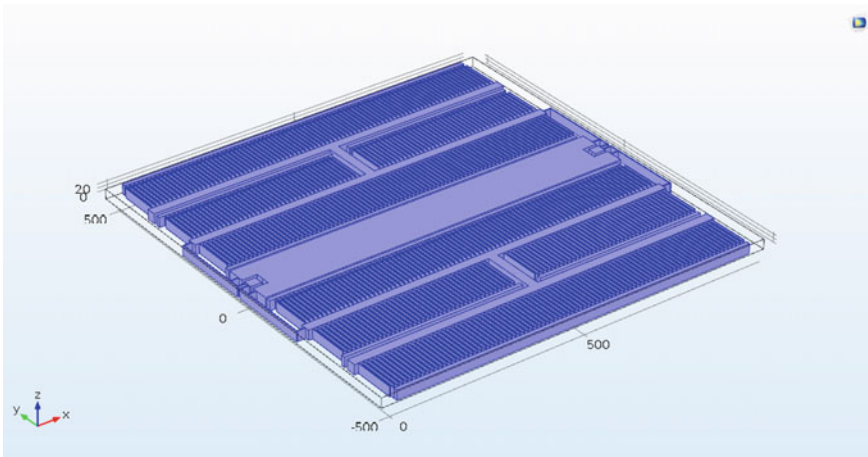
Fig. 5 Three-dimensional accelerometer model in COMSOL MULTIPHYSICS 4.2

Table 1 The dimensions of the accelerometer

Geometrical parameters	Dimension (μm)
Proof mass length	130
Proof mass width	954
Sensing finger length	96
Sensing finger width	2
Gap between sensing fingers	2
Length of spring beam (Ls)	207
Width of spring beam	3
Length of pivot arm	20
Width of pivot arm	3

mounted on middle ear bone umbo where high sensitivity is required to sense and measure incoming sound signals [6, 11]. Figure 6 indicates a direct linear relationship between the accelerometer output displacement with respect to the applied umbo acceleration. The analytical and simulation results are compared in Table 2.

The results in Table 2 show that the error in mechanical sensitivity is approximately 1%. Hence, the formula developed in the presented work accurately matches the simulation results.
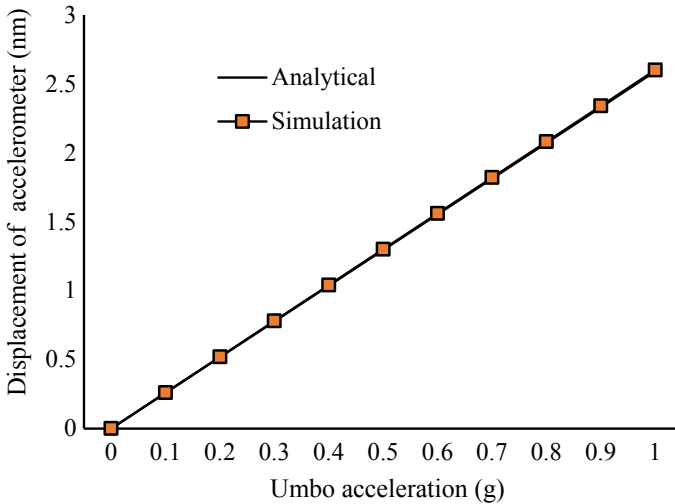
Fig. 6 Variation in the displacement of the accelerometer with respect to umbo acceleration

Table 2 Comparison of analytical and simulation values of mechanical sensitivity

Umbo acceleration (g)	Analytical values of mechanical sensitivity	Simulation values of mechanical sensitivity	Error (%)
0.0	0	0	0
0.1	0.259	0.260	1
0.2	0.518	0.520	1
0.3	0.777	0.780	1
0.4	1.036	1.040	1
0.5	1.295	1.300	1
0.6	1.554	1.560	1
0.7	1.813	1.820	1
0.8	2.072	2.080	1
0.9	2.331	2.340	1
1.0	2.590	2.600	1

4   Conclusion

The paper presents an accelerometer with microlevers for enhanced sensitivity. The paper studies the influence of the width of the microlever on the displacement of the accelerometer and introduces a correction factor for the same. It is observed that for enhanced mechanical sensitivity, the width of the lever needs to be as narrow as possible. The analytical results are compared with simulations implemented in COMSOL MULTIPHYSICS. The mechanical sensitivity of 2.60 nm/g is obtained.

References

1. Krishnan, G., Kshirsagar, C.U., Ananthasuresh, G.K., Bhat, N.: Micromachined high-resolution accelerometers. J. Indian Inst. Sci. **87**(3), 333–361 (2007)
2. Monajemi, P., Ayazi, F.: Design optimization and implementation of a microgravity capacitive HARPSS accelerometer. IEEE Sens. J. **6**(1), 39–46 (2006)
3. Chen, S., Chenyang, X., Wendong, Z., Jijun, X., Binzhen, Z., Jie, H.: A new type of MEMS two axis accelerometer based on silicon. In: Proceedings of 3rd IEEE International Conference NEMS, pp. 959–964, China (2008)
4. Abdolvand, R., Amini, B.V., Ayazi, F.: Sub-micro-gravity in-plane accelerometers with reduced capacitive gaps and extra seismic mass. J. Microelectromech. Syst. **16**(5), 1036–1043 (2007)
5. Dwivedi, A., Khanna, G.: Sensitivity enhancement of a folded beam MEMS capacitive accelerometer-based microphone for fully implantable hearing application. Biomed. Eng./Biomedizinische Technik (2017). https://doi.org/10.1515/bmt-2016-0181
6. Dwivedi, A., Khanna, G.: Optimization of device geometry of a fully implantable MEMS microphone for hearing aid by genetic algorithm. In: Proceedings of COMSOL Conference Bangalore, Bangalore (2016)
7. Chia-Pao, H., Yi-Chang, H., Ming-Chuen, Y., Weileun, F.: A novel SOI based single proof-mass 3-axis accelerometer with gap-closing differential capacitive electrodes in all sensing directions. In: Proceedings of IEEE Sensors, pp. 1188–1191, USA (2010)
8. Amini, B.V., Abdolvand, R., Ayazi, F.: A 4.5-mW closed-loop microgravity CMOS SOI accelerometer. IEEE J. Solid-State Circuits **41**(12), 2983–2991 (2006)
9. Qu, H., Fang, D., Xie, H.: A monolithic CMOS-MEMS 3-axis accelerometer with a low-noise, low-power dual-chopper amplifier. IEEE Sens. J. **8**(9), 1511–1518 (2008)
10. Yazdi, N., Najafi, K.: An all-silicon single-wafer micro-g accelerometer with a combined surface and bulk micromachining process. J. Microelectromech. Syst. **9**(4), 544–550 (2000)
11. Dwivedi, A., Khanna, G.: Numerical simulation and modelling of a novel MEMS capacitive accelerometer based microphone for fully implantable hearing aid. Microsystem Technologies (2018). https://doi.org/10.1007/s00542-018-4003-22018
12. Khan, S., Ananthasuresh, G.K.: Improving the sensitivity and bandwidth of in-plane capacitive micro-accelerometers using compliant mechanical amplifiers. J. Microelectromech. Syst. **23**(40), 871–887 (2014)

Noise and Linear Distortion Analysis of Analog/RF Performance in a Two Dimensional Dielectric Pocket Junctionless Double Gate (DP-JLDG) MOSFET to Control SCEs

Abhinav Gupta, Amrish Kumar, Sanjeev Rai and Rajeev Tripathi

Abstract In this paper, a dielectric pocket junctionless double gate (DP-JLDG) MOSFET has been proposed with an example of n-channel and with a 20 nm channel length. The characteristics of the proposed device have been compared with a junctionless double gate (JLDG) MOSFET. The proposed device offers excellent analog/RF behavior and can operate over a wide range of frequency with low power dissipation. This paper also analyses the impact of dielectric pockets on analog and digital performance of the device. Further, in this paper an exclusive effort to investigate the noise and distortion analysis of DP-JLDG MOSFET has been incorporated. The proposed device is an appropriate alternative for low power analog and digital circuits due to its higher device gain, larger operating range, and lower power dissipation.

Keywords JLDG · SCEs · Dielectric pocket · Analog and RF FOMs

1 Introduction

The continuous miniaturization of MOS device at nanometer scale imposes many challenges such as augmented gate leakage current and numerous short channel effects (SCEs) [1]. To extend the scaling limits of MOS devices, multiple gates (MG) MOSFETs have been considered as a foremost candidate. As the MG MOSFETs offers enhanced current driving competency and exceptional control over SCES [2]. The researchers have investigated and proposed various types of MG MOSFETs such as double gate [3], triple gate [4], quadruple gate [5], dual material double gate [6], triple material double gate [7], and junctionless double gate MOSFETs [8], etc.

A. Gupta (✉)
Electronics Engineering Department, Rajkiya Engineering College, Sonbhadra 231206, India
e-mail: abhinavkit87@gmail.com

A. Kumar · S. Rai · R. Tripathi
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India

To overwhelm the limitation of MOS device and to enhance the scaling limit of MOS devices, lateral channel engineering techniques such as halo implantation [9] and pocket implantation [10] has been incorporated by the researchers. The electric field penetration from drain to the source has been shielded by pocket implantation [11]. The presence of dielectric pocket (DP) reduces the charge sharing between source and drain regions. Because the DP near the sidewalls of the channel serves as diffusion stopper for highly doped source and drain regime. Hence the SCEs are reduced through DP. However, DP will enhance the band to band tunneling current and the risk of avalanche breakdown. It will also reduce the on current (I_{on}) of the device as the effective channel length of the device increases [12].

The Dielectric Pocket Junctionless Double Gate (DP-JLDG) MOSFET has been proposed as a leading candidate to come up with the aforementioned challenges by incorporating DP in a junctionless double gate (JLDG) MOSFET. Section 2 describes the JLDG device structure with all possible cases of DP-JLDG MOSFETs. The simulation setup has also been discussed in Sect. 2. In Sect. 3, the electrostatic, analog and RF analysis has been studied. Further, the linearity analysis has been discussed in Sect. 4. The advantages of DP-JLDG MOSFET over conventional JLDG MOSFET have been discussed in Sect. 5 in terms of electrostatic, analog/RF and linearity performance matrices.

2 DP-JLDG Device Structure and Simulation Setup

In 2006, Jayanarayanan et al. [13] and Gili et al. [14] have already proposed the dielectric pocket double gate MOSFET concept for a junction based DG architectures having 50 nm and 70 nm channel length respectively. In the present work, a similar device concept has been adopted for JLDG MOSFET to reduce the SCEs. The six possible structures have been obtained by changing the location of dielectric pocket (DP). Figure 1 shows the JLDG device structure (D1) having 20 nm channel length and all possible DP-JLDG device structures have been illustrated in Fig. 2. Device D2 is the DP-JLDG MOSFET incorporating DP at both source and drain side. Devices D3 is the DP-JLDG MOSFET considering DP at the boundary of channel and source/drain at both source and drain side. Device D4 is the DP-JLDG MOSFET with DP inside the channeling regime at both source and drain end. Device D5 is having DP only at the drain end. Device D6 includes DP at the edge of channel and drain only at drain side. Device D7 is the DP-JLDG MOSFET incorporating DP inside the channel region only at the drain side.

The silicon film thickness must be scaled down with channel length in order to minimize the SCEs [15]. Hence the channel thickness considered in this work is 5 nm. The doping concentration of channel, source, and drain regions have been considered as 1×10^{19} cm^{-3}. The work function of the metal gate is 5.2 eV. The thicknesses of the top and bottom gate oxides are 1 nm each. The DP length and DP thickness have been taken as 7 nm and 1.5 nm respectively. In numerical simulations, Shockley
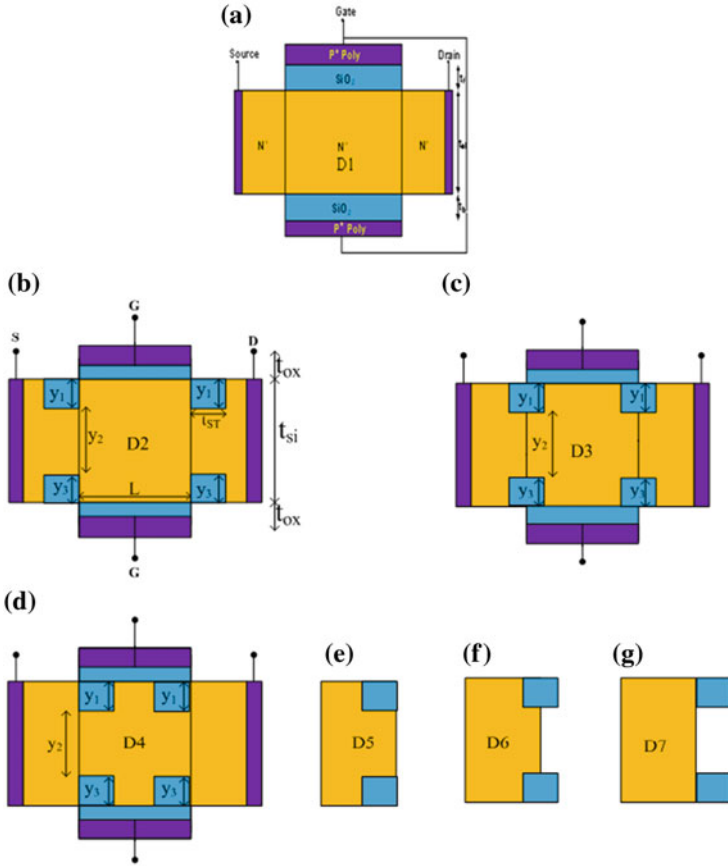
Fig. 1 a JLDG MOSFET, **b–g** all possible DP-JLDG MOSFET device structures

Read Hall (SRH) model has been included, to care about minority carrier recombination. The field mobility models (FLDMOB), constant mobility model (CONMOB), and Boltzmann model are used to imitate electrostatic, analog/RF and linearity performance matrices of the DP-JLDG MOSFET. The various scattering mechanisms cause the mobility degradation in the device. To consider these degradation effects, constant mobility (CVT) model has been incorporated for temperature variation.

3 Noise and Linear Distortion Analysis

If the terminal voltage of a MOS device varies with time, the on current of the device also varies with time. But this assumption is not true as the on current reveals minute fluctuations that are referred to as noise. Hence, an accurate noise description is
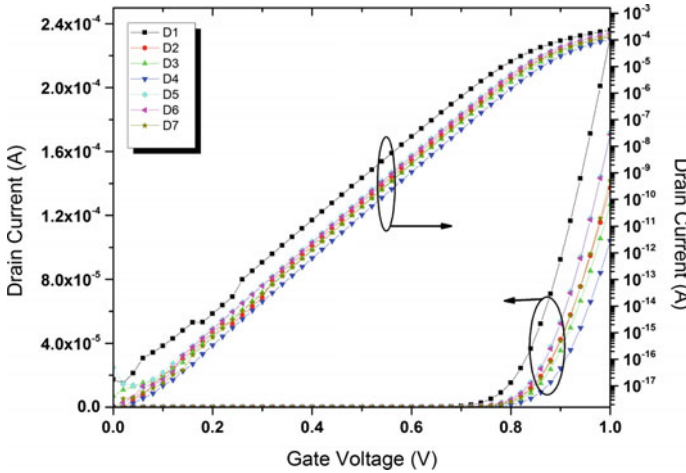
Fig. 2 Plot of drain current with gate voltage for all device structures (i.e. D1, D2, D3, D4, D5, D6, and D7) for both linear and logarithmic scale

important for optimum circuit design [16, 17]. The thermal noise is an important parameter of a device. It is produced by the arbitrary movement of charge carriers in a conducting medium. This arbitrary movement accounts for the capability of material to store energy in the form of random agitation. The expression for thermal noise can be given as;

$$S_{vt} = 4 \cdot K \cdot T \cdot f_T \cdot R_{DS} \tag{1}$$

where K is Boltzmann's constant, T is operating device temperature, f_T is the cutoff frequency (i.e. operating bandwidth), S_{vt} is the thermal noise power spectral density and R_{DS} is the drain to source resistance.

In analog/RF circuits, the nonlinear behaviors such as higher order harmonics and intermodulation distortions are introduced at the output which results in the wastage of useful output power. Hence the reduction of harmonic distortion turns out to be a primary objective for analog/RF circuit application based on MOSFETs. The linear distortion characteristics can be analyzed by higher order transconductance coefficients, higher order voltage, impedance interception points, and intermodulation distortion. For an extremely linear device, the higher order transconductance should be insignificant. Mathematically its coefficients can be expressed as;

$$g_m = \frac{\partial^n I_{DS}}{\partial V_{GS}^n} \tag{2}$$

The identical extrapolated input voltage for the first-second and first-third harmonics are represented by voltage intercept points (i.e. VIP2 and VIP3) and are given as

$$VIP2 = 4 \times \frac{g_{m_1}}{g_{m_2}} \tag{3}$$

$$VIP3 = \sqrt{24 \times \frac{g_{m_1}}{g_{m_3}}} \tag{4}$$

The input intercept point of the third harmonic (IIP3) can be defined as the extrapolated input power at which the third harmonic and first harmonic powers become comparable and is expressed as

$$IIP3 = \frac{2 \times g_{m_1}}{3 \times g_{m_3} \times R_s} \tag{5}$$

The intermodulation distortion of third harmonic can be expressed as in Eq. (6) and is defined as intermodulation harmonic power, where third and first-order intermodulation harmonic powers becomes comparable. For enhanced device characteristics all intercept points should be as high as possible whereas IMD3 should be low enough such that it will provide lower distortion.

$$IMD3 = \left\{ \frac{9 \times (VIP3)^3 \times g_{m_3}}{2} \right\}^2 \times R_s \tag{6}$$

4 Noise and Linear Distortion Analysis

In this section, the characteristics of the dielectric pocket junctionless double gate (DP-JLDG) MOSFET has been investigated and compared with a JLDG MOSFET by using ATLAS^{TM} device simulator [18]. The plot of drain current for all device structures (i.e. D1, D2, D3, D4, D5, D6, and D7) for both linear and logarithmic scale is shown in Fig. 2. From Fig. 2 it can be observed that the on current of a JLDG MOSFET degrades with the application of dielectric pocket (DP).

Figure 3 illustrates the plot of on current and output conductance of all device cases for V_{GS} fixed at 1 V and varying drain bias voltage. The ON current of JLDG MOSFET reduces due to the presence of DP at the boundary of the source/drain regime. Hence, the output conductance also reduces with the application of DP. The plot of transconductance generation factor (TGF) and transconductance with gate to source voltage for all cases of DP-JLDG MOSFET is illustrated in Fig. 4. From Fig. 4, it has been examined that with the application of DP, TGF increases, i.e., the maximum gain can be achieved with lower power dissipation with the device. Figure 5 demonstrates the plot of early voltage and intrinsic gain (dB) with gate to source bias voltage for all device cases. With the application of DP at the boundary, both early voltage, and intrinsic gain increases.
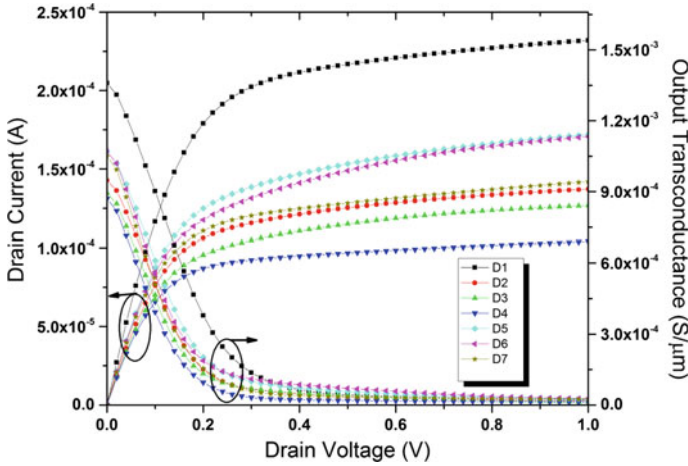
Fig. 3 Plot of drain current and output conductance with gate voltage for all device structures (i.e., D1, D2, D3, D4, D5, D6, and D7) for both linear and logarithmic scale
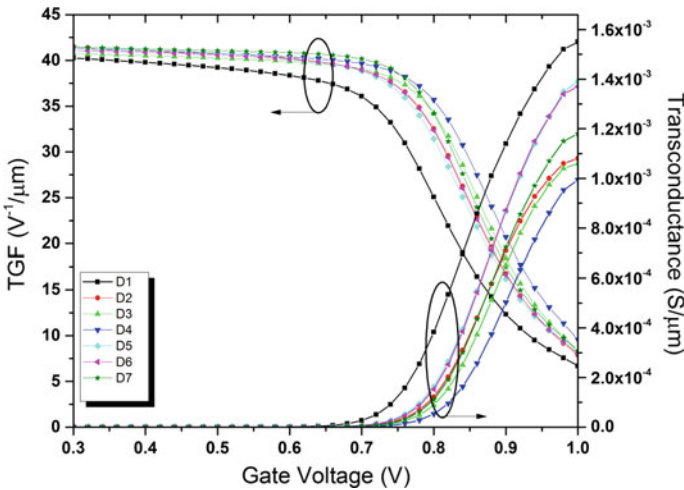
Fig. 4 The plot of transconductance generation factor (TGF) and transconductance with gate to source voltage for all cases of DP-JLDG MOSFET

The electrostatic performances of all cases of DP-JLDG MOSFET have been compared in Table 1. From Table 1 it has been observed that the device D7 shows the highest immunity to short channel effects (SCEs). As the threshold voltage and on current of device D7 is highest among all other device cases and the subthreshold swing (SS) and drain induced barrier lowering (DIBL) of device D7 is lowest among all device cases. Hence, the DP-JLDG MOSFET incorporating DP inside the channel region only at the drain side can be used for low power circuit applications.
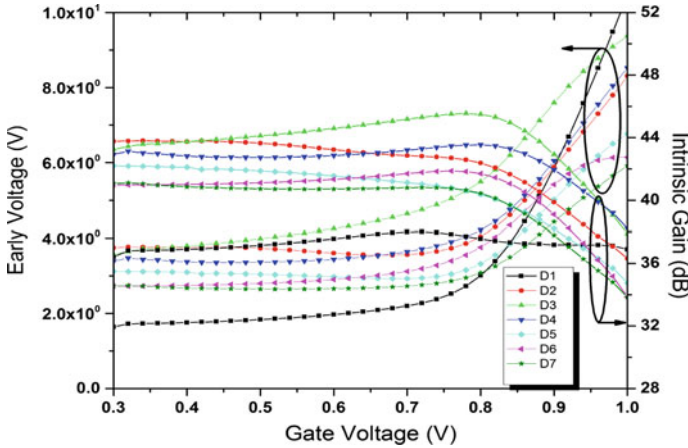
Fig. 5 The plot of early voltage and intrinsic gain (dB) with gate to source bias voltage for all device cases

Table 1 Electrostatic parameters of DP-JLDG MOSFET for all device cases

Parameters	Devices						
	D1	D2	D3	D4	D5	D6	D7
Threshold voltage (V)	0.649	0.603	0.601	0.585	0.568	0.563	0.701
DIBL (mV/V)	8.866	9.848	10.281	19.367	17.214	14.734	6.742
SS (mV/decade)	63.056	64.142	65.187	65.42	65.504	65.805	61.531
I_{on}/I_{off} ($\times 10^{12}$)	33.8	5.48	2.86	2.09	2.02	1.55	71.5

The comparison of analog performance parameters for all device cases of DP-JLDG MOSFET is shown in Table 2. From Table 2 it has been observed that the transconductance, output conductance, and early voltage of device D1 is highest among all device cases. The intrinsic gain of device D3 is highest among all cases of DP-JLDG MOSFET and the TGF of device D7 is highest among all device cases.

Table 2 Analog performance parameters for all device cases of DP-JLDG MOSFET

Parameters	Devices						
	D1	D2	D3	D4	D5	D6	D7
g_m (mA/V)	1.55	1.08	1.06	0.994	1.39	1.37	1.18
g_d (mA/V)	1.36	0.948	0.894	0.870	1.07	1.07	1.05
V_{EA} (V)	10.47	8.31	9.38	8.52	6.76	6.14	5.92
A_V	38.05	43.78	45.54	43.52	42.20	41.70	41.07
TGF (V^{-1})	40.25	41.22	40.72	41.31	41.28	41.07	41.45

The plot of intrinsic capacitances with the gate to source voltage is shown in Fig. 6. From Fig. 6, it has been observed that the intrinsic capacitance increases with application of DP at the boundary of source/drain regime. The plot of cutoff frequency and gain transconductance frequency product (GTFP) with gate to source voltage for all device cases is shown in Fig. 7. With the application of DP cutoff frequency
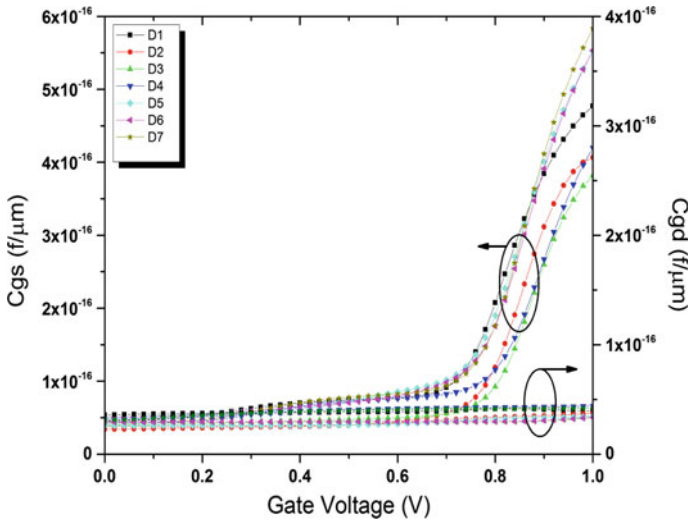
Fig. 6 The plot of intrinsic capacitances with gate to source voltage
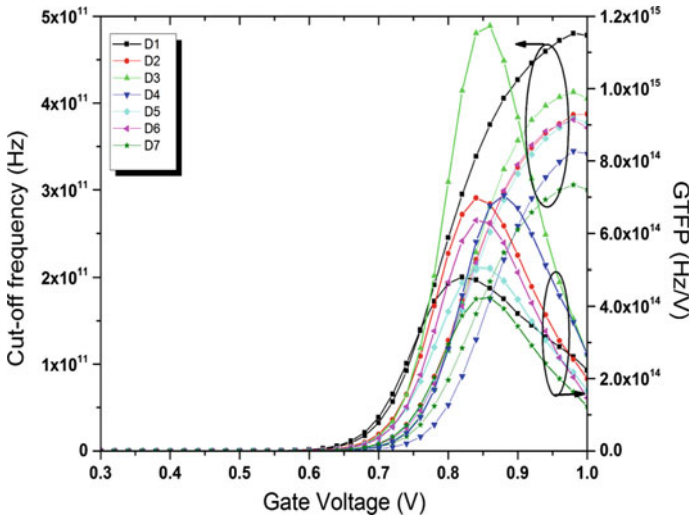
Fig. 7 The plot of cutoff frequency and gain transconductance frequency product (GTFP) with gate to source voltage
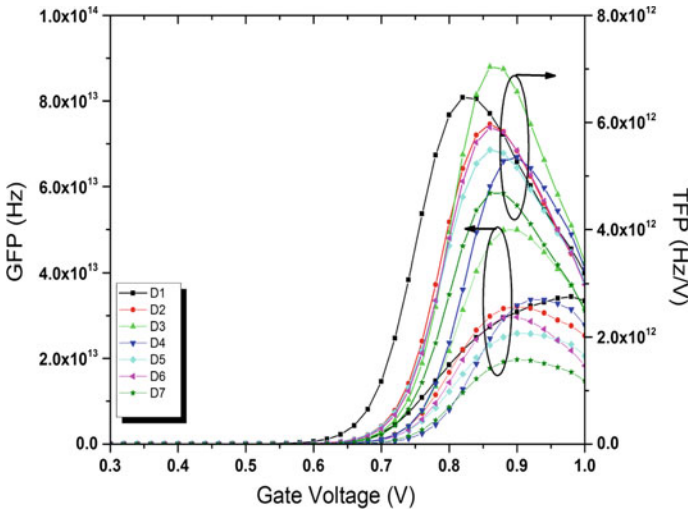
Fig. 8 The plot of gain frequency product (GFP) and transconductance frequency product (TFP) with gate to source voltage

decreases whereas GTFP increases. Figure 8 demonstrates the gain frequency product (GFP) and transconductance frequency product (TFP) with gate to source voltage. From Fig. 8, it has been examined that both GFP and TFP increases with application of DP. The speed of a device can be measured with the carrier transit time. The transit time of a MOSFET is inverse of its cutoff frequency. Figure 9 shows the variation of transit time with gate t source voltage for all device cases and with the application of DP transit time of the device increases.

The RF performance parameters of DP-JLDG MOSFET for all device cases have been compared in Table 3. From Table 3 it is observed that the parasitic capacitances of device D3 are lowest among all devices. The cutoff frequency of device D1 is highest, i.e., the cutoff frequency reduces with the application of DP. Further, it has also been observed from Table 3 that GFP, TFP, and GTFP of device D3 is highest among all device cases as the reduction in the parasitic capacitances results in the enhancement of these RF performance parameters.

The plot of second-order transconductance (g_{m2}) and third-order transconductance (g_{m3}) are shown in Figs. 10 and 11 respectively. From Figs. 10 and 11 it is observed that both g_{m2} and g_{m3} reduces with the application of DP because of the reduction in the on current of the device. The plots of voltage intercept point (VIP2 and VIP3) are shown in Figs. 12 and 13 respectively. From Figs. 12 and 13 it has been examined that the linearity performance of the device degrades with the application of DP. Figures 14 and 15 demonstrate the plot of IIP3 and IMD3 with gate to source bias voltage for all device cases. With the application of DP both IIP3 and IMD3 decreases which results in the nonlinear behavior of the device.

Figure 16 demonstrates the plot of thermal noise with a gate to source bias voltage. It has been observed from Fig. 16 that the thermal noise of a device is almost constant
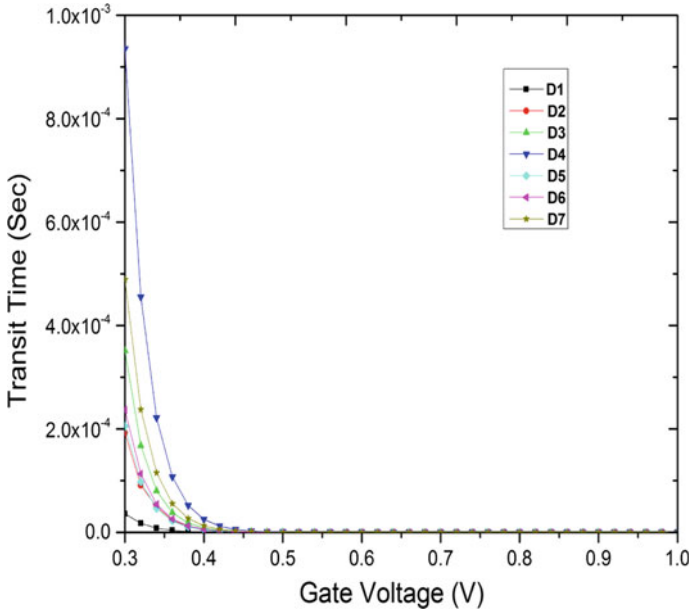
Fig. 9 The variation of transit time with gate t source voltage for all device cases

Table 3 Comparison of RF performance parameters of DP-JLDG MOSFET for all device cases

Parameters	Devices						
	D1	D2	D3	D4	D5	D6	D7
Cgs (aF)	477	407	382	420	552	553	583
Cgd (aF)	416	374	346	439	346	336	419
f_T (GHz)	480	387	513	345	382	381	306
GFP (THz)	34.4	32.2	49.9	33.7	25.9	29.7	19.7
TFP (THz/V)	6.47	5.97	7.04	5.35	5.49	5.91	4.68
GTFP (THz/V)	481	698	1170	705	505	636	424

in subthreshold regime and it starts diminishing in the saturation region. It has also been examined from Fig. 16 that the thermal noise of a JLDG MOSFET increases with application of dielectric pocket at the boundary of source/drain regime.

5 Conclusion

In this paper, we have investigated the noise and linear distortion analysis of DP-JLDG MOSFET considering all possible applications of DP at the edge of
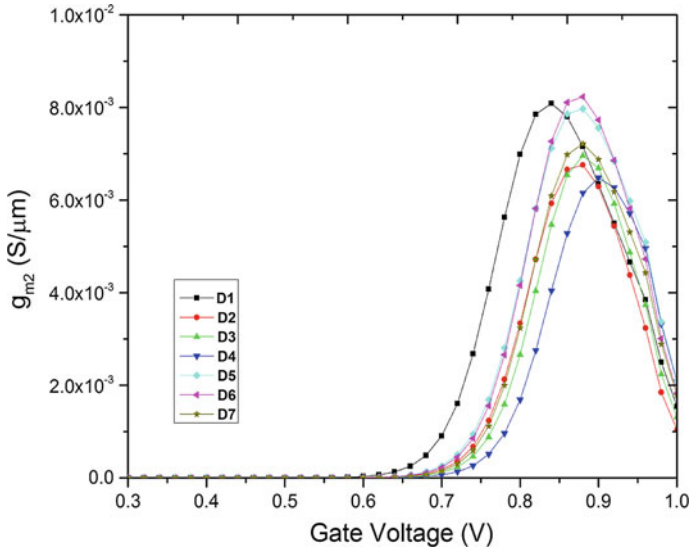
Fig. 10 The plot of g_{m2} with gate voltage
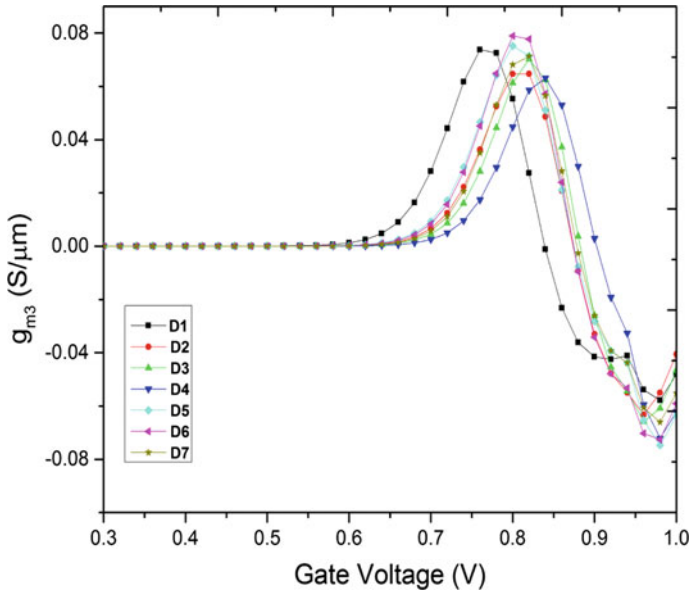
Fig. 11 The plot of g_{m3} with gate voltage
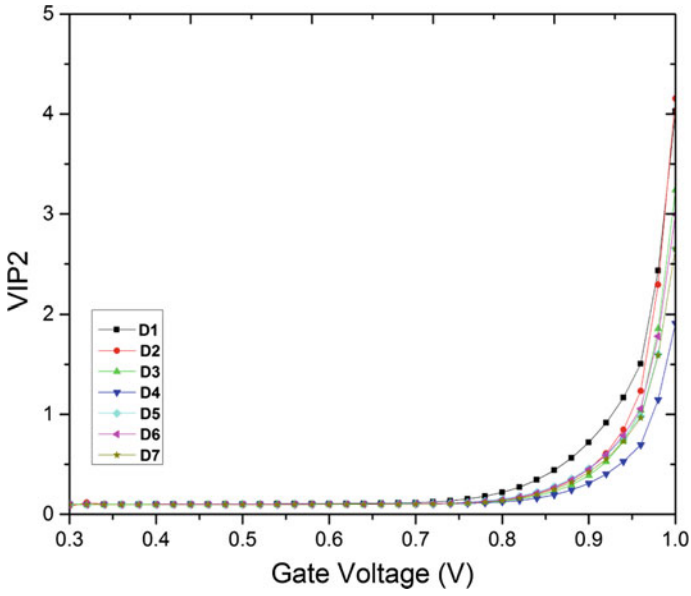
Fig. 12 The plot of VIP2 with gate voltage
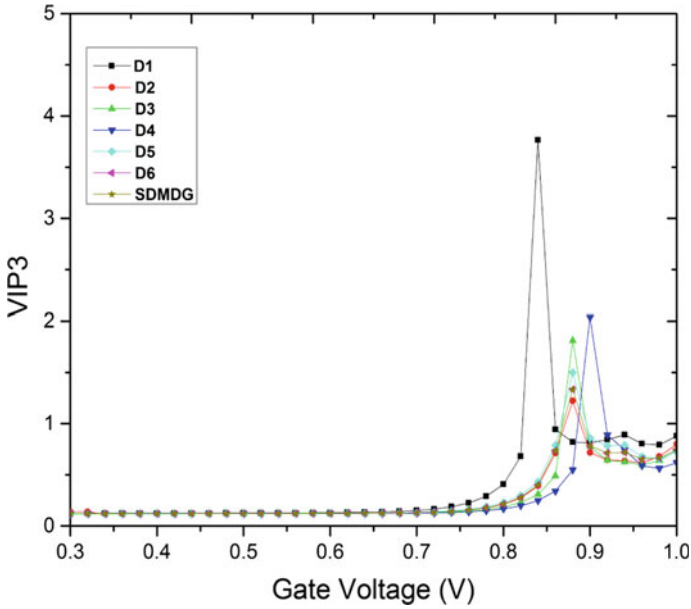
Fig. 13 The plot of VIP3 with gate voltage
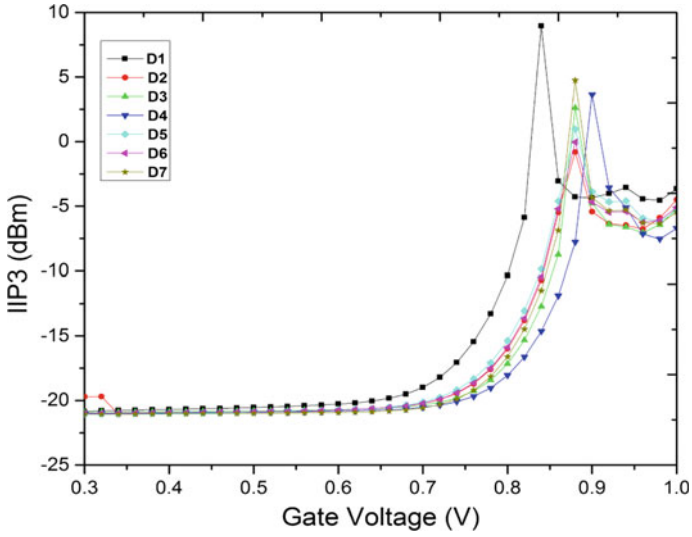
Fig. 14 The plot of IIP3 with gate voltage
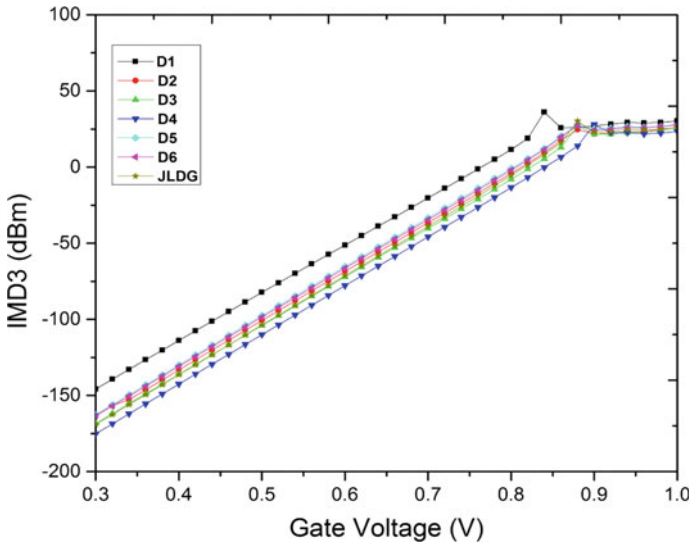
Fig. 15 The plot of IMD3 with gate voltage
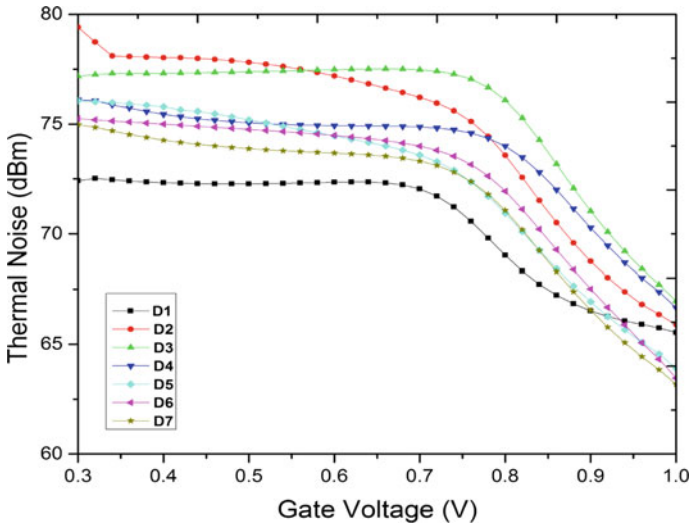
Fig. 16 The plot of thermal noise with gate to source voltage for all device cases of DP-JLDG MOSFET

source/drain regime. The electrostatic, analog/RF, linearity parameters, and thermal noise of DP-JLDG MOSFET have been compared with the JLDG MOSFET through extensive device simulation. The results reveal that the DP-JLDG MOSFET device D7, i.e., the DP-JLDG MOSFET incorporating DP inside the channel region only when the drain side offers an excellent electrostatic performance compared to all other structures. The early voltage of device D1 is highest among all other devices which show that there is very less impact of channel length modulation in the saturation region. The intrinsic gain and gain-bandwidth product of device D3 is highest among all DP cases which prove its capability to operate for higher frequency applications. Hence, the proposed device offers excellent analog/RF behavior and can operate over a wide range of frequency with low power dissipation. The results also show that the thermal noise and the linearity performance of the device degrade with the application of the DP. The proposed device is an appropriate alternative for low power analog and digital circuits due to its higher device gain, larger operating range, and lower power dissipation.

References

1. Wang, Y., Shan, C., Dou, Z., Wang, L., Cao, F.: Improved performance of nanoscale junctionless transistor based on gate engineering approach. Microelectron. Reliab. **55**(2), 318–325 (2015)
2. Colinge, J.-P.: Multiple-gate SOI MOSFETs. Solid-State Electron. **48**(6), 897–905 (2004)
3. Kim, J.-J., Roy, K.: Double gate-MOSFET subthreshold circuit for ultralow power applications. IEEE Trans. Electron Devices **51**(9), 1468–1474 (2004)

4. Lemme, M.C., Mollenhauer, T., Henschel, W., Wahlbrink, T., Baus, M., Winkler, O., Granzner, R., Schwierz, F., Spangenberg, B., Kurz, H.: Subthreshold behavior of triple-gate MOSFETs on SOI material. Solid-State Electron. **48**(4), 529–534 (2004)
5. Chaudhry, A., Jagadesh Kumar, M.: Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review. IEEE Trans. Device Mater. Reliab. **4**(1), 99–109 (2004)
6. Sharma, R.K., Gupta, M., Gupta, R.S.: TCAD assessment of device design technologies for enhanced performance of nanoscale DG MOSFET. IEEE Trans. Electron Devices **58**(9), 2936–2943 (2011)
7. Gupta, S.K., Baidya, A., Baishya, S.: Simulation and analysis of gate engineered triple metal double gate (TM-DG) MOSFET for diminished short channel effects. Int. J. Adv. Sci. Technol. **38**, 15–24 (2012)
8. Roy, N.C., Gupta, A., Rai, S.: Analytical surface potential modeling and simulation of junction-less double gate (JLDG) MOSFET for ultra low-power analog/RF circuits. Microelectron. J. **46**(10), 916–922 (2015)
9. Reddy, G.V., Kumar, M.J.: Investigation of the novel attributes of a single-halo double gate SOI MOSFET: 2D simulation study. Microelectron. J. **35**(9), 761–765 (2004)
10. Wu, J.-W., Cheng, C.-C., Chiu, K.-L., Guo, J.-C., Lien, W.-Y., Chang, C.-S., Huang, G.-W., Wang, T.: Pocket implantation effect on drain current flicker noise in analog nMOSFET devices. IEEE Trans. Electron Devices **51**(8), 1262–1266 (2004)
11. Kumari, V., Saxena, M., Gupta, R.S., Gupta, M.: Two-dimensional analytical drain current model for double-gate MOSFET incorporating dielectric pocket. IEEE Trans. Electron Devices **59**(10), 2567–2574 (2012)
12. Jurczak, M., Skotnicki, T., Gwoziecki, R., Paoli, M., Tormen, B., Ribot, P., Dutartre, D., Monfray, S., Galvier, J.: Dielectric pockets—a new concept of the junctions for deca-nanometric CMOS devices. IEEE Trans. Electron Devices **48**(8), 1770–1775 (2001)
13. Jayanarayanan, S.K., Dey, S., Donnelly, J.P., Banerjee, S.K.: A novel 50 nm vertical MOSFET with dielectric pocket. Solid State Electron. **50**(5), 897–900 (2006)
14. Gili, E., Uchino, T., Hakim, M.M.A., De Groot, C.H., Buiu, O., Hall, S., Ashburn, P.: Shallow junctions on pillar sidewalls for sub-100-nm vertical MOSFETs. IEEE Electron Device Lett. **27**(8), 692–695 (2006)
15. Park, J.-T., Colinge, J.-P.: Multiple-gate SOI MOSFETs: device design guidelines. IEEE Trans. Electron Devices **49**(12), 2222–2229 (2002)
16. Panda, D.K., Lenka, T.R.: Oxide dependent compact model of channel noise for Emode AlGaN/GaN MOS HEMT. AEU Int. J. Electron. Commun. **82**, 467–473 (2017)
17. Mohammadi, H., Naderi, A.: A novel SOI-MESFET with parallel oxide metal layers for high voltage and radio frequency applications. AEU Int. J. Electron. Commun. **83**, 541–548 (2018)
18. Atlas User's Manual: Device Simulation Software (2008)

Finite Element Modeling of a Wideband Piezoelectric Energy Harvester for Ambient Vibration Extraction

Prateek Asthana, Apoorva Dwivedi and Gargi Khanna

Abstract The work proposes a novel piezoelectric energy harvester (PEH) based on the seesaw mechanism. The seesaw structure has been utilized in PEH for extracting a wide range of ambient vibrations and converting them into usable electrical energy. The proposed structure operates on the first two resonant frequencies in comparison to the conventional cantilever system that works on the first resonant frequency. Eigen frequency, as well as response to a varying input vibration frequency, is carried out showing better performance of seesaw cantilever design. The harvester efficiently converts mechanical energy into electrical energy while reaching a power output of 0.55 mW with a wider band of operation.

Keywords Finite element modeling · Piezoelectric energy harvester · Vibration · Analytical modeling · Resonant frequency · Renewable energy

1 Introduction

The need for self-powered devices, has arisen in recent times due to the limited capacity of the battery and inconvenience of replacing or recharging batteries, which have a fixed lifetime as compared to energy harvesting techniques lasting the system lifetime. A new perspective to power supply for wireless devices is energy harvesters. There are limited energy sources available on earth, hence energy harvesting techniques provide a nonconventional energy resources being researched at an increasing rate. The advantage of microscale energy harvesting is package minimization which leads to ease of integration with other systems [4, 5, 9]. Real-life applications require continuous power supply to function throughout their lifetime; hence, energy harvesting is the best power supply for these applications. Macro-energy harvesters

P. Asthana (✉) · A. Dwivedi · G. Khanna
Department of Electronics and Communication Engineering,
National Institute of Technology, Hamirpur, India
e-mail: prateekasthana1989@gmail.com

549
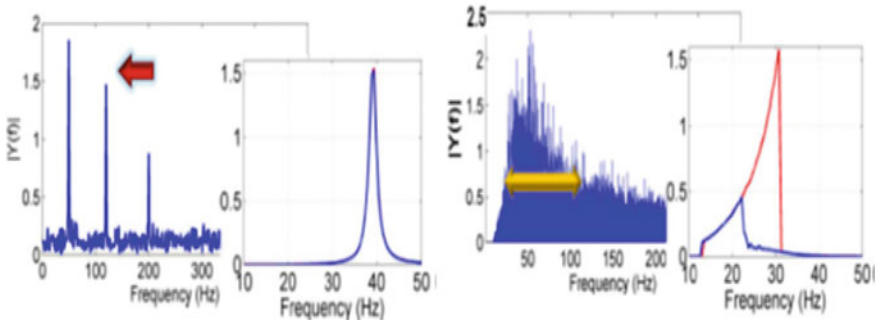
Fig. 1 Frequency band for ambient vibration extraction. **a** Narrowband frequency extraction. **b** Wideband frequency extraction [1]

have been in practical use for a long time, but for independent miniature electronic devices, a need for sufficient power supply to help them function has been an area of research for the past few years. Vibration is a viable source of energy. There are many ambient vibration sources from which energy can be extracted and utilized to power microelectronic devices like wireless sensor nodes. Wideband extraction is an important criterion for extraction of ambient vibrations. Ambient vibrations are random in nature; hence, in order to harvest power from these sources, a wideband energy harvester is essential. Figure 1 shows the need for wideband extraction as compared to narrowband extraction. Some of the sources for wideband vibration extraction are busy street, busy railway platforms, train tracks, bridges, etc. While the sources for narrowband extraction include car engines, electrical appliances, etc.

A study of various harvesters utilized for wideband vibration extraction have been described in Table 1. It can be analyzed from the Table that in order to provide a wideband, nonlinearities were being introduced in the system, these nonlinearities

Table 1 Piezoelectric energy harvesters

Mechanism	Volume (cm^3)	Frequency (Hz)	Bandwidth (Hz)	Acceleration	Peak power (μW)
Linear	4.38	58.6	1–2	$0.5g$, $0.1g$	1153
Monostable nonlinearity	4.38	58	10@$0.5g$	$0.5g$, $0.3g$	1330
Combined bistable and monostable nonlinearity	7.2	65.8	10@$0.5g$	$0.5g$, $1g$, $0.2g$	1403
Bistable nonlinearity	2.56	35.2	6.2@$0.5g$	$0.5g$, $0.2g$	29.2
MEMS/magnetically tunable	0.11	230	12@$0.125g$	$0.125g$	8.45
MEMS/monostable nonlinearity	0.14	630	82@$0.5g$	$0.5g$	2.5
3D printed/softening monostable nonlinearity	6	150	4.5@$1g$	$0.1g$	2500
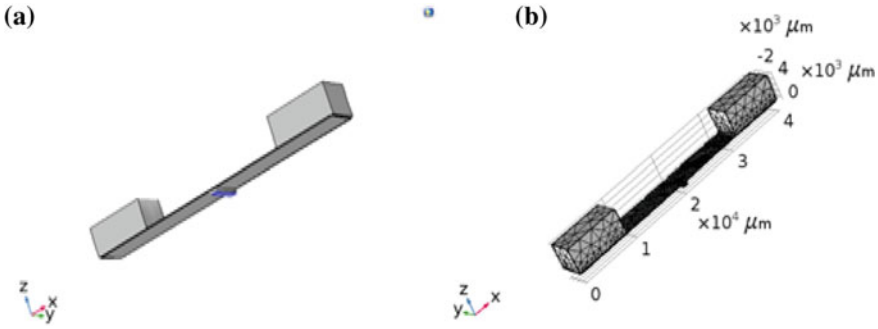
Fig. 2 a Finite element model of seesaw piezoelectric energy harvester. **b** Meshed model of PEH

increase the bandwidth but at the same time make the system hybrid and bulky. In this work, we have proposed a linear wideband energy harvester. The harvester is based on a seesaw-type lever as shown in Fig. 2.

The paper is further subdivided into two main sections. The first section introduces the dynamics of the device and the second discusses the results and findings.

2 Device Dynamics

Device dynamics consist of operation of device, device dimensions, and resonant frequency analysis of device carried out on COMSOL Multiphysics.

2.1 Operation of Device

A seesaw device is implemented to harvest ambient mechanical vibration and convert them into usable electrical energy. A seesaw device has two arms with both contributing equally in power generation; when one arm goes up, the other goes down like a seesaw swing in a playground. The first two resonant frequency together form the bandwidth of the design [3].

Numerical Model

The working of device depends on its resonant frequency. When the resonant frequency matches the ambient frequency, then the maximum power is obtained. In case of seesaw mechanism, the wide bandwidth is obtained because the first two resonant frequencies of the structure are very close [6, 7]. The resonant frequency is derived with the help of Euler–Bernoulli beam theory. Following assumptions are made in the analytical modeling on the constitutive piezoelectricity relations: (a) Beam assumption for Euler–Bernoulli; (b) negligible air damping for external

excitation; (c) proportional damping (i.e., viscous air damping and strain rate damping are assumed to be proportional to bending stiffness and mass per length); and (d) uniform electric field through the piezoelectric thickness [12]. The piezoelectric constitutive equations are given as

$$\sigma = sX + dE \tag{1}$$

$$D = dX + \epsilon E \tag{2}$$

where s is elasticity modulus (Pa) and σ is stress (Pa) that is dependent on electric field E (N/C) and strain X, while d is the piezoelectric constant (m/V). Electric field and strain determine the dielectric displacement D. ϵ is absolute permittivity of the piezoelectric material (F/m). Assumptions (a) and (b) determine the governing equation of mechanical motion as

$$YI\frac{\partial^4 w_r el(x,t)}{\partial x^4} + c_a\frac{\partial w_{rel}(x,t)}{\partial t} + kv(t)(\frac{d\delta(x)}{dx} - \frac{d\delta(x-L)}{dx}) +$$
$$m\frac{\partial^2 w_{rel}(x,t)}{\partial t^2} + c_{sI}\frac{\partial^5 w_{rel}(x,t)}{\partial t \partial x^4} = -[m + M_t\delta(x-L)\frac{\partial^2 w_b(x,t)}{\partial t^2} \tag{3}$$

where I is moment of inertia of beam; $M(x,t)$ is the internal bending moment of the beam; $w_{rel}(xt)$ and $w_b(xt)$ are deflections relative to the base motion and base excitation, respectively; c_a and c_s are coefficients for viscous air damping and strain rate damping, respectively, and m is mass per unit length [10–13]. The bending moment of the piezoelectric layer [14, 15] is given by

$$D_p = \frac{E_p^2 t_p^4 + E_{np}^2 t_{np}^4 + E_p t_{np}(2t_p^2 + 2t_{np}^2 + 3t_p t_{np})}{12(E_p t_p + E_{np} t_{np})} \tag{4}$$

The natural transverse vibration is written as

$$f_i = \frac{v_i^2}{2\pi\sqrt{2}}(\frac{h}{l_b^2})\sqrt{\frac{e_T}{\rho}} \tag{5}$$

The Bernoulli–Euler equation can be derived in a term related to bending modulus as

$$f_i = \frac{v_i^2}{2\pi l_b^2}\sqrt{\frac{D}{m_w}} \tag{6}$$

with v_i for the first three modes are: $v_1 = 1.8751$, $v_2 = 4.6941$, and $v_3 = 7.857$. The first natural frequency of the unimorph structure is calculated as

$$f_N = \frac{0.1615}{l_b^2}\sqrt{\frac{D_p}{m}} \tag{7}$$

The natural frequency of cantilever with proof mass, f_M is

$$f_M = f_N \sqrt{\frac{m_{eff}}{m_{eff} + M_t}} \tag{8}$$

M_t is the additional proof mass while the effective mass at the tip of the cantilever structure is m_{eff} [8, 15]. Effective mass is given by

$$m_{eff} = 0.236 \rho_b w_b h_b l_b \tag{9}$$

2.2 Device Dimensions

Device dimensions play a key role in determining the performance of piezoelectric energy harvester. The device dimensions of the seesaw structure have been described in Table 2. Device parameters have been designed to target the natural frequency of around 100 Hz. Finite element model and meshed model of the design have been shown in Fig. 2.

2.3 Resonant Frequency Analysis

The models are simulated using COMSOL Multiphysics in 3D. The physics included piezoelectric devices, which implements both electrostatics and solid mechanics, and an additional electric circuit study is added to determine the voltage generated across the resistor. Resonant frequency of the design is shown in Fig. 3. The geometrical parameters of the model are given in Table 2. Corresponding to ambient vibrations,

Table 2 Device parameters for piezoelectric energy harvester

Parameter name	Parameter symbol	Parameter value (mm)
Length of device	L	40
Length of piezo	L_p	40
Width of device	w_d	4
Thickness of piezo	t_p	0.06
Thickness of substrate	t_s	0.04
Length of proof mass	L_m	9
Thickness of proof mass	t_m	4
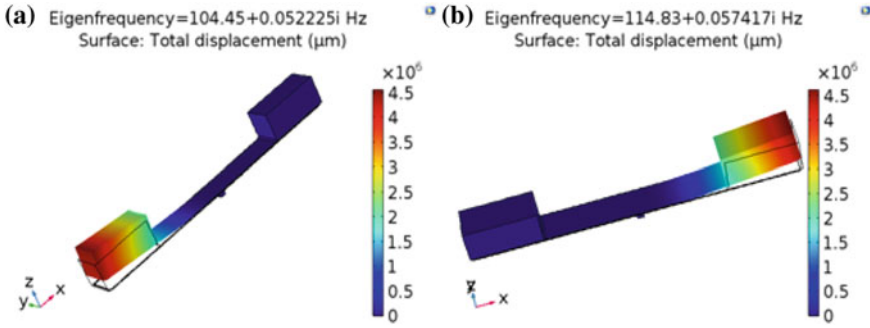
Fig. 3 **a** First mode shape of seesaw piezoelectric energy harvester. **b** Second mode shape of seesaw piezoelectric energy harvester

the acceleration value of $1g$ is applied to the designed structure [5]. The structure has been designed using silicon as substrate material and Zinc Oxide as piezoelectric material.

3   Result and Discussion

Piezoelectric energy conversion is based on converting mechanical energy into electrical energy. Mechanical energy is represented as the elastic strain energy of the device, while the electrical energy is the stored energy on the piezoelectric layer. The efficiency of the device is the ratio of $E_{harvested}$ to that of E_{stored}. Total elastic strain energy and electrical stored energy are shown in Fig. 4.

$$Eff. = \frac{E_{harvested}}{E_{stored}} = \frac{3.604}{4.561} = 79.03\% \tag{10}$$

The conversion of stress on the piezo layer to electrical power is displayed by electromechanical coupling. Higher is the stress on the piezo layer, higher is the displacement leading to significant conversion to electrical power. Higher power is generated by the piezoelectric layer of the device which leads to an increase in the quantity of accumulated charges and larger potential being developed on the layer. The drop across the 4 Mohm resistor has been plotted in Fig. 5. A peak output voltage for the harvester is produced at a peak current of around 11 A while the peak voltage is around 45 V and a maximum power of 0.57 mW across the load resistor. The peak power across the load resistor is around 0.55 mW. The bandwidth of the harvester is around 20 Hz compared to a conventional harvester of around 6 Hz [2].
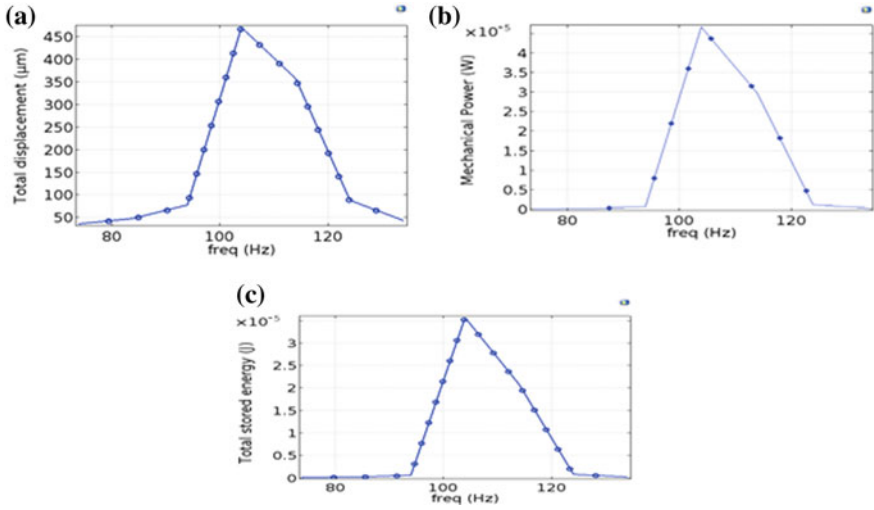
Fig. 4 Energy response of PEH. **a** Displacement of free ends. **b** Elastic strain energy. **c** Electrical stored energy
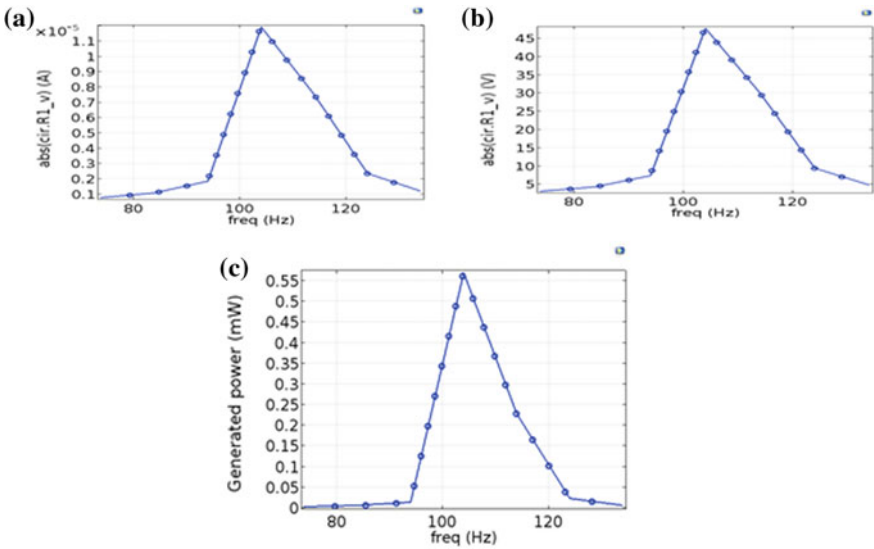
Fig. 5 Output characteristics of a seesaw PEH. **a** Current output. **b** Voltage output. **c** Generated power output

4 Conclusion

A wideband energy harvester with resonant frequency of 104 and 114 Hz is simulated in COMSOL. The harvester has a bandwidth of around 20 Hz while a peak power of around 0.55 mW is generated from ambient vibrations. This prototype can be further optimized to produce more power and better performance. A further energy harvesting circuitry techniques like synchronized switch harvesting on inductor (SSHI) and synchronous charge extraction (SCE) can be applied on it to further amplify the output and improve the performance of the energy harvester.

References

1. Aghakhani, A., Basdogan, I.: Equivalent impedance electroelastic modeling of multiple piezo-patch energy harvesters on a thin plate with ac-dc conversion. IEEE/ASME Trans. Mech. **22**(4), 1575–1584 (2017)
2. Asthana, P., Khanna, G.: Design and analysis of slotted cantilever structure for piezoelectric energy harvesting. In: 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), pp. 387–390. IEEE (2017)
3. Asthana, P., Khanna, G.: Finite element modelling of piezoelectric energy harvesters using lead-based and lead-free materials for voltage generation. J. Asian Ceram. Soc. (just-accepted) (2018)
4. Beeby, S.P., Tudor, M.J., White, N.: Energy harvesting vibration sources for microsystems applications. Meas. Sci. Technol. **17**(12), R175 (2006)
5. Caliò, R., Rongala, U.B., Camboni, D., Milazzo, M., Stefanini, C., De Petris, G., Oddo, C.M.: Piezoelectric energy harvesting solutions. Sensors **14**(3), 4755–4790 (2014)
6. De Giuseppe, G., Centuori, A., Malvasi, A.: An improved PZT cantilever SPICE model for practical energy harvesting circuits simulations and measurements. Measurement **98**, 374–383 (2017)
7. Erturk, A., Inman, D.J.: A distributed parameter electromechanical model for cantilevered piezoelectric energy harvesters. J. Vib. Acoust. **130**(4), 041002 (2008)
8. Hehn, T., Manoli, Y.: Piezoelectricity and energy harvester modelling. In: CMOS Circuits for Piezoelectric Energy Harvesters, pp. 21–40. Springer (2015)
9. Khanna, G., Asthana, P.: A comparative study of circuit for piezo-electric energy harvesting. In: 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1689–1694. IEEE (2016)
10. Kim, S.G., Priya, S., Kanno, I.: Piezoelectric mems for energy harvesting. MRS Bull. **37**(11), 1039–1050 (2012)
11. Li, Z., Zhou, G., Zhu, Z., Li, W.: A study on the power generation capacity of piezoelectric energy harvesters with different fixation modes and adjustment methods. Energies **9**(2), 98 (2016)
12. Liu, H., Tay, C.J., Quan, C., Kobayashi, T., Lee, C.: Piezoelectric MEMS energy harvester for low-frequency vibrations with wideband operation range and steadily increased output power. J. Microelectromech. Syst. **20**(5), 1131–1142 (2011)
13. Sodano, H.A., Inman, D.J., Park, G.: Comparison of piezoelectric energy harvesting devices for recharging batteries. J. Intell. Mater. Syst. Struct. **16**(10), 799–807 (2005)
14. Tilmans, H.A.: Equivalent circuit representation of electromechanical transducers: I. lumped-parameter systems. J. Micromech. Microeng. **6**(1), 157 (1996)
15. Yi, J.W., Shih, W.Y., Shih, W.H.: Effect of length, width, and mode on the mass detection sensitivity of piezoelectric unimorph cantilevers. J. Appl. Phys. **91**(3), 1680–1686 (2002)

Impact of Oxide Engineering on Analog/RF Performance of Doping-Less DMDG MOSFET

Abhinav Gupta, Amrish Kumar, Sanjeev Rai and Rajeev Tripathi

Abstract In order to reduce short channel effects, a dual metal concept has been employed in MOSFETs. But to eliminate the problem of a gate tunneling dual material double gate (DMDG) MOSFET is not sufficient. Hence, to overcome gate tunneling oxide engineering technique has been employed in a DMDG MOSFET. In this paper, a doping-less dual material double gate (DL-DMDG) MOSFET has been analyzed using oxide engineering technique. To induce an n-type substrate in a doping-less MOSFET, charge plasma concept has been incorporated. Using 2D ATLAS simulator various analog/RF parameters have been investigated for this device with different oxide materials. The basic purpose of this paper is to improve analog/RF performance of the device and to increase immunity to SCEs.

Keywords DL-DMDG MOSFET · Charge plasma · SCEs · Analog/RF

1 Introduction

The MOSFETs have become the most important and basic building block of VLSI circuits. The demand for higher packing density, high-speed and low cost requires scaling of the MOSFET. Scaling of MOSFETs in nano-scale regime results in the reduction of the distance between the source and drain due to which gate electrode fails to control over the complete channel region because of charge sharing between the drain and source extension [1, 2]. Hence, continuous scaling of MOSFET increases the short channel effects (SCEs) [3, 4] and thus degrades the performance of the device. As the electron drift characteristics in the channel and the threshold voltage variation is limited by the SCEs. Various methods such as channel engineering, strain engineering and work function engineering techniques have been

A. Gupta (✉)
Electronics Engineering Department, Rajkiya Engineering College, Sonbhadra 231206, India
e-mail: abhinavkit87@gmail.com

A. Kumar · S. Rai · R. Tripathi
Department of Electronics and Communication Engineering,
Motilal Nehru National Institute of Technology, Allahabad 211004, India

employed by many researchers to enhance the performance of the device and to avoid the undesirable SCEs [5, 6].

Due to excellent immunity to SCEs, double gate (DG) devices provide better scalability in the sub-nano-meter regime. In 2010, Mohankumar et al. [7] presented double gate MOSFET using the gate and channel engineering and analyzed its analog/RF performance parameters. DG MOSFETs shows improvement in analog/RF parameters compared to the single gate. In 2011, Sharma [8] proposed a device called Nano-scale DG MOSFET and analyzed this device using 2D ATLAS simulator. However, if channel length of the device reduced below 100 nm, double gate device still shows significant drain induced barrier lowering (DIBL) and threshold voltage roll-off. In order to eliminate above problem, Dual metal gate concept has been employed in MOSFETs in which two different materials of different work function has been used as a gate electrode. Further, the major issues associated with conventional doped devices are random dopant fluctuations (RDF) [9] which degrades the device performance and increases gate leakage current. To overcome above problem and to improve mobility and current driving capability of the device doping-less technology [8] has been incorporated. The doping-less MOSFET is basically an MOS transistor in which intrinsic substrate has been used and with the help of charge plasma [10] concept, an n-type plasma has been induced in the channel for conduction mechanism.

In this paper, analog/RF performance of DL-DMDG MOSFET [11] using oxide engineering has been investigated. The concept of the Dual metal gate has been incorporated in the device and due to work function difference between two gate metal, there is a step in potential profile in channel region due to which SCEs reduces and trans-conductance improves. Due to scaling oxide thickness reduces, which leads to gate leakage current due to tunneling phenomena. Hence, to overcome above problem double gate oxide (DGO) [12] concept has been incorporated in this device, in which the oxide is divided into two regions in such a manner that the permittivity of region 2 (near the drain) is greater than permittivity of region 1, i.e., near the source end ($\varepsilon_2 > \varepsilon_1$). Si–Ge material has been used as a channel for enhancement of mobility and drive current of the device [13]. Charge plasma has been used as the main conduction mechanism in a doping-less MOSFET. When metal of different work function and semiconductor are brought in contact with each other then, p-type or n-type charge plasma is generated in the semiconductor film based on the metal work function [14]. As the device is doping-less it is free from impurities due to which leakage current has been reduced which further reduces the SCEs of the device as compared to the conventional doped device.

In this paper, numerical simulation has been performed in order to analyze the impact of oxide engineering on electrostatic, analog/RF parameters of DL-DMDG MOSFET by using 2-D Atlas simulator. Along with the introduction, in Sect. 2 device structure and its specifications have been described. Section 3 consists of simulation results for electrostatic and analog/RF parameters of doping-less DMDG MOSFET. Finally, in Sect. 4 conclusion has been conferred.

2   Device Structure and Simulation Setup

A doping-less dual material double gate (DL-DMDG) MOSFET is shown in Fig. 1.
In this device charge plasma concept has been employed to induce carriers for con-
duction. When metal and intrinsic semiconductors are amalgamating together then,
at the equilibrium condition, Fermi level of semiconductor and metal are aligned.
If metal work function is lesser than the semiconductor, electrons are transferred to
semiconductor from metal and in this manner n-type plasma has been generated in
an intrinsic semiconductor. And if semiconductor work function is lesser than that
of metal, electrons are transferred from semiconductor to metal and in this manner
p-type plasma has been generated in an intrinsic semiconductor. In order to generate
n-type plasma, work function should be small. Hence, for source/drain contact work
function of 3.9 eV has been taken. In order to decrease gate leakage current, two
different materials as a gate oxide have been used. Different work functions gate
metal (M_1 and M_2) are used, such an arrangement offers simultaneous rise in trans-
conductance and reduced SCEs due to step in the surface potential profile [15]. Si–Ge
material is used as a channel to enhance mobility and current driving capability.

   As the device is symmetrical, the same voltage has been applied to both the front
and back gates. The arrangement of the gate oxide in DL-DMDG MOSFET is 1:1
ratio of SiO_2:HfO_2 with a thickness of 0.9 nm on both sides. Further, channel length
and channel thickness are 20 nm and 5 nm respectively. The device specifications
are shown in Table 1. Work functions employed for two metal gate i.e. control gate
and screening gate are 5.3 eV (M_1) and 4.6 eV (M_2) respectively.
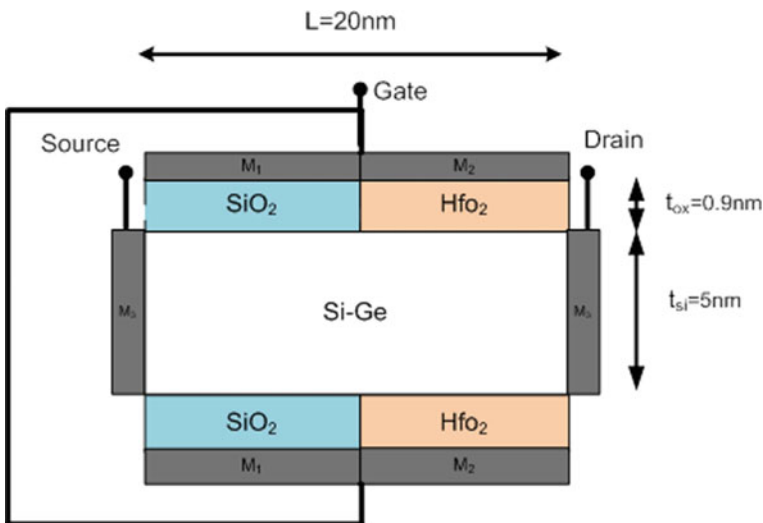
Fig. 1   Cross-sectional view of DL-DMDG MOSFET

Table 1 Device specification for DL-DMDG MOSFET

Parameters	Values
Gate work function (M_1)	5.3 eV
Gate work function (M_2)	4.6 eV
Gate oxide thickness (t_{ox})	0.9 nm
Region 1 oxide length	10 nm
Region 2 oxide length	10 nm
Doping concentration (N_d)	10^{15} cm^{-3}

2-D ATLAS device simulator [16] has been used for evaluating electrostatic and analog/RF parameter of the device. In order to study temperature deviation in simulation, inversion layer constant voltage temperature (CVT) model has been used. Constant mobility model, field mobility model and Boltzmann model are used to simulate electrostatic and analog/RF performance of the device. In addition to this Shockley–Read–Hall (SRH) has been used for generation and recombination of minority carrier to consider the minority carrier leakage currents and Fermi–Dirac model has been used with a rational Chebyshev approximation to get precise results. Parameters like sub-threshold swing, I_{on}/I_{off} ratio, drain induced barrier lowering (DIBL), threshold voltage shows that doping-less device is higher immunity to short channel effects (SCEs) [17].

3 Simulation Results

3.1 Electrostatic Parameters

Figure 2 shows the comparison of drain current of DL-DMDG MOSFET for different region 2 oxide material. From Fig. 2 it has been observed that drain current is high for higher oxide permittivity of region 2. Table 2 shows the electrostatic performance of DL-DMDG MOSFET for different region 2 oxide material. From Table 2 it has been observed that higher the oxide permittivity of region 2, better will be the electrostatic performance of the device. Further, it has been concluded that as oxide permittivity of region 2 increases, the device will be more immune to short channel effects (SCEs).

3.2 Analog/RF Parameters

The most important figure of merit (FOM) for the analog circuit are trans-conductance (g_m), output conductance (g_d), early voltage (V_{ea}) and intrinsic gain (A_v) [18]. Figure 3 represents the variation of drain current and trans-conductance with respect to gate bias voltage for different region 2 oxide materials and from Fig. 3 it has been
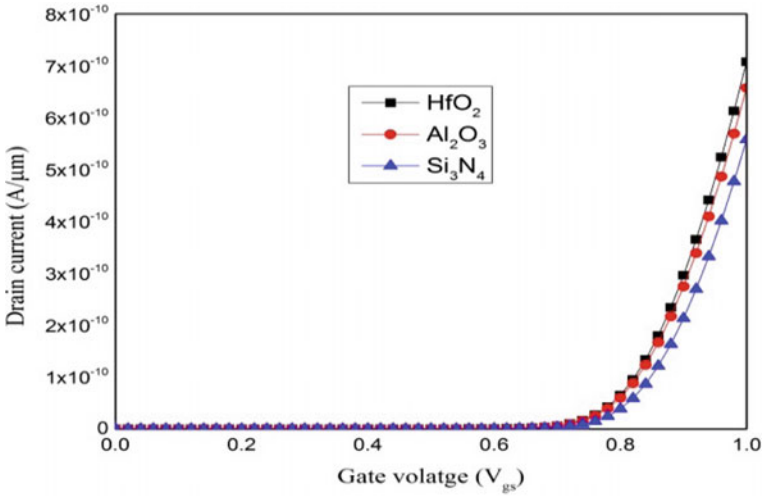
Fig. 2 Comparison of drain current of DL-DMDG MOSFET for different region 2 oxide material

Table 2 Electrostatic parameters of DL-DMDG MOSFET for different region 2 oxide material	Parameters	HfO_2	Al_2O_3	Si_3N_4
	DIBL	3.04	5.66	6.87
	Sub-threshold swing	62.17	79.68	79.89
	I_{on}/I_{off}	8.32×10^{10}	8.02×10^{10}	7.88×10^{10}
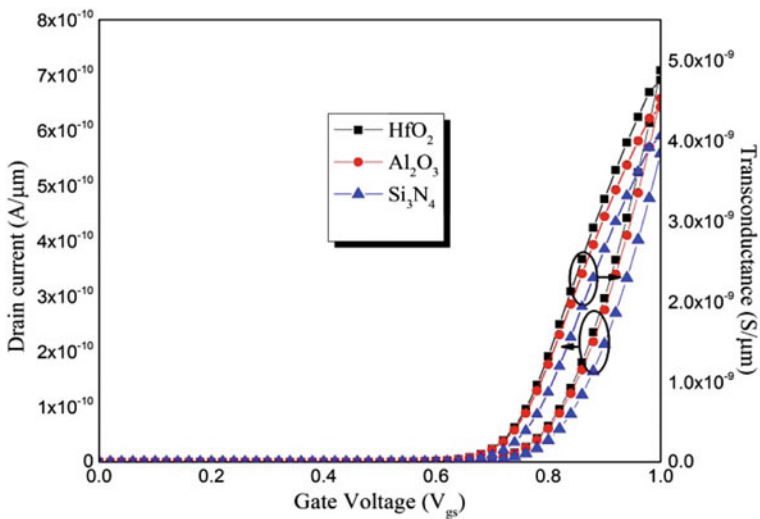
Fig. 3 Variation of drain current and trans-conductance as a function of gate bias voltage for different region 2 oxide material
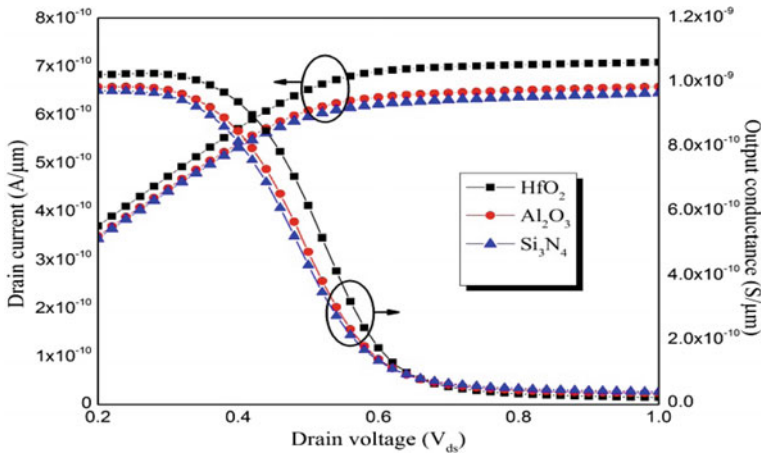
Fig. 4 Variation of drain current and output conductance as a function of gate bias voltage for different region 2 oxide material

concluded that as the oxide permittivity of region 2 increases, drain current as well as trans-conductance increases because leakage current reduces. Figure 4 shows the variation of drain current and output conductance with respect to drain bias voltage for different region 2 oxide materials and here also with increase in oxide permittivity of region 2, output conductance increase and large output-conductance means small resistance which leads to increase drain current.

Figure 5 shows the variation of trans-conductance and Trans-conductance generation factor (TGF) as a function of gate bias voltage for different region 2 oxide material. Figure 6 shows the variation of intrinsic gain and early voltage with gate bias voltage for different region 2 oxide material. From the curve, it has been concluded that for small gate voltage early voltage is high and as the gate voltage increases early voltage decreases and finally become constant. Table 3 shows the maximum values of analog FOMs such as trans-conductance (g_m), output conductance (g_d), early voltage (V_{ea}), intrinsic gain (A_v) and trans-conductance generation factor (TGF) for different region 2 oxide material. From Table 3 it has been concluded that as oxide permittivity of region 2 increases analog performance of the device improves.

The gate to source capacitance (C_{gs}) and gate to drain capacitance (C_{gd}) has been evaluated from ATLAS simulator using 1 MHz frequency with the step size of 0.02 V. Figure 7 shows the variation of intrinsic C_{gs} and C_{gd} as a function of gate voltage for different region 2 oxide material. From Fig. 7 it has been observed that C_{gd} shows little variation with respect to gate bias voltage while C_{gs} increases with increase in gate bias voltage. Further, both the capacitances (C_{gs} and C_{gd}) increases with increase in region 2 oxide permittivity .
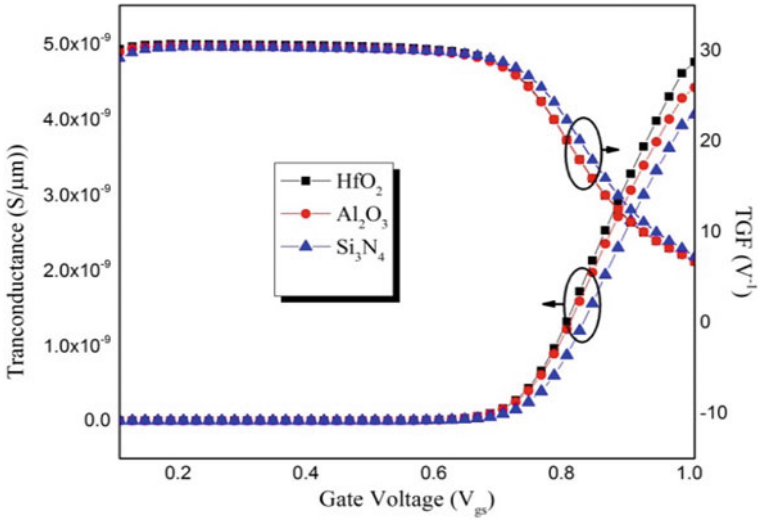
Fig. 5 Variation of trans-conductance and TGF as a function of gate bias voltage for different region 2 oxide material
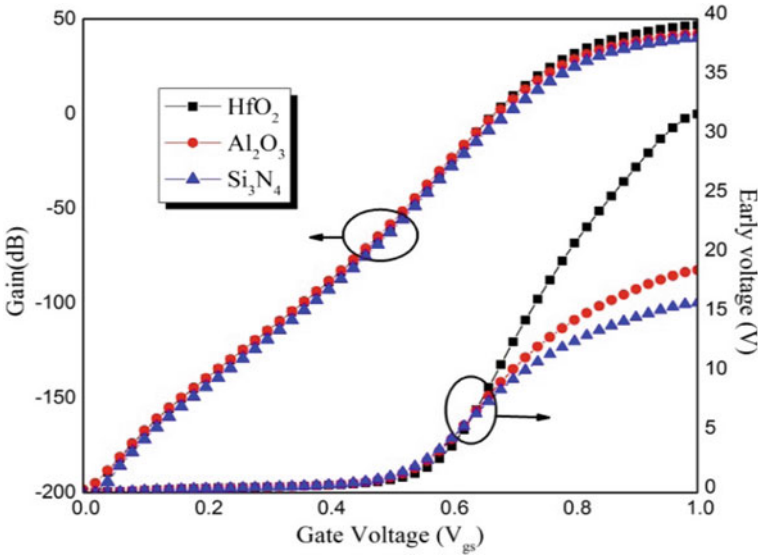
Fig. 6 Variation of intrinsic gain and early voltage as a function of gate bias voltage for different region 2 oxide material

Table 3 Simulated analog FOM for different region 2 oxide material

Parameters	HfO$_2$	Al$_2$O$_3$	Si$_3$N$_4$
Trans-conductance (S)	4.75×10^{-3}	4.1×10^{-3}	4.05×10^{-3}
Output conductance (S)	3.8×10^{-3}	3.5×10^{-3}	3.4×10^{-3}
Early voltage (V)	31.96	18.78	16.00
Intrinsic gain (dB)	46.62	42.00	40.05
TGF (V^{-1})	30.71	30.49	30.31
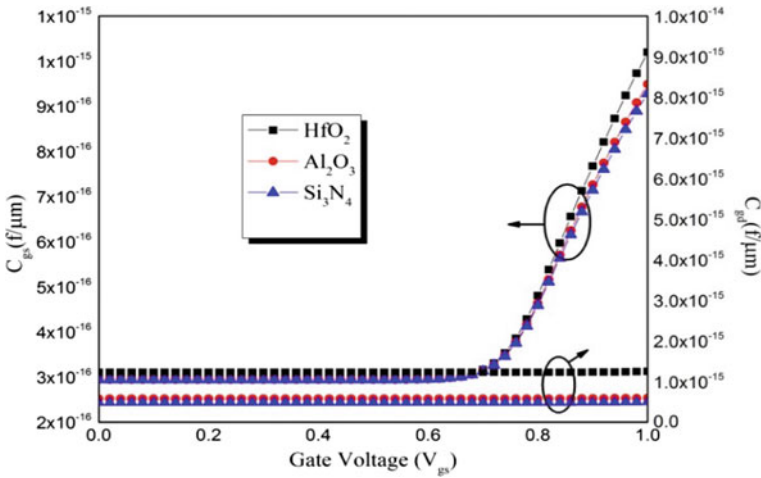
Fig. 7 Variation of C$_{gs}$ and C$_{gd}$ as a function of gate bias voltage for different region 2 oxide material

Figure 8 shows the variation of cut-off frequency (f$_t$) and GTFP as a function of gate bias voltage for different region 2 oxide material. Figure 9 shows the variation of TFP and GFP as a function of gate bias voltage for different region 2 oxide material. Table 4 shows the maximum values of RF FOMs such as cut-off frequency (f$_t$), trans-conductance frequency product (TFP), gain frequency product (GFP), gain trans-conductance frequency product (GTFP) for different region 2 oxide material. From Table 4 it has been observed that with the increase in oxide permittivity of region 2, RF performance of device improves.
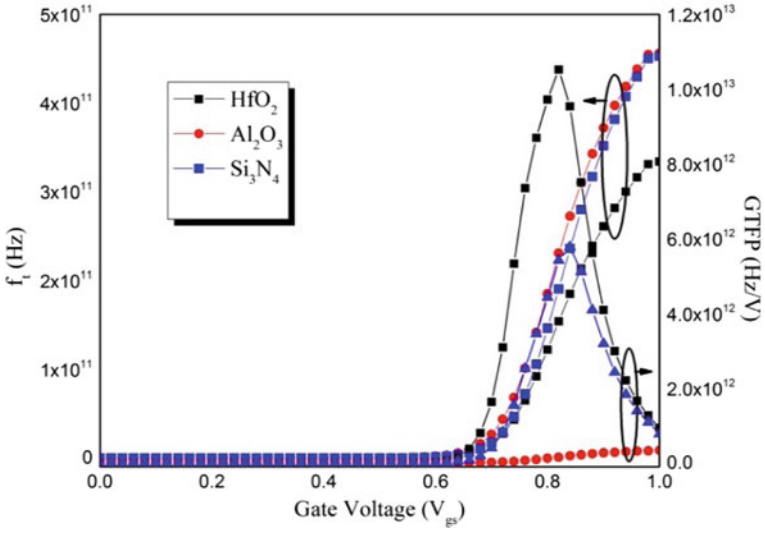
Fig. 8 Variation of Cut-off frequency and GTFP as a function of gate bias voltage for different region 2 oxide material
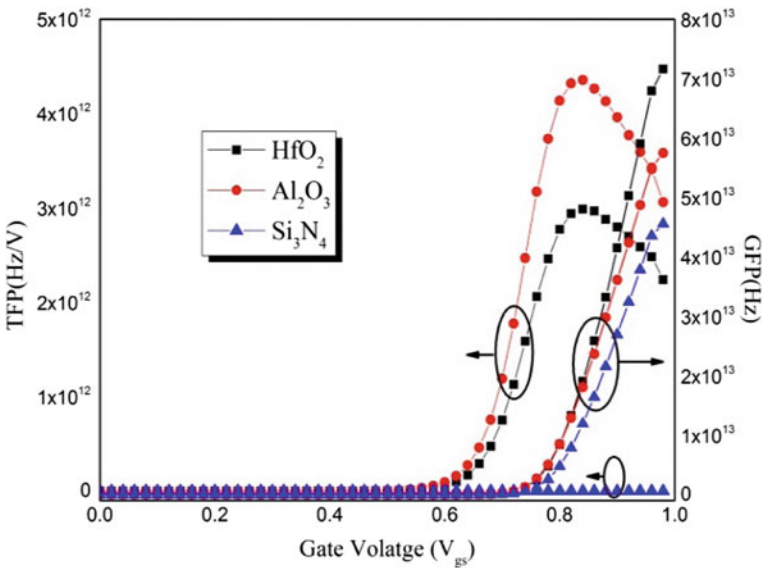
Fig. 9 Variation of TFP and GFP as a function of gate bias voltage for different region 2 oxide material

Table 4 Simulated RF FOM for different region 2 oxide material

Parameters	HfO_2	Al_2O_3	Si_3N_4
Cut-off frequency (Hz)	3.37×10^{11}	4.56×10^{11}	4.52×10^{11}
C_{gs} (F)	2.94×10^{-16}	2.10×10^{-16}	1.94×10^{-16}
C_{gd} (F)	1.22×10^{-15}	5.76×10^{-16}	4.83×10^{-16}
GFP (Hz)	2.98×10^{12}	4.26×10^{12}	5.98×10^{12}
TFP (Hz/V)	7.15×10^{13}	5.74×10^{13}	4.55×10^{13}
GTFP (Hz/V)	1.05×10^{13}	3.37×10^{12}	5.77×10^{12}

4 Conclusion

In this paper, a doping-less dual material double gate (DL-DMDG) MOSFET using charge plasma has been investigated. Using 2D Atlas simulator analog/RF parameter of DL-DMDG MOSFET has been evaluated for different region 2 oxide materials. From the results, it has been observed that device shows better performance in terms of sub-threshold swing, I_{on}/I_{off} ratio, DIBL etc. for high oxide permittivity of region 2 oxide materials. Further, a peak value of analog/RF parameters has been studied for different region 2 oxide material. From the tabular results, it has been concluded that DL-DMDG MOSFET shows better performance for higher region 2 oxide permittivity. Finally, it has been concluded that DL-DMDG MOSFET shows better analog/RF performance for high oxide permittivity and provide better immunity to SCEs.

References

1. Balamurugan, N.B.: A novel scaling theory for effective conductive path effect of double gate (DG) MOSFETs for nano-scale CMOS circuit design. In: International Conference on Signal Processing, Communications and Networking, 2008. ICSCN'08. IEEE (2008)
2. Wang, X., Roy, S., Asenov, A.: High performance MOSFET scaling study from bulk 45 nm technology generation. In: 9th International Conference on Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. IEEE (2008)
3. Hiblot, G.: Impact of short-channel effects on velocity overshoot in MOSFET. In: 2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS). IEEE (2015)
4. Xie, Q., Jun, X., Taur, Y.: Review and critique of analytic models of MOSFET short-channel effects in subthreshold. IEEE Trans. Electron Devices **59**(6), 1569–1579 (2012)
5. Panigrahy, S., Sahu, P.K.: Performance enhancement and reduction of short channel effects of nano-MOSFET by using graded channel engineering. In: 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT), IEEE (2013)
6. Raad, B.R.: Drain work function engineered doping-less charge plasma TFET for ambipolar suppression and RF performance improvement: a proposal, design, and investigation. IEEE Trans. Electron Devices **63**(10), 3950–3957 (2016)

7. Mohankumar, N., Syamal, B., Sarkar, C.K.: Influence of channel and gate engineering on the analog and RF performance of DG MOSFETs. IEEE Trans. Electron Devices **57**(4), 820–826 (2010)
8. Sharma, R.K.: Analog/RF figures of merit of advanced DG MOSFETs. In: 2012 8th International Caribbean Conference on Devices, Circuits and Systems (ICCDCS). IEEE (2012)
9. Wang, H., Zang, C.: Impact of random dopant fluctuation effect on surrounding gate MOSFETs: from atomic level simulation to circuit performance evaluation. J. Nanosci. Nanotechnol. **11**(12), 10429–10432 (2011)
10. Kumar, M.J., Janardhanan, S.: Doping-less tunnel field effect transistor: design and investigation. IEEE Trans. Electron Devices **60**(10), 3285–3290 (2013)
11. Gupta, S.K., Baishya, S.: Effect of high-k gate materials on analog and RF performance of Double Metal Double Gate (DMDG) MOSFETs. In: 2013 Annual IEEE India Conference (INDICON). IEEE (2013)
12. Salmani-Jelodar, M.: Optimum high-k oxide for the best performance of ultra-scaled double-gate MOSFETs. IEEE Trans. Nanotechnol. **15**(6), 904–910 (2016)
13. Voinigescu, S.P.: 2D numerical investigation of the impact of compositional grading on the performance of submicrometer Si-SiGe MOSFET's. IEEE Trans. Electron Devices **42**(6), 1039–1046 (1995)
14. Loan, S.A., Rafat, M.: A high performance charge plasma based lateral bipolar transistor on selective buried oxide. Semicond. Sci. Technol. **29**(1), 015011 (2013)
15. Raad, B.R., Sharama, D.: Drain work function engineered doping-less charge plasma TFET for ambipolar suppression and RF performance improvement: a proposal, design, and investigation. IEEE Trans. Electron Devices **63**(10), 3950–3957 (2016)
16. ATLAS Device Simulator Software, Silvaco, Santa Clara, CA, USA (2015)
17. Kumar, P., Bhowmick, B.: A physics based threshold voltage model for hetero-dielectric dual material gate Schottky barrier MOSFET. Int. J. Numer. Modell. Electron. Netw. Devices Fields e2320 (2018)
18. Kumar, Prashanth, Bhowmick, Brinda: Suppression of ambipolar conduction and investigation of RF performance characteristics of gate-drain underlap SiGe Schottky barrier field effect transistor. IET Micro Nano Lett. **13**(5), 626–630 (2018)

Phosphorene: A Worthy Alternative of Graphene and MoS$_2$ in Surface Plasmon Resonance Sensor

J. B. Maurya (iD)**, Alka Verma and Y. K. Prajapati**

Abstract In this paper, Graphene, MoS$_2$, and Phosphorene-based surface plasmon resonance sensor are analysed numerically and theoretically. Transfer matrix method is used for the formulation of attenuated total reflection intensity. The surface plasmon resonance curves are obtained by using angular interrogation. The number of layers of Graphene, MoS$_2$, and Phosphorene are optimized with respect to performance defining parameters; shift in resonance angle, minimum reflection intensity, and beam width of surface plasmon resonance curve. It is found that senor will have best performance for 10 layers of Graphene, 4 layers of MoS$_2$, and 10 layers of Phosphorene. Further, at these optimized number of layers, Phosphorene has highest sensor performance in terms of Sensitivity, Resolution, and Quality factor. Hence, it is believed that Phosphorene can replace Graphene and MoS$_2$ in surface plasmon resonance sensor.

Keywords Surface plasmon resonance sensor · Graphene · MoS$_2$ · Phosphorene · Sensitivity

J. B. Maurya
Department of Electronics and Communication Engineering, Bundelkhand Institute of
Engineering and Technology, Jhansi 284128, Uttar Pradesh, India
e-mail: jitendra.maurya06@gmail.com

A. Verma
Departement of Electronics Engineering, Institute of Engineering and Rural Technology,
Allahabad 211002, Uttar Pradesh, India
e-mail: alkapra25@iert.ac.in

Y. K. Prajapati (✉)
Departement of Electronics and Communication Engineering,
Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, Uttar Pradesh,
India
e-mail: yogendrapra@mnnit.ac.in

1 Introduction

The properties like high sensitivity, real time monitoring, label free, fast, reliable, repeatable, reusable, and use of small amount of sample, etc., of surface plasmon resonance sensor (SPR), an optical sensor, make it best choice to sense the change in refractive index (RI), e.g., for the detection of gas, chemical, biochemical, biomolecule, etc. [1]. Because of the extraordinary detection property and performance, SPR sensor has vast application in the study of interaction between different biomolecules e.g., proteins, nucleic acids, peptides, receptors, antibodies, and lipids, in the investigation of viral binding for surface functionalization, in antiviral drug discovery tools, in food safety, in mines to sense the toxic and hazardous gas leakages, etc. [2].

The SPR sensor was first configured by H. Reather and E. Kretschmann in 1968 [3]. The Kretschmann configured SPR sensor has a thin surface plasmon (SP) active metal layer in the range of 40–50 nm which is deposited on the substrate. This configuration needs a light coupler, e.g., prism, waveguide, or grating to couple the light at the bottom surface of metal layer in order to excite SP at the top surface (interface of metal-dielectric) of metal surface. A. Otto, contemporaries of E. Kretschmann, had also configured a slightly different SPR sensor [4]. The Otto's configuration had a gap of approximately 100 nm between the substrate and SP metal. As per the practical aspect, it was very difficult to maintain such a thin air gap between substrate and a thin SP metal, which seems to be a key reason for the lack of popularity of Otto's configuration. The resonance condition, known as SPR, is achieved when the wave vector of incident light becomes equal to the wave vector of surface plasmon. This resonance condition can be achieved by four type of the interrogation; angle, wavelength, phase, and intensity. Out of these interrogations, the angle interrogation is mostly used because of its easy operation with the help of goniometer. In addition, angle interrogation requires single monochromatic laser source which makes the experimental setup much economic. Hence, in this paper angle interrogation is adopted.

Life sciences are being potentially facilitated by SPR sensor because of its capability to detect the presence of single biomolecule. The single-stranded Deoxyribonucleic acid (ssDNA) virus based human pathogens such as Human Boca virus (HBoV) which is responsible for lower respiratory tract infections [5, 6], and Parvoviridae B19 responsible for infectious disease in the pediatric population which is also known as fifth disease or slapped cheek syndrome [7, 8], must be detected in the early stage of infections before the severity of infections in order to protect the patient from death. The DNA hybridization detection capability of SPR sensor makes it suitable to detect these ssDNA-based human pathogens [9]. In DNA hybridization detection, the complimentary ssDNA is immobilized on the top surface of sensor, and then target ssDNA is allowed to make pair via hybridization with complementary ssDNA, which results in double stranded DNA (dsDNA), which eventually increases the RI near the top surface of SPR sensor. This local increase in RI can be detected by SPR sensor.

Bare SP metals, e.g., gold, silver, copper, aluminum have less binding ability with the complimentary ssDNA in order to get immobilize on their surface [10]. Also, these metals are prone to the oxidation and corrosion, which degrades the performance of SPR sensor [11]. Hence, to achieve the immobilization of ssDNA, the surface of these metals must be functionalize to have higher affinity toward ssDNA and to protect them against oxidation and corrosion. These requirements can be fulfilled by several layered nanomaterials; e.g., Graphene (invented in 2004) [12], MoS$_2$ (invented in 2010) [13], and black Phosphorene (invented in 2014) [14] because of their chemical inert nature and natural binding property with ssDNA. The Graphene can bind the ssDNA with the π-stacking bonding property with carbon rings of nucleobases present in ssDNA [15]. The MoS$_2$ can bind the ssDNA with the van der Waals force between nucleobases of ssDNA and the basal plane of MoS$_2$ [16]. Phosphorene can bind the ssDNA after the functionalization with cationic polymer poly-L lysine (PLL) [17]. Here, PLL will work as linker between Phosphorene and ssDNA.

As it is seen that Graphene, MoS$_2$, and Phosphorene can bind ssDNA on their surface, which is required to detect infectious human pathogen, the analysis of SPR sensor having SP metal covered with any of these nanomaterials is essential. Hence, in this paper effects of adding these nanomaterials on sensor performance are analyzed.

2 Theory

2.1 Sensor Structure and Design Parameter

The schematic diagram of nanomaterial-based SPR sensor structure is presented in Fig. 1. Here, the silver is considered as SP active metal layer because of its highest performance among gold, silver, copper, aluminum [18]. At a time, anyone of the Graphene, MoS$_2$, and Phosphorene can be placed on the silver layer. At the operating wavelength 633 nm, the refractive indices of the BK-7 prism, silver, Graphene, MoS$_2$, and Phosphorene are 1.5151 [18], $0.0803 + 4.2347i$ [18], $3 + 1.1487i$ [18], $5.9 + 0.8i$ [18], and $3.5 + 0.01i$ [19] respectively. The refractive index (RI) of the sensing medium (n_c) is considered to vary from 1.33 (pure water) to 1.335 (impure water after adsorption of biomolecules). The thickness of silver layer, Graphene, MoS$_2$, and Phosphorene are 50 nm, (G × 0.34) nm, (M × 0.65) nm, and (P × 0.56) nm respectively, where G, M, and P are the number of Graphene, MoS$_2$, and Phosphorene layers respectively.
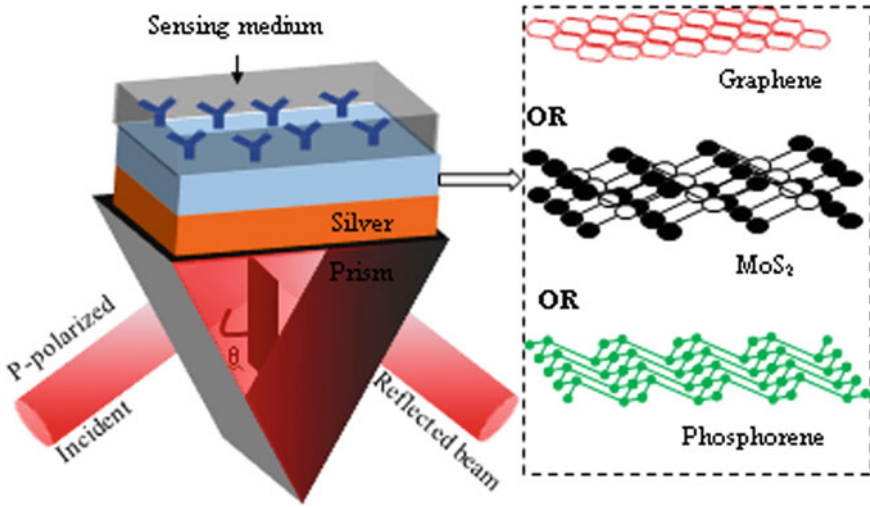
Fig. 1 Schematic of nanomaterial based surface plasmon resonance sensor

2.2 Mathematical Modeling of Reflectance Curve

The matrix method for N-layer model is applied for the calculation of reflection intensity of the reflected light [20]. This method is efficient and does not consider any approximation. The thicknesses of the layers, d_k, are considered along the z-axis. The dielectric constant and RI of the kth layer is considered as ε_k and n_k respectively. By applying the boundary condition, the tangential fields at $Z = Z_1 = 0$ are presented in terms of the tangential field at $Z = Z_{N-1}$ as follows;

$$\begin{bmatrix} U_1 \\ V_1 \end{bmatrix} = H \begin{bmatrix} U_{N-1} \\ V_{N-1} \end{bmatrix} \tag{1}$$

where U_1 and V_1 represent the tangential components of electric and magnetic fields respectively at the boundary of the first layer and U_{N-1} and V_{N-1} are the corresponding fields for the boundary at Nth layer. The H_{ij} presents the characteristics matrix of the combined structure of the sensor, and for p-polarized light it can be given as

$$H_{ij} = \left(\prod_{k=2}^{N-1} H_k \right)_{ij} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \tag{2}$$

with,

$$H_k = \begin{bmatrix} \cos \beta_k & (-i \sin \beta_k)/q_k \\ -iq_k \sin \beta_k & \cos \beta_k \end{bmatrix} \tag{3}$$

where

$$q_k = \left(\frac{\mu_k}{\varepsilon_k}\right)^{1/2} \cos\theta_k = \frac{\left(\varepsilon_k - n_1^2 \sin^2\theta_1\right)^{1/2}}{\varepsilon_k} \tag{4}$$

and

$$\beta_k = \frac{2\pi}{\lambda} n_k \cos\theta_k (z_k - z_{k-1}) = \frac{2\pi d_k}{\lambda}\left(\varepsilon_k - n_1^2 \sin^2\theta_1\right)^{1/2} \tag{5}$$

After some straightforward mathematical steps, one can obtain the reflection coefficient for p-polarized light which is given below;

$$r_p = \frac{(H_{11} + H_{12}q_N)q_1 - (H_{21} + H_{22}q_N)}{(H_{11} + H_{12}q_N)q_1 + (H_{21} + H_{22}q_N)} \tag{6}$$

The reflection intensity R_p of the defined multilayer configuration is given as;

$$R_p = |r_p|^2 \tag{7}$$

2.3 Principle of Operation

A p-polarized light wave of 633 nm wavelength, generated from He–Ne laser source, is focused on the lateral plane of the prism. This prism couples the light to the deposited silver thin film. With the effect of attenuated total reflection (ATR), the evanescent field of coupled light excites the plasmons at silver surface. These excited plasmons propagate along the interface of metal-dielectric at the resonance angle, and termed as surface plasmon wave (SPW). The wave vector of this SPW can be varied with the incidence angle (θ) of light at the prism. The resonance condition is achieved after incidence angle greater than critical angle $\{\theta c = \sin^{-1}(n_c/n_{prism})\}$ at which wave vector of SPW is exactly matched with the incident light. The incidence angle at which resonance condition is achieved known as resonance angle (θ_{res}), i.e., $\theta_{res} > \theta_c$. As the RI of sensing medium is changed, the resonance condition is disturbed, hence again one has to change the incidence angle to achieve the resonance condition. Thus, the change in RI of sensing medium due to adsorption of biomolecules can be measured by measuring the shift in resonance angle ($\Delta\theta_{SPR}$). The intensity of ATR light ($R = R_p$) corresponding to different incident angle can be measured on the opposite face of the prism by the photo-detector followed by Lock-in-amplifier, and the same is plotted as incidence angle versus reflection intensity. This plot is known as reflectance curve or SPR curve which has dip at resonance angle. The incidence angle corresponding to the minimum reflection intensity is known as the resonance angle (θ_{SPR}).

2.4   Performance Parameters

The main performance parameters of the SPR sensor are sensitivity, resolution (detection accuracy), and the quality factor. For a good sensor, higher value of all of these performance parameters is always desire.

Sensitivity (S) to the sensing medium RI change is defined as the ratio of shift in resonance angle of incidence ($\Delta\theta_{SPR}$) to the RI change in the sensing medium (Δn_c);

$$S = \frac{\Delta\theta_{SPR}}{\Delta n_c} \tag{8}$$

The beam width is determined by the method adopted by J. B. Maurya and Y. K. Prajapati and the determined beam width is termed as Jitednra-Yogendra beam width (JYBW) [21]. The resolution is defined as the reciprocal of the beam width of the reflectance curve;

$$Resolution = \frac{1}{JYBW} \tag{9}$$

The quality factor (Q) is defined as the ratio of sensitivity to the beam width of the reflectance curve;

$$Q = \frac{S}{JYBW} \tag{10}$$

3   Result and Discussion

In Fig. 2, the SPR curves are plotted for monolayer of Graphene, MoS_2, and Phosphorene at sensing medium refractive index 1.33 and 1.335. Since, the adsorption of biomolecules accounts very less change in refractive index of sensing medium, Δn_c is considered very less. As the n_c is changed from 1.33 to 1.335, right shift of resonance angle for all of the nanomaterial can be easily observed from Fig. 1, which signifies the adsorption of biomolecules on the surface of nanomaterials. Since the Δn_c is constant for all the nanomaterials, as per the (8) amount of right shift, i.e., $\Delta\theta$ will define the sensitivity order of different nanomaterial. However, the sensitivity will also depend on the physisorption property of different nanomaterials with respect to particular biomolecule, which is out of scope for numerical analysis. The $\Delta\theta$ for Graphene, MoS_2, and Phosphorene are; 0.63, 0.63, and 0.57, respectively. Hence, from (8) the sensitivity order for monolayer; Graphene = MoS_2 > Phosphorene. But, as per the practical point of view, the synthesis of large uniform monolayer of Graphene, MoS_2, and Phosphorene are very tough and not always possible. Hence, the variation of $\Delta\theta_{res}$, minimum reflection intensity (Rmin), and JYBW along with
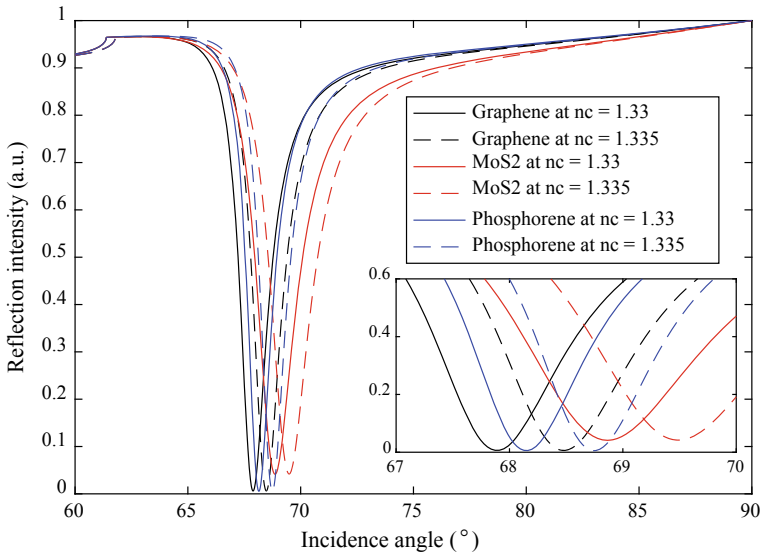
Fig. 2 Reflectance curve for monolayer of Graphene, MoS$_2$, and Phosphorene at sensing medium refractive index 1.33 and 1.335 indicating refractive index sensing

number of layers of nanomaterial are essential, which are accomplished in the following paragraphs.

In Fig. 3, $\Delta\theta$, Rmin, and JYBW are varied corresponding to the number of layers of nanomaterials. Here, the $\Delta\theta$ is calculated corresponding to $\Delta n_c = 0.005$, whereas Rmin and JYBW are calculated at $n_c = 1.335$ (after the adsorption of biomolecule). It is seen from this figure that the $\Delta\theta$ for Graphene and Phosphorene increases
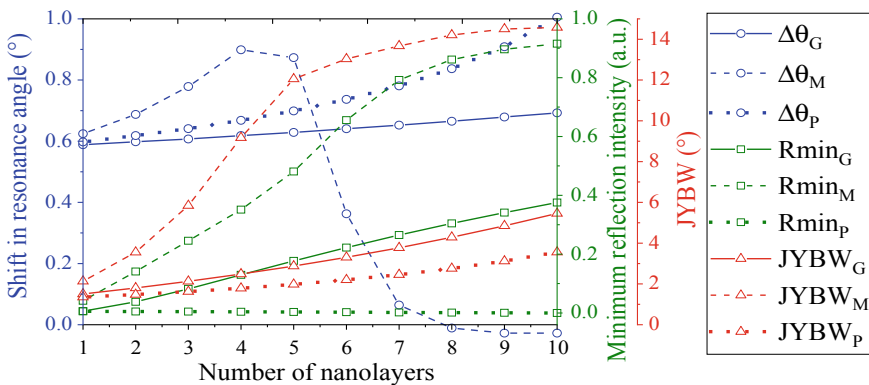
Fig. 3 Optimization of number of layers of Graphene (G), MoS$_2$ (M), and Phosphorene (P) with respect to performance defining parameters; shift in resonance angle, minimum reflectance, and beam width of reflectance curve

monotonically, for the considered range, with the number of nanolayers, whereas for MoS$_2$, the $\Delta\theta$ increases up to four layers and start to decrease rapidly beyond four layers. This can be justified by observing the Rmin of nanomaterials.

The Rmin and JYBW are the function of extinction coefficient (imaginary part of the RI) and thickness of the material [22]. The wave propagation is slowed down in the lossy medium (higher extinction coefficient), and this speed of propagation further decreases as the thickness of lossy medium increase. As a result of this, the surface plasmon damps in the lossy medium which increases with the extinction coefficient and thickness of material. The SPR curve becomes shallower (increasing Rmin) and broader (increasing JYBW), as the surface plasmon damping increases. Hence, the Rmin and JYBW increase with the number of nanolayers (increasing thickness).

The Rmin and JYBW of all nanomaterials are increasing monotonically because of the increasing damping with number of nanolayer. At any particular nanolayer, extinction coefficient has the order; Graphene (1.1487) > MoS$_2$ (0.8) > Phosphorene (0.01), and thickness of monolayer has the order; MoS$_2$ (0.65 nm) > Phosphorene (0.56 nm) > Graphene (0.34 nm). Hence, the product of extinction coefficient and thickness of monolayer nanomaterial follow the order; MoS$_2$ (0.52) > Graphene (0.39) > Phosphorene (0.0056). It seems that Rmin and JYBW follows the order of product of extinction coefficient and thickness of monolayer nanomaterial, i.e., MoS$_2$ > Graphene > Phosphorene. It is just a coincidence nothing else, because if two number less than zero are multiplied the result is less than the individual numbers, but it is not the case for damping in material. Although the extinction coefficients or thickness of any material is less than zero, increasing of either of these quantity increase the damping.

The Rmin of Phosphorene is approximately equal to zero which is completely considerable throughout the range of considered number of nanolayers. Whereas, the maximum Rmin of Graphene is less than 0.4 (at G = 10) which can be considered for the identification of SPR dip. But, the Rmin for MoS$_2$ is increasing at a rapid rate with the number of MoS$_2$ layers, and it reaches approximately at 0.6 for M = 4, at the worst case it can be considered for the identification of SPR dip. But, the Rmin beyond 0.6, for MoS$_2$ (M > 4), signifies the improper SPR dip which is not possible to identify, and hence improper $\Delta\theta$ may be observed.

The JYBW of Graphene and Phosphorene are considerable throughout the considered range, whereas for the MoS$_2$ it can only be considered up to four layers (M = 4), beyond which the JYBW will be very high, and as per (9) and (10), significant decrease in the resolution and quality factor of the sensor can be observed. Hence, on the basis above discussion regarding Fig. 3, it can be said that for having high sensitivity and considerable resolution and quality factor at sufficient SPR dip (considerable Rmin), the optimized number of layers of Graphene (G), MoS$_2$ (M), and Phosphorene (P) are; G = 10, M = 4, and P = 10 respectively.

In Fig. 4, a bar graph is plotted to compare the sensor performance parameter, i.e., sensitivity, resolution, and quality factor for the optimized number of layers of all of the three considered nanomaterials, i.e. Graphene, MoS$_2$, and Phosphorene. It is
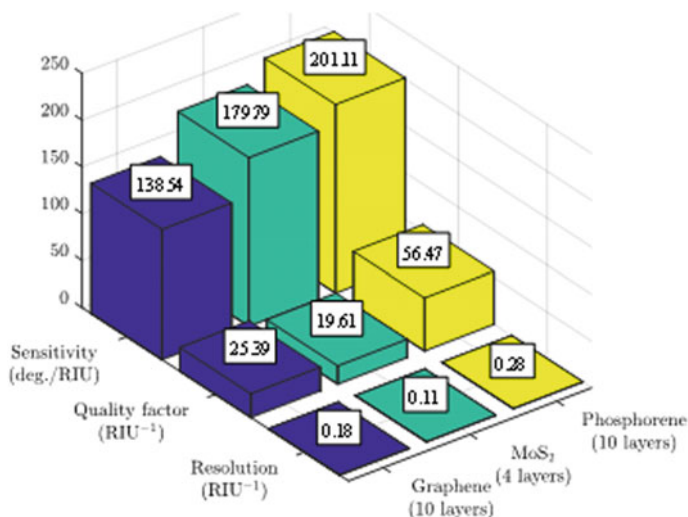
Fig. 4 Surface plasmon resonance sensor performance in terms of sensitivity, quality factor, and resolution for the optimized number of layers of Graphene, MoS$_2$, and Phosphorene

found that Phosphorene has highest all of the three performance parameters, which is always desired for a good sensor.

4 Conclusion

Three layered nanomaterials namely; Graphene, MoS$_2$, and Phosphorene are analyzed theoretically and numerically. It is found that the optimized number of layers for Graphene, MoS$_2$, and Phosphorene are 10, 4, and 10, respectively. And, at these optimized number of layers, the Phosphorene has highest sensor performance in terms of sensitivity, resolution, and quality factor. Hence, Phosphorene can replace the Graphene and MoS$_2$ for the SPR sensing provided that its physisorption property must be suitable for the antibody immobilized on it.

References

1. Homola, J.: Surface plasmon resonance sensors for detection of chemical and biological species. Chem. Rev. **108**(2), 462–493 (2008)
2. Shpacovitch, V.: Application of surface plasmon resonance (SPR) for the detection of single viruses and single biological nano-objects. J. Bacteriol. Parasitol. **3**(7), e110 (2012)
3. Kretschmann, E., Raether, H.: Radiative decay of non radiative surface plasmons excited by light. Zeitschrift für Naturforschung A **23**(12), 2135–2136 (1968)

4. Otto, A.: Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. Zeitschrift für Physik A Hadrons and nuclei **216**(4), 398–410 (1968)

5. Ricour, C., Goubau, P.: Human bocavirus, a newly discovered parvovirus of the respiratory tract. Acta Clin. Belgica **63**(5), 329–334 (2008)

6. Allander, T., Jartti, T., et al.: Human bocavirus and acute wheezing in children. Clin. Infect. Dis. **44**(7), 904–910 (2007)

7. Lunardi, C., Tinazzi, E., Bason, C., Dolcino, M., Corrocher, R., Puccetti, A.: Human parvovirus B19 infection and autoimmunity. Autoimmun. Rev. **8**(2), 116–120 (2008)

8. Wawina, T.B., Tshiani, O.M., et al.: Detection of human parvovirus B19 in serum samples from children under 5 years of age with rash–fever illnesses in the Democratic Republic of the Congo. Int. J. Infect. Dis. **65**, 4–7 (2017)

9. Endo, T., Kerman, K., Nagatani, N., Takamura, Y., Tamiya, E.: Label-free detection of peptide nucleic acid–dna hybridization using localized surface plasmon resonance based optical biosensor. Anal. Chem. **77**(21), 6976–6984 (2005)

10. Wu, L., Chu, H.S., Koh, W.S., Li, E.P.: Highly sensitive graphene biosensors based on surface plasmon resonance. Opt. Express **18**(14), 14395–14400 (2010)

11. Kravets, V.G., Jalil, R., Kim, Y.J., Ansell, D., Aznakayeva, D.E., Thackray, B., et al.: Graphene-protected copper and silver plasmonics. Sci. Rep. **4**, 5517 (2014)

12. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D.A., Zhang, Y., et al.: Electric field effect in atomically thin carbon films. Science **306**(5696), 666–669 (2004)

13. Mak, K.F., Lee, C., Hone, J., Shan, J., Heinz, T.F.: Atomically thin MoS_2: a new direct-gap semiconductor. Phys. Rev. Lett. **105**(13), 136805 (2010)

14. Liu, H., Neal, A.T., Zhu, Z., Luo, Z., et al.: Phosphorene: an unexplored 2D semiconductor with a high hole mobility. ACS Nano **8**(4), 4033–4041 (2014)

15. McGaughey, G.B., Gagné, M., Rappé, A.K.: π-Stacking interactions alive and well in proteins. J. Biol. Chem. **273**(25), 15458–15463 (1998)

16. Zhu, C., Zeng, Z., Li, H., Li, F., Fan, C., Zhang, H.: Single-layer MoS_2-based nanoprobes for homogeneous detection of biomolecules. J. Am. Chem. Soc. **135**(16), 5998–6001 (2013)

17. Kumar, V., Brent, J.R., Shorie, M., Kaur, H., Chadha, G., Thomas, et al.: Nanostructured aptamer-functionalized black phosphorus sensing platform for label-free detection of myoglobin, a cardiovascular disease biomarker. ACS Appl. Mater. Interfaces **8**(35), 22860–22868 (2016)

18. Maurya, J.B., Prajapati, Y.K.: A comparative study of different metal and prism in the surface plasmon resonance biosensor having MoS_2-graphene. Opt. Quant. Electron. **48**(5), 280–291 (2016)

19. Wu, L., Guo, J., Wang, Q., Lu, S., Dai, X., Xiang, Y., Fan, D.: Sensitivity enhancement by using few-layer black phosphorus-graphene/TMDCs heterostructure in surface plasmon resonance biochemical sensor. Sens. Actuators B: Chem. **249**, 542–548 (2017)

20. Maurya, J.B., Prajapati, Y.K., Singh, V., Saini, J.P., Tripathi, R.: Improved performance of the surface plasmon resonance biosensor based on graphene or MoS_2 using silicon. Opt. Commun. **359**, 426–434 (2016)

21. Maurya, J.B., Prajapati, Y.K.: A novel method to calculate beam width of SPR reflectance curve: a comparative analysis. IEEE Sens. Lett. **1**, 1–4 (2017)

22. Pockrand, I.: Surface plasma oscillations at silver surfaces with thin transparent and absorbing coatings. Surf. Sci. **72**, 577–588 (1978)

Substrate Integrated Waveguide Wideband and Ultra-Wideband Bandpass Filters Using Multimode Resonator

Naveen Singh, Sandeep Kumar, Sanjay Gupta and R. K. Chauhan

Abstract In this paper, a novel design technique using substrate integrated waveguide (SIW) based on multimode resonator is presented for the implementation of the dual-band bandpass, wideband bandpass and ultra-wideband bandpass filters. Filter design is started by etching four W-shaped slots on the upper conductor layer of SIW cavity in order to produce few resonant peaks in passband. W-shaped slots forming up quintuple mode resonator create wide passband while maintaining overall size of the filter to be much less. Implemented wide bandpass filter achieves an insertion loss lower than 0.8 dB, return loss better than 12 dB in the passband and 3 dB fractional bandwidth of 44%. The structure is further modified into two H-shaped slots operating as sextuple mode resonator. This design produces an ultra-wideband bandpass filter with an insertion loss smaller than 1.1 dB, return loss better than 11 dB in the passband and 3 dB fractional bandwidth of 55%.

Keywords Substrate integrated waveguide (SIW) · Multimode resonator (MMR) · Ultra-wideband (UWB)

1 Introduction

Recently, the demand of filters in the field of wireless radio frequency (RF) communication systems has been growing at a very fast pace. This puts challenges before the designers to design and develop receiver compatible filters with reduced size.

N. Singh (✉) · S. Kumar · S. Gupta · R. K. Chauhan
Department of Electronics and Communication Engineering,
Madan Mohan Malaviya University of Technology, Gorakhpur, UP, India
e-mail: naveensg1995@gmail.com

S. Kumar
e-mail: sndpkmr925@gmail.com

S. Gupta
e-mail: sanjay095gupta@gmail.com

R. K. Chauhan
e-mail: rkcece@mmmut.ac.in

Ultra-wideband bandpass filters are among such filters that are preferred for supporting high data rate while providing several narrow transmission bands. The proposed filters have been studied for wideband of frequency ranging from 6 to 14 GHz. This ultra-wide frequency spectrum has numerous high speed wireless applications such as broadcasting satellite, radiolocation, etc. Interestingly, SIW technology has been successfully employed to design such filters with high quality factor, low loss, high power capability and low cost.

Substrate integrated waveguide (SIW) is composed of two rows of periodic metallic via holes embedded in a dielectric substrate that connect two parallel metal plates [1, 2]. It has many similarities with conventional rectangular waveguides [2]. It preserves advantages of conventional rectangular waveguides, namely, high Quality factor. Losses of SIW components are lower than corresponding microstrip devices. SIWs have attracted much attention in the recent years, because it allows the design of several microwave components with easy fabrication with planar circuits, including filters, directional couplers, circulators, oscillators, mixers, power amplifiers, slot array, and leaky antennas.

In order to make overall size of SIW bandpass filter to be much compact with a wide passband, a multimode resonator (MMR) is the best possible element to be considered in the design of filters. If each multimode resonator has N resonant modes within the desired passband, the proposed filter is considered to be Nth order filter with very much compact size. Multimode resonators can be classified on the basis of number of modes such as dual-mode resonators, triple-mode resonators, quadruple-mode resonators, quintuple-mode resonators, and so on [3].

The following sections construct W and H-shaped slots etched on the upper metallic layer of SIW in order to produce a number of resonant peaks in this way implementing a dual-band bandpass filter which is further modified into a wideband and ultra-wideband SIW bandpass filters with very much reduced size, respectively. In Sect. 2, design and analysis of dual-band bandpass filter is introduced by etching four W-shaped slots. This design is further examined with variation of a few parameters, then a wideband bandpass filter is obtained with return loss higher than 12 and 3-dB fractional bandwidth (FBW) of 44.2%. This section extends its design procedure by etching two H-shaped slots on the upper metallic plate of SIW cavity which produces an ultra-wideband filter with 3 dB fractional bandwidth of 55% and return loss better than 11 dB in Sect. 3, Results are discussed after extensive parametric studies and numerical optimization in a high-frequency electromagnetic simulator (HFSS). This simulation tool is based on algorithm implementing finite element method.

2   Design and Analysis of Proposed Filters

In general, three design techniques are used to design a bandpass filter using multimode resonator. Coupling resonators topology [4] is employed as the first design technique. Dual-mode resonator filters are generally preferred to be designed using this technique. Second design technique deals with distributing the fundamental

resonant frequencies of the resonator according to the insertion loss function of the chebyshev pole in passband [5]. The third design method, typically called 3 dB matching technique, is more common for higher mode resonator filters design, namely the quadruple-mode resonators filter [6], the quintuple-mode resonators filter. Technique used in [5] is utilized to implement the proposed filters in this paper.

The design procedure can be started by first choosing the desired center resonant frequency of TE_{m0k} mode given by (1). The initial sizes of this cavity can be obtained when m as 1 and k are already decided. It is noteworthy to say that the number of modes of resonator filter is determined by k. Other resonance frequencies in the passband can be evaluated using (3).

$$f_{\mathrm{TE}_{m0k}} = \frac{c_0}{2\sqrt{\varepsilon_r}}\sqrt{\left(\frac{m}{w_{\mathit{eff}}}\right)^2 + \left(\frac{k}{l_{\mathit{eff}}}\right)^2} \tag{1}$$

In (1), c_0 is speed of light in vacuum, ε_r is relative permittivity of substrate, w_{eff} is the equivalent width of SIW and l_{eff} is the equivalent length. w_{eff} And l_{eff} can be determined using the formulas given in [1]:

$$w_{\mathit{eff}} = B - 1.08\frac{d^2}{p} + \frac{d^2}{L}$$

$$l_{\mathit{eff}} = L - 1.08\frac{d^2}{p} + \frac{d^2}{L} \tag{2}$$

$$f_j = f_c\left(1 + x_j \times \frac{FBW}{2}\right) \tag{3}$$

$$x_j = \cos\left(\frac{2k + 1 - 2j}{2k}\pi\right) \quad j = 1\,to\,k \tag{4}$$

The constituted slots behaving as multimode resonator is created to produce few resonant peaks in the passband and a transmission zero in the upper cutoff frequency, thus improving the upper stopband performance. The resonant frequencies exist between lower cutoff frequency and transmission zero frequency in this way creating a wide passband. It is worth to note that different types of shapes for the slot behaving as multimode resonator can be studied from [3]. From [3], four W-shaped slots are initially considered in our design which is further modified into two H-shaped slot for getting ultra-wideband response.

2.1 Design of Dual-Band Bandpass Filter

A single-layer Rogers RT/duroid 5880 (tm) substrate is utilized to fabricate the proposed dual-band bandpass filter with a thickness of 0.8 mm using linear arrays of

metallized via-holes with a diameter of 1 mm and a gap of 1.75 mm between them. Design of dual-band bandpass filter can be initiated by first fixing the dimensions of each W-shaped slot at D, t, w, u at 4.50 mm, 0.75 mm, 0.20 mm, 0.325 mm respectively in order to make simple and effective way of design. The length h is allowed to vary from 3.8 to 4.6 mm and length m is also allowed to vary from 1 to 2.5 mm. The optimized value for the h and m was obtained to be 4.2 mm and 1.5 mm respectively. The desired slots will have W-shaped that will constitute resonant peaks in both passbands as shown in Fig. 1a.

The simulated return loss for the filter in lower and upper passband is better than 46 dB and 16 dB respectively. The insertion losses which is mainly attributed to conductor and dielectric losses is found to be lower than 1.2 dB. Insertion loss is improved by choosing diameter of metallic cylinders half to the center-to-center distance between them [1, 2] (Table 1).
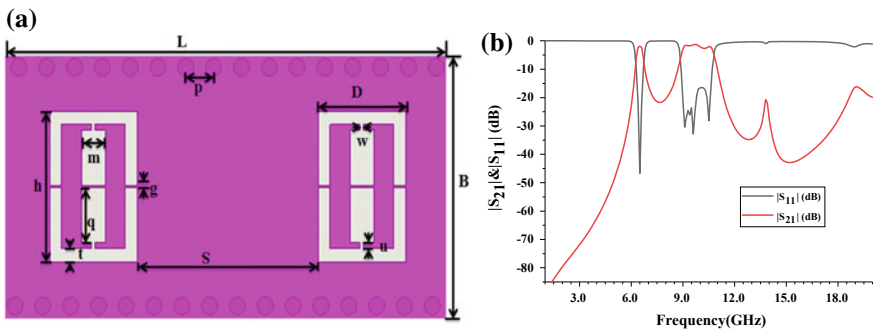
Fig. 1 **a** Schematic of the SIW based dual-band bandpass filter with four W-shaped slots. **b** Simulated frequency responses of ultra-wide bandpass filter

Table 1 Dimensions of dual-band bandpass filter

Symbols	Value (mm)	Symbols	Value (mm)
L	31.75	B	15.00
D	5.50	S	11.40
m	1.50	t	0.75
r	0.25	g	0.10
u	0.30	w	0.20
p	1.75	h	12.70
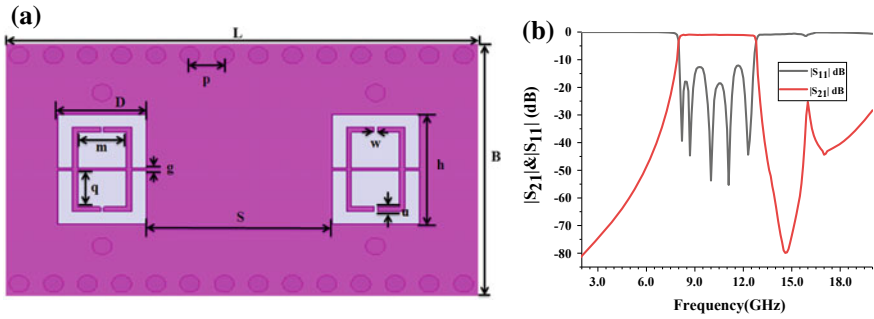
Fig. 2 a Schematic of the SIW based wideband bandpass filter with four W-shaped slots. **b** Simulated frequency responses of ultra-wide bandpass filter

Table 2 Dimensions of wide bandpass filter

Symbols	Value (mm)	Symbols	Value (mm)
L	28	B	15
D	4.50	S	9.50
m	2.45	t	0.75
r	0.50	g	0.10
u	0.30	w	0.20
p	1.75	h	5.76

2.2 Design of Wideband Bandpass Filter Using Quintuple Mode Resonator

With a slight variation in dimensions of h and m in W-shaped slots of SIW dual-bandpass filter, a wideband bandpass filter can be obtained as shown in Fig. 2a. More specifically, h should be kept below 2,6 mm and m at 2.45 mm. This proposed filter has a sharp transition at upper cutoff frequency with a fractional bandwidth of 44.2%.

We observe from the return loss frequency response that the filter has five resonant peaks in the passband due to constituted quintuple mode resonator. These resonant frequencies constitute a wide passband with an insertion loss of 0.8 dB and return loss higher than 12 dB (Table 2).

2.3 Design of Ultra-Wideband Bandpass Filter Using Sextuple Mode Resonator

Ultra-wide band bandpass filter can be implemented on W-slotted SIW cavity by extending the slot to H-shaped as depicted in Fig. 3a. In this structure, six resonant modes exist between lower cutoff frequency and transmission zero frequency, thus

(a)

(b)
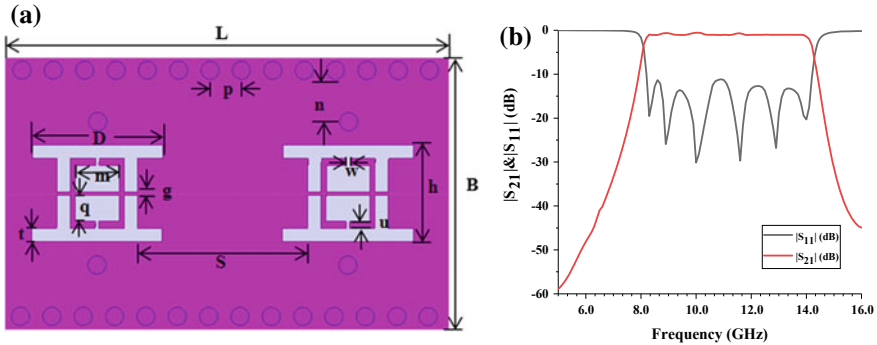
Fig. 3 a Schematic of the SIW based ultra-wideband bandpass filter with two H-shaped slots. **b** Simulated frequency responses of ultra-wide bandpass filter

Table 3 Dimensions of ultra-wide bandpass filter

Symbols	Value (mm)	Symbols	Value (mm)
L	28.00	B	15.00
D	7.30	S	9.50
m	2.45	t	0.75
r	0.50	g	0.10
u	0.30	w	0.20
p	1.75	h	3.20

creating a wide sixth order bandpass filter due to presence of sextuple mode resonator. These resonant frequencies can be approximately evaluated by (3) if we have an evaluated center frequency by taking average of cutoff frequency and transmission zero frequency.

With varying h and m and keeping rest of the dimensions fixed, we get desired UWB filter with a fractional bandwidth of 55%. Figure 3b shows the simulated frequency responses of the proposed bandpass filter. The simulated passband insertion loss achieves better than to be 1.1 dB with return loss higher than 11 dB (Table 3).

3 Results and Discussion

All the proposed bandpass filters are fabricated on a Rogers RT/duroid 5880 (tm) substrate with a dielectric loss tangent of 0.0009 and thickness of 0.8 mm. Filter design is initiated with dual-band bandpass filter by creating four W-shaped slots on upper metallic layer of SIW cavity. Its return loss in lower and upper passband is better than 46 dB and 16 dB, respectively. Its Insertion loss is achieved to be better than 1.02 dB in lower passband and 1.2 dB in the upper passband. With a slight variation of dimensions of slots, a wide bandpass filter is realized with slots behaving

Table 4 Comparison with other works

Ref.	FBW (%)	IL (dB)	RL (dB)	Size $\lambda_g * \lambda_g$
[7]	61	<1.54	>10.0	2.5 * 0.7
[8]	77	<2.00	>10.2	1 * 0.7
[9]	114	<2.5	>9.5	3.3 * 0.4
[10]	5	<3.0	>10	0.19 * 0.21
[11]	13.9	<2.5	>9.6	0.107 * 0.11
[12]	15.4	<1.2	>10.4	0.19 * 0.25
[13]	117	<3	>11	0.514 * 0.312
This work (wideband bandpass filter)	44.2	<0.8	>12	1.03 * 0.55
This work (ultra-wideband bandpass filter)	55	<1.1	>11	1.03 * 0.55

as quintuple mode resonator with a 3 dB fractional bandwidth of 44%. Its return loss is greater than 12 dB and insertion loss is better than 0.8 dB in the passband. This filter is further designed as ultra-wideband bandpass filter by modifying the slots into H-shaped producing six resonant peaks in passbands. Its return loss is greater than 11 dB and insertion loss is better than 1.1 dB in the passband. 3 dB ultra-wide passband typically ranges from 8.1 to 14.1 GHz with a fractional bandwidth of 55%. Results are compared with other ultra-wideband filter works in Table 4. Although proposed wideband filters have narrow fractional bandwidth, they are superior than others in terms of small insertion loss, better return loss with reduced size and easy fabrication with planar circuits.

4   Conclusion

Dual band, wideband and ultra-wideband bandpass filters have been presented in this paper using multimode resonator. Size of the wideband and ultra-wideband bandpass filters has been greatly reduced by etching W-shaped and H-shaped slots on the top metallic layer of SIW cavity respectively. Slots behaves as multimode resonator producing resonant peaks in the passband and a transmission zero at the upper cutoff frequency. This study clearly manifests a simple and effective method to design ultra-wideband bandpass filter with insertion loss better than 1.1 dB and return loss higher than 11 dB. FBW is 55%. Such a compact UWB bandpass filters can be effectively utilized in modern wireless communication systems with a high reception data rate.

References

1. Xu, F., Wu, K.: Guided-wave and leakage characteristics of substrate integrated waveguide. IEEE Trans. Microw. Theory Technol. **53**(1), 66–73 (2005)
2. Deslandes, D., Wu, K.: Accurate modeling, wave mechanisms, and design considerations of a substrate integrated waveguide. IEEE Trans. Microw. Theory Technol. **54**(6), 2516–2526 (2005)
3. Lee, K.C., Su, H.T., Haldar, M.K.: A review of centrally loaded multimode microstrip resonators for bandpass filter design. AEU-Int. J. Electron. Commun. **69**(10), 1533–1540 (2015)
4. Liao, C.-K., Chi, P.-L., Chang, C.-Y.: Microstrip realization of generalized Chebyshev filters with box-like coupling schemes. IEEE Trans. Microw. Theory Technol. **55**(1), 147–153 (2007)
5. Chiou, Y.-C., Kuo, J.-T., Cheng, E.: Broadband quasi-Chebyshev bandpass filters with multimode stepped-impedance resonators (SIRs). IEEE Trans Microw. Theory Technol. **54**(8), 3352–3358 (2006)
6. Huang, X., Feng, Q., Xiang, Q.: Bandpass filter with tunable bandwidth using quadruple-mode stub-loaded resonator. IEEE Microw. Wirel. Compon. Lett. **22**(4), 176–178 (2012)
7. Hong, W., Hao, Z.-C., Chen, X.-P., Chen, J.-X., Wu, K.: Compact super-wide bandpass substrate integrated waveguide filters. IEEE Trans. Microw. Theory Technol. **53**(9), 2968–2977 (2005)
8. Geng, L., Che, W.-Q., Deng, K.: Wideband bandpass filter of folded substrate integrated with stripline compact resonant cell. Microw. Opt. Technol. Lett. **50**(2), 390–393 (2008)
9. Wu, L.-S., Zhou, X.-L., Yin, W.-Y.: Ultra-wideband bandpass filter using half-mode T-septum substrate integrated waveguide with electromagnetic bandgap structures. Microw. Opt. Technol. Lett. **51**(7), 1751–1755 (2009)
10. Tang, W., Hong, J.-S., Chun, Y.-H.: Compact tunable microstrip bandpass filters with asymmetrical frequency response. In: Proceedings of the 38th European Microwave Conference, pp. 599–602(2008)
11. Deng, H.-W., Zhao, Y.-J., Zhang, L., Zhang, X.-S., Zhao, W.: Compact triple-mode stub-loaded stepped impedance resonator and bandpass filter. Microw. Opt. Technol. Lett. **53**(4), 701–703 (2011)
12. Lee, K.C., Su, H.T., Haldar, M.K.: A novel compact triple-mode resonator for microstrip bandpass filter design. In: 2010 Asia-Pacific Microwave Conference Proceedings, pp. 1871–1874, Tokyo (2010)
13. Chu, Q.-X., Wu, X.-H., Tian, X.-K.: Novel UWB bandpass filter using stubloaded multilple-mode resonator. IEEE Microw. Wirel. Compon. Lett. **21**(8), 403–405 (2011)
14. Bin, W.G., Jian, C.Y.: Broadband printed-circuit-board characterization using multimode substrate integrated waveguide resonator. IEEE Trans. Microw. Theory Technol. **65**(6), 2145–2152 (2017)
15. Pozar, D.M.: Microwave Engineering, 4th edn. Wiley

Impact of Dimensional Effects on Subsurface Leakage Current of a Low-V_{TH} Nanoscale MOSFET Under Accumulation Bias

Vadthiya Narendar and Ashutosh Kumar Pandey

Abstract In portable battery-operated electronics, power dissipation has become a critical parameter. The presence of various types of leakage current components in nanoscale MOSFET influences the leakage power dissipation. There is another type of leakage current present in nanoscale MOSFET which is due to V_{DS}, the reduction of barrier height between the source and drain at distance away from the semiconductor-oxide interface is known as subsurface leakage current. The subsurface leakage current is clearly noticeable in low-threshold voltage (V_{TH}) transistors. This paper investigates the reliability issues with dimensional parameter such as gate length (L_G), source/drain junction depth (X_J), and drain-to-source voltage (V_{DS}) on subsurface leakage current of a nanoscale MOSFET. It is observed that the subsurface leakage current is exponentially dependent on L_G. Subsurface leakage current increases with X_J. Dimensional parameter dependent subsurface leakage current is extracted using ATLAS TCAD simulator.

Keywords Leakage current · Nanoscale device · Subsurface leakage current · Zero-V_{TH} MOSFET

1 Introduction

The major performance metrics of digital integrated circuits (ICs) are power dissipation, delay (speed), and area. Power dissipation is predominating parameter in battery operated electronic components. In complementary metal oxide semiconductor (CMOS) circuits, the prime source for power dissipation in nanometer regimes is due to the presence of various types of leakage current in MOSFETs. CMOS devices are scaled down by following Moore's law for every technological node to attain

V. Narendar (✉)
National Institute of Technology Warangal, Warangal 506004, Telangana, India
e-mail: narendarv@nitw.ac.in

A. K. Pandey
Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, UP, India
e-mail: chandan270697@gmail.com

30% decrease in delay time of a transistor which results in higher packing density and higher performance along with reduced threshold voltage, channel length and gate oxide thickness [1–9]. The amount of current flow in a transistor when the gate voltage is zero is known as off-state (I_{off}) leakage current and it is influenced by other leakage components such as subthreshold leakage, gate induced drain leakage and punchthrough current. Moreover, I_{off} also depends on device physical dimensions, channel doping, supply voltage, gate oxide thickness [5, 6]. Aforementioned various types of leakage currents in bulk MOSFET are modeled and available in the BSIM6 model. In addition to that, there is a leakage current path between the source and the drain underneath the channel region is called as subsurface leakage current. From the literature, it has not received much attention. The subsurface leakage current is clearly noticeable in zero-threshold voltage transistors [10].

The continuous increase in demand of low-power electronics pushes the conventional MOS devices into an ultra-low-threshold voltage (V_{TH}) MOS devices. A new breakthrough zero- threshold voltage device enables the electronic industry to introduce a very low supply voltage (0.1 V) with high precession circuits for the first time [11]. They enable countless applications starts from ultra-sensitive remote sensors to implantable medical devices. They also facilitate to design relay sensors and amplifier circuits which demands long term power back-up for uninterrupted operation. In addition to that, they also allow the electronic industry to design the circuits which takes 100 times less operating power than aggressive MOSFETs [12, 13]. Moreover, they offer to design sensitive and critical security applications in extremely remote locations, even in implantable medical devices for long-lasting operational life to avoid invasive surgery. They also useful in wide verity of analog circuits such as computers and its peripherals, hand held instrumentation, industrial control, security and alarm systems and sensor and detection systems. The other special analog circuits such as level translators, multiplexers, and buffers circuits are also designed with them. The applications of ultra-low-V_{TH} devices are diverse into Analog/RF, communication, and biomedical applications [14–17]. In addition to that, other verity of circuits such as: input/output, current reference, bandgap reference, high performance, input buffers and voltage regulators, etc. [18–22]. This paper investigates the impact of dimensional effect such as gate length (L_G), substrate doping (N_{BULK}), gate oxide thickness (T_{ox}), gate-to-source voltage (V_{GS}), drain-to-source voltage (V_{DS}), and temperature on subsurface leakage current of a nanoscale zero-V_{TH} MOSFET. This paper is organized as follows. Section 2 describes the device structure and simulation setup. The impact of device parameters on subsurface leakage current is explained in Sect. 3. Section 4 includes results and discussion in which the numerical simulation data obtained from ATLAS device simulator is compared with BSIM6 [23, 24]. Finally, Sect. 4 concludes the paper.

Subsequent paragraphs, however, are indented.

2   Device Structure and Simulation Setup

Figure 1 shows cross-sectional view of nMOSFET structure used in ATLAS device simulator with parameters of gate length (L_G), gate oxide thickness (T_{ox}), bulk substrate doping (N_{BULK}), and source/drain junction depth (X_J). The source and drain regions are n+-doped (Gaussian) with 1×10^{20} cm^{-3} and extension regions of source and drain are doped with 1×10^{19} cm^{-3} to suppress the GIDL current [25]. The numerical simulations of ATLAS include Shockley–Read–Hall and band-to-band tunneling models. Concentration based mobility model (CONMOB) along with field dependent mobility model (FLDMOB) has been incorporated for carrier mobility. For carrier transport drift-diffusion model is included. To measure the shift in energy band gap at semiconductor-oxide interface and its corresponding transverse electric field quantization model is included. Moreover, Auger model is also included. The device simulations are carried out at 300 K unless stated otherwise. The lightly doped zero-threshold voltage devices are promising candidates to show the subsurface leakage current. It has been used in various applications such as analog and RF, electrostatic discharge (ESD) and I/O.

The drain current variation with respect to gate to source voltage characteristics of nMOSFET at W = 1 µm, T_{ox} = 2 nm, V_{GS} = −3 V and V_{DS} = 0.2 V for different gate length (L_G) is depicted in Fig. 2. It can be noticed from Fig. 2 that, in strong accumulation regime (V_{GS} = −3 to −1 V) the leakage current (in a log scale) is independent of gate bias and depends on gate length. The subsurface leakage current shown in Fig. 2 is not due to the leakage current components occurs in conventional MOSFET such as p-n junction leakage current, GIDL current and parasitic BJT current. The drain-to-body p-n junction leakage current is independent of gate length [5]. The GIDL current is more sensitive to the negative gate bias, GIDL rises with increasing negative bias and it can be minimized by lightly doped drain (LDD) architecture [25]. The contribution of parasitic BJT current is added when source-body junction is forward biased. In order to justify the subsurface leakage current, device parameter based model need to be studied.

Fig. 1 Cross-sectional view of n-channel MOSFET device structure
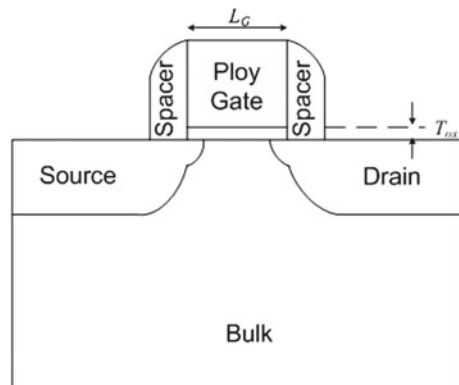
Fig. 2 Drain current (I_{DS}) variation with respect to gate to source voltage (V_{GS}) characteristics of nMOSFET for different gate length (L_G) at $X_J = 60$ nm, $W = 1$ μm, $T_{ox} = 2$ nm and $V_{DS} = 0.2$ V
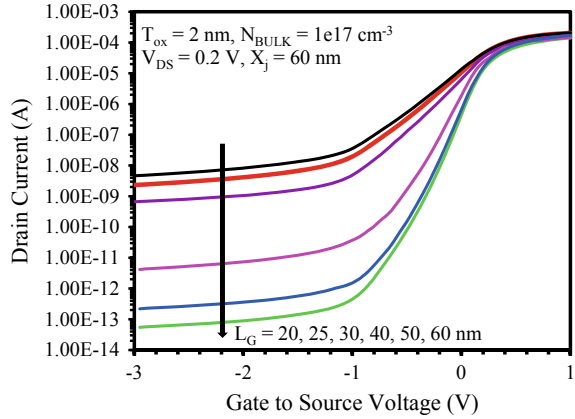
3 Impact of Dimensional Effects on Subsurface Leakage Current

3.1 Drain-to-Source Voltage and Gate Length

The electron current density contour plots of nMOSFET for different gate lengths ($L\_G = 20, 25, 30, 40$ and 50 nm) is depicted in Fig. 3a–e. It shows that, away from Si/SiO$_2$ interface beneath the channel region the subsurface leakage current paths have been established between source and drain. From Fig. 3 it can be clearly noticed that subsurface leakage current path is going away Si/SiO$_2$ interface and interaction of leakage path is also decreasing as gate length increases from 20 to 50 nm. In other words, the gate control over the channel is increases and leakage current suppressed. Thus, subsurface leakage current is sensitive to L_G as demonstrated in Fig. 2.
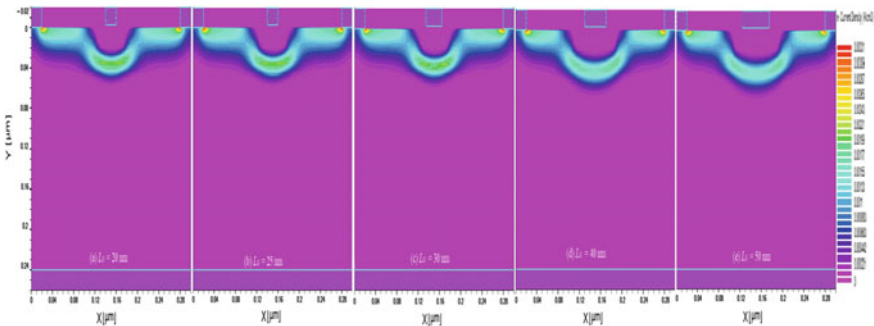
Fig. 3 Electron current density contour plots of nMOSFET at $T_{ox} = 2$ nm, $X_J = 40$ nm, $W = 1$ μm, $V_{GS} = -3$ V and $V_{DS} = 0.1$ V with gate length (L_G) **a** 20, **b** 25, **c** 30, **d** 40, and **e** 50 nm

The subsurface leakage current also reduces the barrier height by increasing V_DS, this effect is analogous to DIBL with diverse intensity. In short channel device, the gate control over the channel is reduced by decrease in barrier height considerably for a smaller value of V_DS. As a result, the electron from source can easily surmount the barrier. With the help conduction band energy diagram the impact of V_DS is demonstrated in Fig. 4.

The drain current variation with respect to gate to source voltage of nMOSFET for different drain-to-source voltages at L_G = 30 nm, Tox = 2 nm, X_J = 60 nm and W = 1 μm is shown in Fig. 5. It can be observed from Fig. 5 that the gate length sensitive subsurface leakage current to be exponential and it is clear evident from the slopes in log scale for different V_DS values are almost same. The variation of drain current (subsurface leakage) at V_GS = −3 V in log scale with respect to gate length (L_G) for different drain-to-source voltages is depicted in Fig. 6. The leakage current increases with decreasing gate length and increasing drain-to-source voltage. Figure 7 illustrates the drain current variation with respect to drain-to-source voltage for different gate lengths (L_G = 20, 30, 40, 50, and 60 nm) at V_GS =

Fig. 4 Conduction band energy diagram of nMOSFET for different drain-to-source voltage at L_G = 30 nm, T_{ox} = 2 nm, X_J = 60 nm, V_{GS} = –3 V
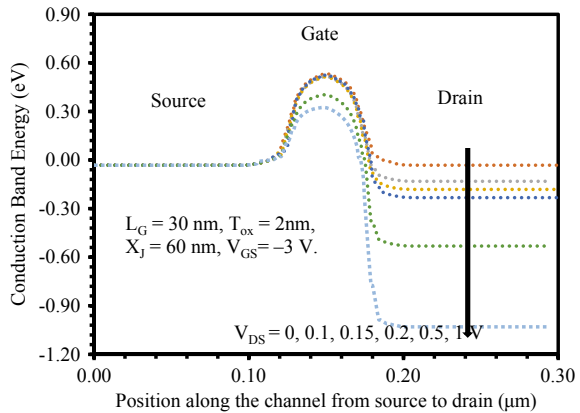
Fig. 5 Drain current variation with respect to gate to source voltage for different drain-to-source voltage of nMOSFET at L_G = 30 nm, T_{ox} = 2 nm, X_J = 60 nm
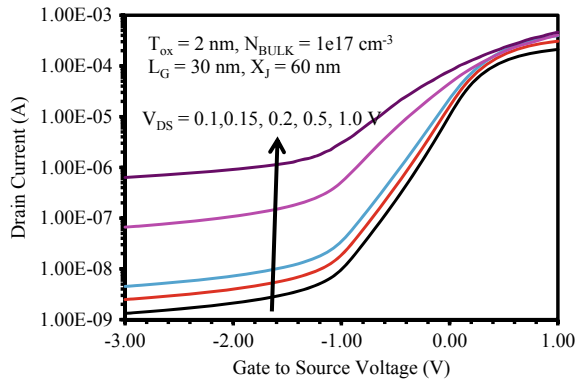
Fig. 6 Drain current (I_{DS}) at $V_{GS} = -3$ V variation with respect to gate length (L_G) for different drain-to-source voltage (V_{DS}) at $N_{BULK} = 1 \times 10^{17}$ cm^{-3}, $X_J = 60$ nm, $T_{ox} = 2$ nm
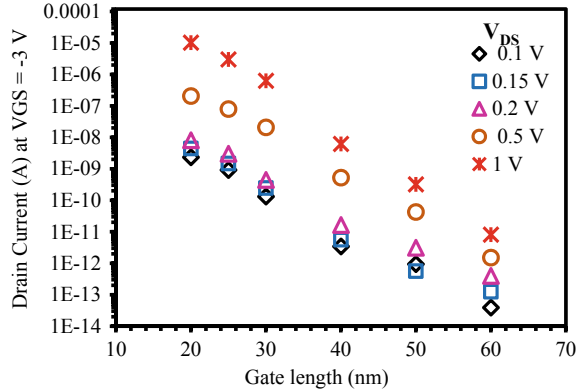
Fig. 7 Drain current variation with respect to drain-to-source voltage for different gate length at $W = 1$ μm, $T_{ox} = 2$ nm, $X_J = 60$ nm
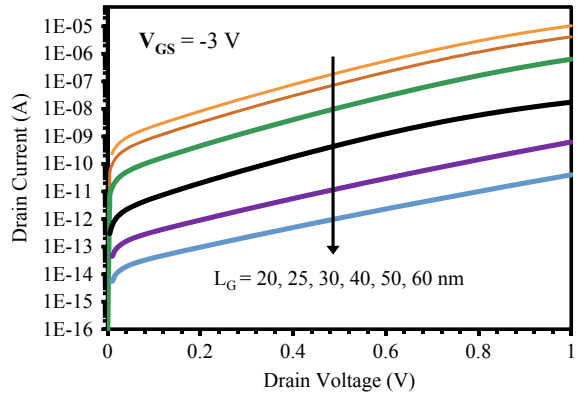
-3 V. The drain current exponentially depends on V_DS. However, the curves are marginally deviating from exponential trend at higher V_DS which is due to the nonlinear barrier height lowering. The high injected electrons influence the space-charge effect, because it screen out the electric field by drain terminal as a result, reduction of barrier height lowering [26]. The low mobility implies the minimized drift current across drain-body reverse bias junction which results into more linear barrier lowering. However, the linear approximation is adopted for interpretation and subsurface leakage current mentioned in [27].

It can be observed that, when VDS = 0 V, the leakage current is also zero. Moreover, for larger gate length (LG > 70 nm) the drain current is independent of gate length which means that the reverse-biased junction leakage dominate the drain current.

Fig. 8 Drain current (I_{DS}) variation with respect to gate-to-source voltage (V_{GS}) characteristics of nMOSFET for different source/drain junction depth (X_J) at $L_G = 40$ nm, W = 1 μm, $T_{ox} = 2$ nm, and $V_{DS} = 0.1$ V
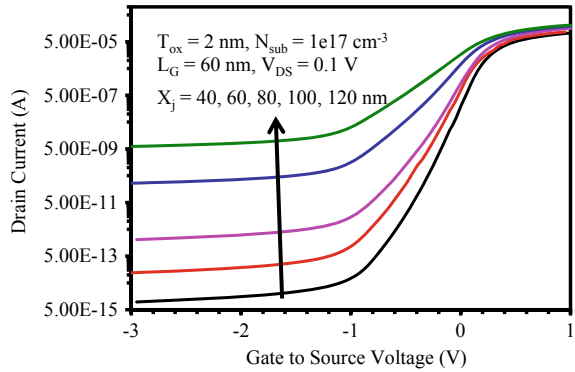
3.2 Source/Drain Junction Depth

In long channel MOSFETs, the source/drain junction depth (XJ) is larger than the gate depletion layer width. The XJ influences the device performance such as threshold voltage electron drift, and subthreshold characteristics. In addition to that it modulates the subsurface leakage current as well. The drain current variation with respect to gate to source voltage is demonstrated in Fig. 8. The electron current density cotour plots of nMOSFET for different source/drain junction depth (XJ) at $L_G = 40$ nm, W = 1 μm, $T_{ox} = 2$ nm, $V_{GS} = -3$ V and $V_{DS} = 0.1$ V is illustrated in Fig. 9. It can be noticed from Fig. 8 that in case of shallower X_J, subsurface leakage current is less which is due to the fact that, the leakage current path is close to the surface and better gate control. It is clear evident that as X_J increase from 40 to 120 nm the subsurface leakage current increases. We can observe that the subsurface leakage current dependence on X_J to be in exponential relationship from Fig. 8. The subsurface leakage current is higher for higher X_J. It can be observed that the exponential increase in subsurface leakage current is inversely proportional to technology node.
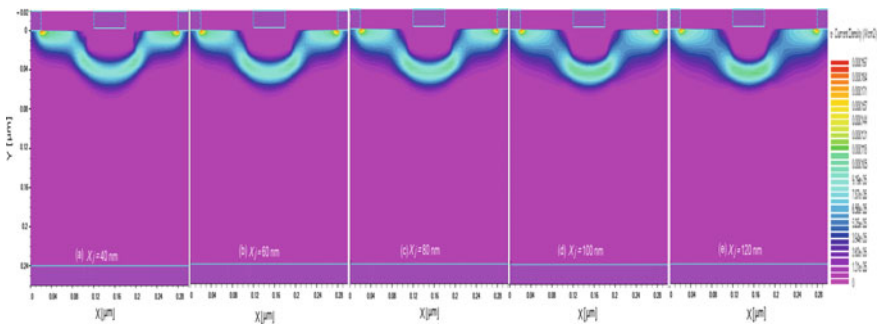
Fig. 9 Electron current density contour plots of nMOSFET at $L_G = 40$ nm, W = 1 μm, $T_{ox} = 2$ nm, $V_{GS} = -3$ V and $V_{DS} = 0.1$ V with source/drain junction depth (X_J) **a** 40, **b** 60, **c** 80, **d** 100, and **e** 120 nm

4 Conclusion

Due to V_{DS}, the reduction of barrier height between the source and drain region at a distance away from semiconductor–oxide interface is known as subsurface leakage current. A device dimensional parameter dependence subsurface leakage current extracted from ATLAS TCAD simulation results. The subsurface leakage current is influenced by device dimensional parameters. The subsurface leakage current is estimated including gate to source voltage (V_{GS}), gate length (L_G), source/drain junction depth (X_J). As device scaled down into sub-45 nm regime (short channel), the interaction between source and drain increases which leads to higher leakage current. Further, subsurface leakage current increases with X_J.

References

1. Frank, D.J., Dennard, R.H., Nowak, E., Solomon, P.M., Taur, Y., Wong, H.: Device scaling limits of Si MOSFETs and their application dependencies. Proc. IEEE **89**(3), 259–288 (2001)
2. Chang, L., Choi, Y.-K., Ha, D., Ranade, P., Xiong, S., Bokor, J., Hu, C., King, T.-J.: Extremely scaled silicon nano-CMOS devices. Proc. IEEE **91**(11), 1860–1873 (2003)
3. Skotnicki, T., Hutchby, J.A., King, T.-J., Philip Wong, H.-S., Boeuf, F.: The end of CMOS scaling—toward the introduction of new materials and structural changes to improve MOSFET performance. IEEE Circ. Dev. Mag. **21**(1), 16–26 (2005)
4. Dennard, R.H., Gaensslen, F.H., Yu, H.-N., Rideout, V.L., Bassous, E., Leblanc, A.R.: Design of ion-implanted MOSFET's with very small physical dimensions. IEEE J. Solid-State Circ. **SC-9**, 256–268 (1974)
5. Roy, K., Mukhopadhyay, S., Mahmoodi-Meimand, H.: Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. Proc. IEEE **91**(2), 305–327 (2003)
6. Lo, S.-H., Buchanan, D.A., Taur, Y., Wang, W.: Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultrathin-oxide nMOSFET's. IEEE Electron Device Lett. **18**(5), 209–211 (1997)
7. Vasudev, P.K.: Ultra-thin silicon-on-insulator for high speed sub-micrometer CMOS technology. Solid State Technol. 61–65 (1990)
8. Swaina, S.K., Dutta, A., Adakb, S., Pati, S.K., Sarkar, C.K.: Influence of channel length and high-K oxide thickness on subthreshold analog/RF performance of graded channel and gate stack DG-MOSFETs. Microelectron. Reliab. **61**, 24–29 (2016)
9. Veeraraghavan, S., Fossum, J.G.: Short-channel effects in SOI MOSFET's. IEEE Trans. Electron Dev. **36**(3), 522–528 (1989)
10. Ohguro, T., et al.: 0.18 μm low voltage/low power RF CMOS with zero V_{TH} analog MOSFETs made by undoped epitaxial channel technique. IEDM Tech. Dig. 33.5.1–33.5.4 (1997)
11. Junior, C.S., Deotti1, D., da Ponte1, R.M., Machado1, M.B., Schneider, M.C.: Zero-Threshold-Voltage MOSFETs: a survey. (2012). http://www.lbd.dcc.ufmg.br/colecoes/sforum/2012/0015.pdf
12. Burr, J.B., Schott, J.: A 200 mV self-testing encoder/decoder using Stanford ultra-low-power CMOS. In: IEEE International Solid-State Circuits Conference (1994)
13. de Sousa, F.R., Machado, M.B., Galup-Montoro, C.: A 20 mV colpitts oscillator powered by a thermoelectric generator. In: IEEE International Symposium on Circuits and Systems, Seoul, Korea (2012)
14. Voldman, S.H.: ESD: circuits and Devices, 2nd edn. Wiley, West Sussex, UK (2015)

15. Paliouras, V., Vounckx, J., Verkest, D., ed.: Integrated circuit and system design: power and timing modeling, optimization and simulation. Springer, Heidelberg (Lecture Notes in Computer Science), vol. 3728 (2005)
16. Yuan, F.: CMOS Circuits for Passive Wireless Microsystems. Springer, New York (2010)
17. Lee, E.K., Lam, A., Li, T.: A 0.65 V rail-to-rail constant g_m opamp for biomedical applications. In: IEEE International Symposium on Circuits and Systems, pp. 2721–2724 (2008)
18. Li, K.-T., Lin, C.-C.: High voltage tolerant I/O circuit using native NMOS transistor for improved performance. US Patent No. 7,113,018 (2006)
19. Eberlein, M.: A technology compensated current reference in standard CMOS. In: 7th International Conference on ASIC, pp. 522–525 (2007)
20. Banba, H., Shiga, H., Umezawa, A., Miyaba, T., Tanzawa, T., Atsumi, S., Sakui, K.: A CMOS bandgap reference circuit with sub-1-V operation. IEEE J. Solid-State Circ. **34**, 670–674 (1999)
21. Hunt, J.S., Saripella, S.C.: Input buffer circuit. US Patent No. 6,784,700 (2004)
22. C.-Y. Chen, "Power management unit for use in portable applications" US Patent No. 6,879,142 (2005)
23. Device Simulation Software: Silvaco International. Santa Clara, CA (2012)
24. Chauhan, Y.S., et al.: BSIM6: Analog and RF compact model for bulk MOSFET. IEEE Trans. Electron Device **61**(2), 234–244 (2014)
25. Parke, S.A., Moon, J.E., Wann, H.C., Ko, P.K., Hu, C.: Design for suppression of gate-induced drain leakage in LDD MOSFETs using a quasi-two-dimensional analytical model. IEEE Trans. Electron Devices **39**(7), 1694–1703 (1992)
26. Sze, S.M., Ng, K.K.: Physics of Semiconductor Devices, 3rd edn. Wiley, New York, NY, USA (2007)
27. Lin, Y.-K., et al.: Modeling of subsurface leakage current in low VTH short channel MOSFET at accumulation bias. IEEE Trans. Electron Devices **63**(5), 1840–1845 (2016)

Dielectric Pocket (DP) Based Channel Region of the Junction-Less Dual Material Double Gate (JLDMDG) MOSFET for Enhanced Analog/RF Performance

Amrish Kumar, Abhinav Gupta and Sanjeev Rai

Abstract In this paper, Dielectric pocket (DP) based channel region of the junction-less dual material double gate (JLDMDG) MOSFET is proposed. A comparative analysis of the JLDMDG and DPDMDG on the basis of analog/RF performance parameters has been done. Dielectric pockets are introduced in the channel region of JLDMDG. It can be observed that DPDMDG MOSFET has better functional density, higher early voltage and lower trans-conductance for high amplification. It can also be compared RF performance which shows that DPDMDG MOSFET has higher gain bandwidth product, lower values of capacitances for high-speed operation and lower power dissipation. Sensitivity parameters are estimated for DPDMDG and JLDMDG and it is observed that I_{on} and I_{off} in DPDMDG is less sensitive to variation in t_{si} and t_{ox} in compare with JLDMDG. The analog/RF performance has shown that the DPDMDG MOSFET has better characteristics in low-power design and high-frequency applications.

Keywords SCEs · Dielectric pocket (DP) DPDMDG · JLDMDG

A. Kumar (✉) · S. Rai
Department of Electronics & Communication Engineering,
Motilal Nehru National Institute of Technology, Allahabad 211004, India
e-mail: amrish1288@gmail.com

S. Rai
e-mail: srai@mnnit.ac.in

A. Gupta
Electronics Engineering Department, Rajkiya Engineering College Sonbhadra, Sonbhadra 231206, India
e-mail: abhinavkit87@gmail.com

1   Introduction

During the past few decades, the device performance with excellent high speed has been achieved by the upgraded design. The design mainly uses of high class material and scaling device dimensions. However, the biggest challenges in upcoming technology, is to control the short-channel effects while the down scaling of channel length. The short-channel effects are more dominating as we decreasing or down scaling the channel length [1, 2]. The SCEs such as the threshold voltage roll-off, drain induced barrier lowering (DIBL) and hot-carrier effect [3]. Therefore, to overcome the SCEs most of the researchers have been working on the channel engineering with different structure and different materials. In the recent trends, Junction-less (JL) Transistors [4] is more suppress the short-channel effects. It reduces the complex fabrication process which is used to formation the junction. In a junction-less devices have constant doping concentration throughout the source, drain and channel regions, it means there are no junction i.e. (source-gate or drain-gate) junction. Junction-less transistor is behaving as variable resistor and it operates in fully depleted regions. The conduction mechanism of JL transistor is bulk conduction instead of surface channel conduction. Due to this the leakage current in a junction-less device is less compared to conventional double gate (DG) MOSFET. Dielectric pocket (DP) [5, 6] based channel region of the junction-less dual material double gate (JLDMDG) MOSFET is called as DPDMDG MOSFET. The dielectric pockets are simply a SiO_2 layer area exists at the lateral walls of the channel region. DP cause to limit the diffusion which happens between highly doped source and drain region of the MOSFET. Due to this, the charge sharing decreases and improves the device performance. In the DPDMDG device, the importance of dual material gate (DMG) [7] provides high electron velocity, enhance drive current, trans-conductance and reduce the output conductance. A DMG has two different gate electrodes one is control gate M1 (workfunction of control gate M1 is qφM1) and another one is screening gate M2 (workfunction of screening gate M2 is qφM2) where work function (qφM1) is greater than (qφM2) for n-channel MOSFET. DPDMDG improves the analog/RF performance by using the structure of DMG. In Sect. 2 discusses the DPDMDG device structure and simulation model. Section 3 shows the simulation results and discussion with comparison of analog/RF performance of DPDMDG and JLDMDG MOSFET along with their sensitivity analysis. In Sect. 4 paper has been concluded.

2   Device Structure and Simulation

The structure of junction-less dual material double gate (JLDMDG) MOSFET with dielectric pocket (DP) in the channel region has been proposed. JLTs having heavily doping throughout the source, drain, and channel regions. The device is turn OFF condition when it operates in fully depleted mode. For this purpose, a gate metal which has a large work function difference to that of the channel is needed. The
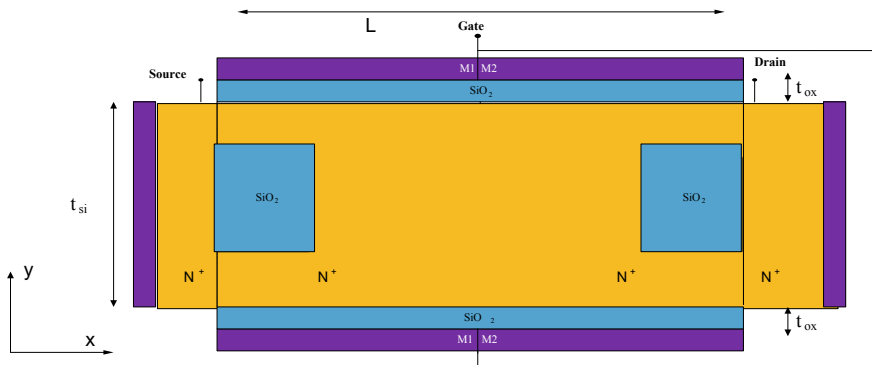
Fig. 1 Device structure of n-channel DPDMDG MOSFET

dielectric pockets are simply a SiO$_2$ layer introduced at the lateral walls of the channel region. DP cause to limit the diffusion which is happen between highly doped source and drain region of the MOSFET.

The device structure of DPDMDG MOSFET is shown in Fig. 1. The thickness of gate oxide t_{ox}, gate length L and channel thickness t_{Si} are 1.5 nm, 50 nm and 10 nm respectively. The work functions for control gate M1 and screening gate M2 are 4.77 eV and 4.6 eV, respectively. The doping in source, channel and drain region is pentavalent to make the transistor N type. The device parameters [8] of DPDMDG transistor are shown in Table 1.

The simulation has been done by Silvaco ATLAS [9] device simulator. During the simulation constant voltage and temperature (CVT) mobility model, Shockley–Read–Hall (SRH) and Auger recombination models for minority carrier combination have been used. In the simulation standard recombination models along with The Fermi–Dirac carrier statistics have been used. For the solving equation of the models we choose Newton's Gummel methods. As the thickness of the proposed model is below 10 nm so Quantum mechanical effect will comes into play. So, drift diffusion model is used in simulation and neglected the Quantum mechanical effect (QME).

Symbol	Quantity	Value
L	Channel length	50 nm
N_D	Doping concentration	2×10^{18} Cm^{-3}
t_{Si}	Channel thickness	10 nm
t_{ox}	Gate oxide thickness	1.5 nm
W_1	Work function for gate1	4.77 eV
W_2	Work function for gate2	4.6 eV
L_d	Length of dielectric pocket	6 nm
W_d	Width of dielectric pocket	5 nm

Table 1 Device parameter of DPDMDG MOSFET

3 Results and Discussion

The analog performance [10–12] of any devices can be defined the following important parameters such as, trans-conductance generation factor (TGF) intrinsic gain (Av), trans-conductance (g_m), early voltage (V_{EA}), and output conductance (g_d), etc., from Fig. 2a shows the variation of trans-conductance (g_m) and drain current (I_D) with gate voltage. Trans-conductance is the rate of change in drain current to change in gate voltage for little period of $I_D - V_{GS}$ curve with all other parameters is constant. If the value of g_m is higher results higher amplification factor of the device. Figure 2a shows that as we increase the gate to source voltage, g_m increases up to 0.7 V, i.e., peak value of g_m after that g_m start to reduce. Trans-conductance is defined as

$$g_m = \frac{dI_D}{dV_{GS}} \tag{1}$$

The variation of trans-conductance generation factor (TGF) with gate to source voltage is shown in Fig. 2b TGF define as the ratio of trans-conductance and drain current and it shows that the how much drain current required to achieve a definite value of trans-conductance. So, maximum value of TGF means that we need minimum value of drain current to reach desired trans-conductance. Hence, TGF is defined as

$$TGF = \frac{g_m}{I_D} \tag{2}$$

Figure 3a shows the variation of early voltage (V_{EA}) and intrinsic gain (A_V) with respect to gate voltage. Early voltage is the ratio of drain current to output conductance. As we increase the gate voltage the V_{EA} is decreases but after certain gate
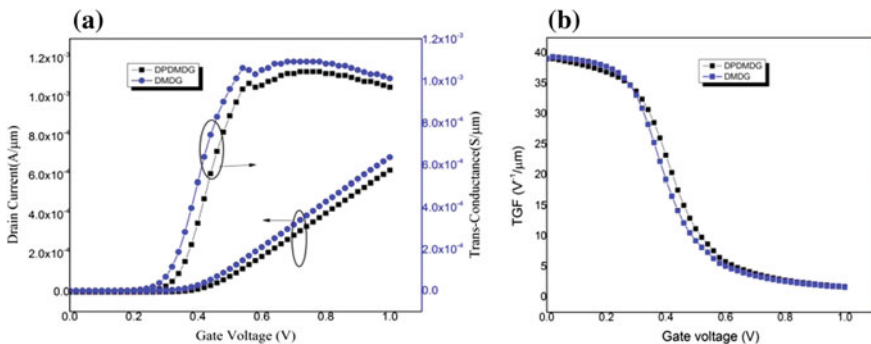
Fig. 2 **a** Variation of Drain Current and trans-conductance as a function of gate voltage. **b** Deviation of TGF with gate voltage (V_{GS}) with fixed drain voltage (V_{DS})
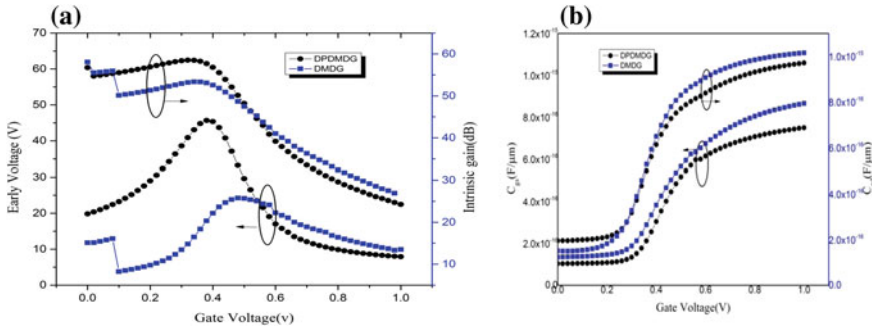
Fig. 3 **a** Deviation of early voltage (V_{EA}) and intrinsic gain (A_V) with gate voltage (V_{GS}) with fixed drain voltage (V_{DS}). **b** Deviation of C_{gs} and C_{gd} as a function of gate voltage with fixed drain voltage

voltage it become constant because of device comes in the saturation region. Intrinsic gain is the ratio of trans-conductance and output conductance. As we increase the gate voltage the A_V is decreases. For the better analog performance of the device is that the early voltage and intrinsic gain both should be higher. The calculation formula for both is given as

$$V_{EA} = \frac{I_D}{g_d} \tag{3}$$

$$A_V = \frac{g_m}{g_d} = \frac{g_m}{I_D} V_{EA} \tag{4}$$

The RF performance [13, 14] of any devices can be defined the following important parameters such as gain frequency product (GFP), cut-off frequency (f_T), trans-conductance frequency product (TFP) and gain trans-conductance frequency product (GTFP) . Firstly, we calculate the capacitances C_{gd} and C_{gs}, i.e., C_{gd} means capacitance between gate to drain terminal, i.e., and C_{gs} means capacitance between gate to source terminal, i.e. Now we apply a DC ramp voltage signal which is varying from 0 to 1 V with 1 MHz frequency. The step voltage is 0.05 V.

Figure 3b shows the variation of C_{gs} and C_{gd} with gate voltage and both C_{gs} and C_{gd} increases when we increase the gate voltage. From the Eq. 5 the capacitances C_{gs} and C_{gd} should be as low as possible to get the enhanced RF performance. Figure 4 shows the variation of cut-off frequency w.r.t. gate voltage (V_{gs}) for different values of V_{ds}. The characteristics of the device behave as an amplifier up to cut-off frequency. The calculation formula for cut-off frequency is given as

$$f_T = \frac{g_m}{2\pi \left(C_{gs} + C_{gd} \right)} \tag{5}$$
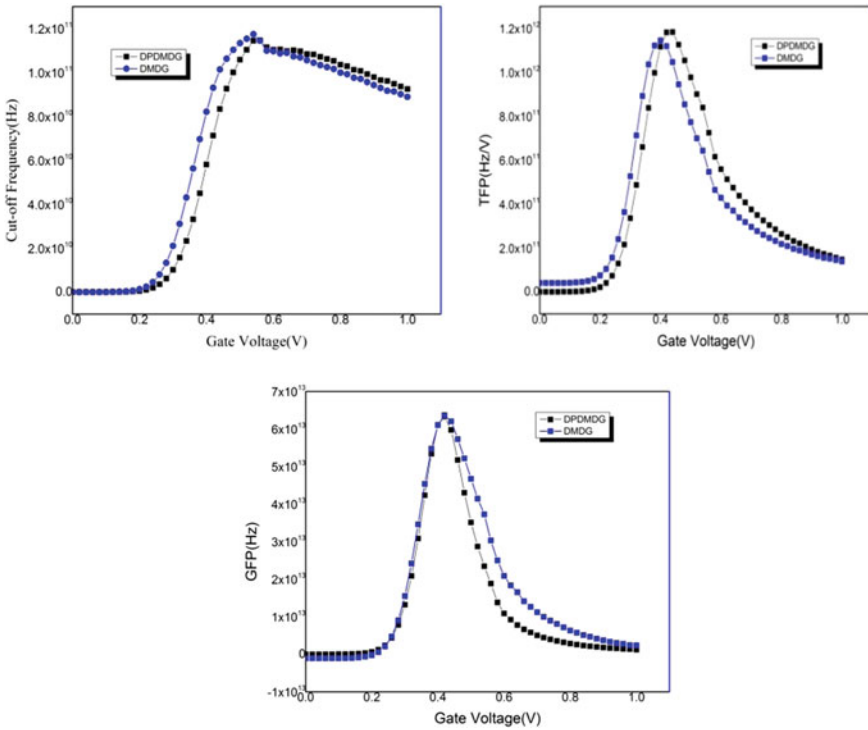
Fig. 4 Deviation of cut-off frequency, TFP and GFP as a function of gate voltage

The GTFP is the products of intrinsic gain and trans-conductance frequency product (TFP). This parameter decides the device switching speed as high as GTFP. It means the high value of GTFP gives the better device RF performance. It is defined as

$$GTFP = \left(\frac{g_m}{g_d}\right) \times \left(\frac{g_m}{I_D}\right) \times f_T \tag{6}$$

$$GFP = \left(\frac{g_m}{g_d}\right) \times f_T \tag{7}$$

$$TFP = \left(\frac{g_m}{I_D}\right) \times f_T \tag{8}$$

Figure 4 shows the variation of TFP and GFP with increasing gate voltage (V_{gs}) for different values of V_{ds}. GFP is important aspect of the device that combines both low as well as high frequency. GFP is the product of intrinsic gain and cut-off frequency. TFP can be defined as product of cut-off frequency and TGP. It represents that power utilization and bandwidth cannot be improved simultaneously. From Fig. 4 when

Table 2 Sensitivity matrix for DPDMDG and JLDMDG

Parameters	DPDMDG	JLDMDG
$S_{Ion}(t_{si})$	0.37	0.5
$S_{Ioff}(t_{si})$	15.9	21.3
$S_{Ion}(t_{ox})$	−0.53	−0.664
$S_{Ioff}(t_{ox})$	11.95	12.05

device operate in the triode region, the TFP and GFP both are increase with gate voltage. Both are decreases in the saturation region. The expression for GFP and TFP is shown in Eqs. 7 and 8 respectively.

3.1 Sensitivity Analysis of the Devices

Sensitivity analysis offers a relative significance of each device parameters on particular performance metrics, for the DPDMDG and DMDG devices, we have to calculate the sensitivity .The sensitivity metric (S) on a parameter (p) is defined as

$$S_S(p) = \frac{\partial S}{S} \Big/ \frac{\partial p}{p} \tag{9}$$

where S is parameter for which we want to calculate sensitivity and p is the device parameter such as t_{si}, t_{ox}. Table 2 compares sensitivity parameter for JLDMDG and DPDMDG. Here we have estimated sensitivity in I_{on} and I_{off} with respect to t_{on} and t_{ox}.

Here we have estimated sensitivity of I_{on} and I_{off} with respect to t_{si} and t_{ox}. It can be observed that DPDMDG MOSFET is less sensitive to variation in t_{si} and t_{ox} in comparison with JLDMDG.

4 Conclusion

In this paper, a dielectric pocket (DP) based channel region of the junction-less dual material double gate (JLDMDG) MOSFET structure has been simulated. A comparative analysis of the JLDMDG and DPDMDG on the basis of analog/RF performance parameters has been done. Form the simulation result, it shows that JLDMDG has better gate controllability over the channel. It can be observed that DPDMDG MOSFET has better functional density, higher early voltage and lower trans-conductance for high amplification. The RF performance shows that DPDMDG MOSFET has higher gain bandwidth product, lower values of capacitances for high speed operation and lower power dissipation. Sensitivity parameters are estimated for DPDMDG and JLDMDG and it is observed that I_{on} and I_{off} in DPDMDG is

less sensitive to variation in t_{si} and t_{ox} in compare with JLDMDG. The analog/RF performance has verified that DPDMDG MOSFET has improved characteristics in low-power design and high-frequency applications.

References

1. Toyabe, T., Kadera, H., Kodera, H.: Generalized scaling theory and its application. **41**(4), 1283–129 (1984)
2. Liang, X., Taur, Y.: A 2-D analytical solution for SCEs in DG MOSFETs. IEEE Trans. Electron Devices **51**(9), 1385–1391 (2004)
3. Li, Cong, et al. "Analytical threshold voltage model for cylindrical surrounding-gate MOSFET with electrically induced source/drain extensions." *Microelectronics Reliability* 51.12 (2011): 2053–2058
4. Colinge, J.P., et al.: Nanowire transistors without junctions. Nat. Nanotechnol. **5**(3), 225–229 (2010)
5. Kumari, V., et al.: Analytical modeling of dielectric pocket double-gate MOSFET incorporating hot-carrier-induced interface charges. IEEE Trans. Device Mater. Reliab. **14**(1), 390–399 (2014)
6. Gupta, A., Maurya, N., Rai, S.: Impact of dielectric pocket on analog/RF performance of short channel double gate MOSFET. In: 2017 4th International Conference on Power, Control & Embedded Systems (ICPCES). IEEE (2017)
7. Reddy, G.V., Jagadesh Kumar, M.: A new DMDG nanoscale SOI MOSFET-two dimensional analytical modeling and simulation. IEEE Trans. Nanotechnol. **4**(2), 260–268 (2005)
8. International Technology Roadmap for Semiconductors. http://public.itrs.net
9. ATLAS Device Simulator Software: Silvaco. Santa Clara, CA, USA (2015)
10. Kumar, A., Gupta, A., Rai, S.: Reduction of self-heating effect using selective buried oxide (SELBOX) charge plasma based junctionless transistor. AEU-Int. J. Electron. Commun. (2018)
11. Kumar, A., Gupta, A., Rai, S.: Charge plasma based graded channel with dual material double gate JLT for enhance analog/RF performance. In: 4th International Conference on Power, Control & Embedded Systems (ICPCES), pp. 1-6-2017/11. IEEE
12. Chauhan, R., Kumar, A., Rai, S.: A comparative study of junctionless dual material double gate silicon on insulator (SOI) and silicon on nothing (SON) MOSFET. In: 4th International Conference on Power, Control & Embedded Systems (ICPCES), pp. 1-7-2017/11 IEEE
13. Srivastava, M., Kumar, A., Rai, S.: Analytical model and performance investigation of electric potential for junctionless cylindrical surrounding gate (JLCSG) MOSFET. In: 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 256–261. IEEE
14. Malik, P., et al.: AC analysis of nanoscale GME-TRC MOSFET for microwave and RF applications. Microelectron. Reliab. **52**(1), 151–158 (2012)

Study and Performance Analysis of Carbon Nanotubes (CNTs) as a Global VLSI Interconnects

Kavindra Kumar Kavi

Abstract This research works are primarily focused on the modeling approach of Carbon Nanotubes (CNTs). Based on the number of conducting channels and different number of shells and their arrangements in bundles an Equivalent Single Conductor (ESC) transmission line model proposed to analyze the effect of power dissipation and propagation delay. Driver Interconnect Load (DIL) system is used to analyze the power dissipation and propagation delay performances. A SPICE simulation is used to compare the performance of bundle Multiwall Carbon Nanotubes (MWCNTs) as compare to conventional bundle Single Wall Carbon Nanotubes (SWCNTs) interconnect. After the performance analysis it is observed that power dissipation and propagation delay increase with the length of interconnects, but the power dissipation and propagation delay decreases for bundle of MWCNT as compare to the conventional bundle SWCNT.

Keywords Carbon nanotube (CNT) · Multiwall CNT bundle (MWCNT) · Power dissipation · Propagation delay · Interconnects · Driver interconnect load (DIL)

1 Introduction

On the silicon chip to provide an electrical connection among nodes of circuit and systems a thin film of conducting material used called as a VLSI interconnect. It plays a major role in 32 nm and below nanoscale technologies Interconnects [1]. In the recent age of high-speed technology, it is much important that device speed should be enhanced in order to provide high speed applications. The speed of VLSI circuits or IC's depends on the main circuit dimensions as well as interconnect used to connect various components [2]. To analyze these dimensional effects, different measures are used like use of thick wires, insertion of repeaters and applying distributed interconnects [3].

K. K. Kavi (✉)
Department of ECED, MNNIT Allahabad, Allahabad, India
e-mail: kavipsu@gmail.com
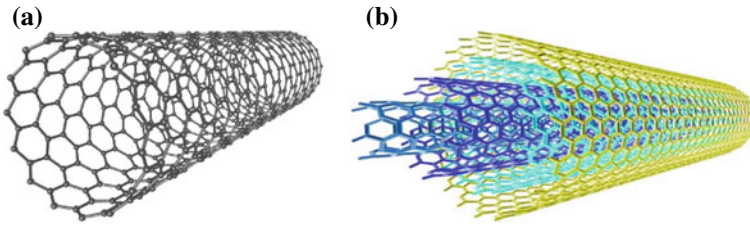
Fig. 1 **a** SWCNT [9]. **b** MWCNT [10]

Carbon nanotubes (CNTs) have many applications due to their unique properties like large current carrying capability [4], high thermal conductivity and high mechanical strength [5] and high thermal stability [6]. CNTs are basically an allotropes of carbon which are made up by rolling up a graphene sheets in cylindrical form [7]. Based on number of shells CNTs are of two types, single walled CNT (SWCNT) and multiwalled CNT (MWCNT) [8]. A single wall Carbon nanotubes (SWCNT) consists of one rolled up graphene sheets with diameter of few nm [9], where multi walled carbon nanotubes (MWCNTs) are formed by several rolled up graphene sheets coaxially with a diameter ranges from 2 nm to several tens of nanometres. The SWCNTs have attracted much more attention than MWCNTs are shown in Fig. 1 due to their simple structure [10].

This research paper produce a hierarchical modeling aspects and analysis of performances for different CNT bundled structures for VLSI interconnect. Depending on the different arrangements of SWCNTs and MWCNTs, there are three various structures are proposed for CNT interconnects.

2 Physical Structures of CNTs Interconnects

2.1 An Isolated SWCNT

An isolated SWCNT situated on ground plane at the height of "h" and the diameter of SWCNT is "d" as shown in Fig. 2 [11]. The diameter of an isolated SWCNT can calculate by the given formula

$$d_{SWCNT} = \frac{0.249}{\pi} \sqrt{m^2 + n^2 + mn} \text{ nm} \tag{1}$$

where m and n are the chiral vectors.

It has been assume that the height SWCNT is almost equal to the length of an isolated SWCNT. The parameter "h" can be replaced by the length of SWCNT, if there is no ground plane.

2.2　SWCNT Bundle

The CNTs are placed in fix dimensional area which having a geometrical significance as well as easy to convert into its equivalent circuit models. The different types of modeling and design parameters to be calculated based on geometry of CNTs. In the following section according to ITRS the SWCNT bundle and MWCNT are situated in 32 nm technology node with *AR* are 2.5.

The modeling of SWCNTs can be done easily due to the simple structure as compare to MWCNTs. In fact most CNTs interconnect fabrication efforts are focused on analysis of an isolated SWCNT or SWCNT bundle. There are great many number of SWCNTs are arrange in a bundle such a way that each SWCNT is by surrounded by six immediate neighbors shown in Fig. 3.

The distances between two consecutive SWCNTs are called as spacing (S_P) between SWCNTs. It is a center to center distance of SWCNTs. The diameter of

Fig. 2 SWCNT of diameter "d" and distance "h" from ground plane
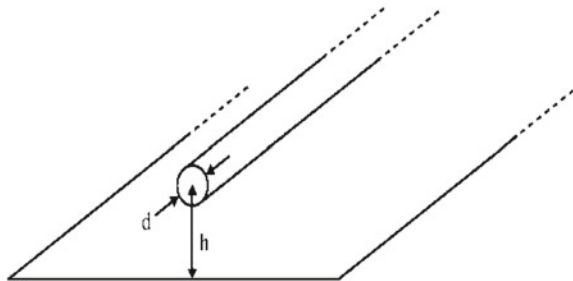
Fig. 3 Geometry of SWCNTs bundle
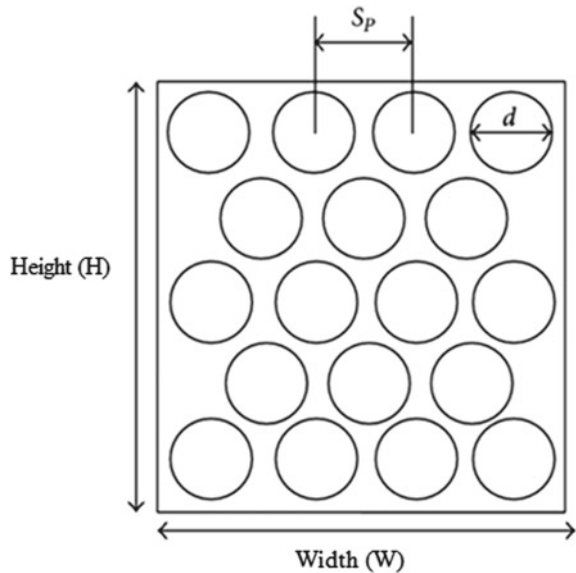
each SWCNT is d (1 nm). The distance between two SWCNTs is called inter shell spacing (δ) which is approximately equal to van der Wall's gap between neighboring carbon atoms. The relation between spacing (S_P) and inter shell spacing (δ) is given by

$$S_P = d + \delta \tag{2}$$

The number of SWCNTs in rows of bundle or along width is N_W and the number of SWCNTs in columns of bundle or along with height is N_H. The total number of SWCNTs can be obtained by using bundle height and width such as total number of SWCNT as below

$$N_{SWCNT} = N_W N_H - integer\left[\frac{N_H}{2}\right], \text{ If } N_H \text{ is even,} \tag{3}$$

$$N_{SWCNT} = N_W N_H - integer\left[\frac{N_H - 1}{2}\right], \text{ If } N_H \text{ is odd,} \tag{4}$$

where, $N_W = \frac{W-d}{S_P}$ and $N_H = \frac{H-d}{S_P}$.

2.3 MWCNT Bundle

MWCNTs are made of many number of SWCNT situated coaxially (as shown in Fig. 4) and each shell in an MWCNT can be metallic or semiconducting depends on different chiralities. To achieve the low resistance with the contribution of inner shells the contact of MWCNTs must be properly implemented [12].

According to the geometry of MWCNT above ground plane it can be observed that number of rolled up graphene sheets with diameters $D_1, D_2 \ldots D_n$ concentrically situated [13]. The gap between two consecutive inter-shells is almost equal to the

Fig. 4 Geometry of MWCNT above ground plane
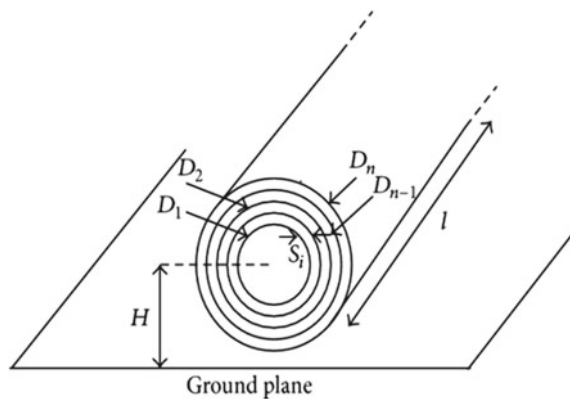
van der Waal's gap (δ) between neighboring carbon atoms such as

$$\delta = \frac{D_n - D_{n-1}}{2} \approx 0.34 \text{ nm} \tag{5}$$

SWCNT having a diameter 1 nm. Since MWCNT having number of shell hence its diameter varies according to the number of shell where outer and inner shell diameters are represented as D_1 and Dn, respectively, where n is the total number of shells. The distance between the ground plane and the center of MWCNT is equivalent to $H = D_n/2 + h_t$, where h_t gives the distance between outermost shell and the ground plane. The diameter of outer most shell of MWCNT can be provides as

$$D_n = D_1 + 2 \times \delta \times (n - 1) \tag{6}$$

where n is No. of shell, δ is 0.34 nm and D_n is Outer shell diameter of MWCNT. According to ITRS 32 technology node, the aspect ratio (AR) of width and height of the bundle is 2.5. The total number of MWCNT can be obtained by using bundle width and height such as

$$N_{MWCNT} = N_W N_H - integer\left[\frac{N_H}{2}\right] \tag{7}$$

where $N_W = \frac{W - D_n}{D_n + \delta} + 1$ and $N_H = \frac{H - D_n}{(D+\delta)\sqrt{3}/2} + 1$; where N_W and N_H represent the number of CNT in vertical and horizontal direction respectively. All the shells of MWCNT conduct current through a conducting channel. The number of conducting channel for each shell is defined as

$$N_n^{channel} = K_1 T D_n + K_2 \quad D_n > \frac{dT}{T} \quad \text{and}$$

$$2/3\, D_n \qquad \leq \frac{dT}{T} \tag{8}$$

where $n = 1, 2, 3 \ldots k$, $T = $ Room Temperature (300 K), expressed in Kelvin, $D_n = $ Radius of the n_{th} shell in nm, $K_1 = 3.87 \times 10^{-4} \text{ nm}^{-1} \text{K}^{-1}$, $K_2 = 0 \cdot 2$, $dT = 1300$ nmK.

$$\frac{dT}{T} = \frac{1300 \text{ nmK}}{300 \text{ K}} = 4.3 \text{ nm} \tag{9}$$

3   Proposed ESC Model

The equivalent single conductor (ESC) transmission line model is an advanced CNT interconnect model which comes after transformation of equivalent distributive RLC model of CNT (Fig. 5). In this model the scattering resistances are varies with value
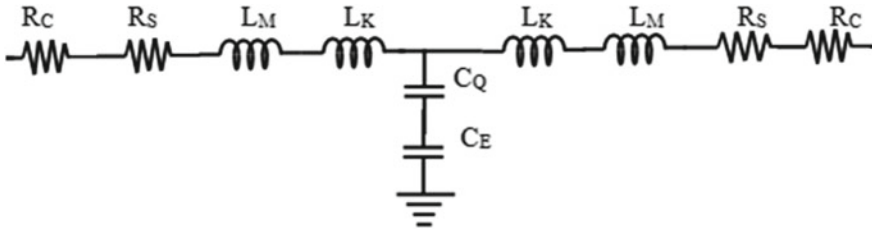
Fig. 5 ESC model of CNTs interconnects

of *mfp* and number of conducting channels. This model is use to connect with driver-interconnect-load (DIL) system [10].

The equivalent resistances for each shell of any CNTs are the summation of the two different types of resistances, contacts resistance (R_C) and scattering resistance (R_S). The contacts resistance which exits among drivers, interconnects and loads at contacts. The contacts resistances for each shell of any CNTs are the summation of the two different types of resistances, i.e., intrinsic contact or quantum contact resistance (R_q) and Imperfect metal nanotubes contact resistance (R_{mc}). The intrinsic contact or quantum contact resistance (R_q) occurs due to quantum confinement of electrons in a nano-wire. The value of intrinsic resistance for single shell in both the side of contacts is 6.45 KΩ. Imperfect metal nanotubes contact resistance (R_{mc}) is due to the imperfection in materials and its structures which are depend on the fabrication process and Its values are varying in the range of 3.2–3.5 KΩ. The equivalent resistances $R_{eq} = Rc + R_S \cdot l$ where,

$$Rc = \left[\left(\frac{1}{(N_{CNT})_{m_n}} \left(\frac{R_q}{\sum_{n=1}^{k} N_n} + R_{mc} \right) \right)^{-1} \right]^{-1} \tag{10}$$

$m_n = m_1, m2, m3 \dots m_k$ are the different type of CNTs in a bundle. $N_n = N_1, N2, N3 \dots N_k$ are different shell of channel. The Scattering resistance (R_S) exist when the nanotubes length (l) exceeding the mean free path (λ) of electrons and the equivalent resistance of the ESC

$$R_S = \left[\left(\frac{1}{(N_{CNT})_{m_n}} \left(\frac{R_q}{\sum_{n=1}^{k} N_n \cdot \lambda_{mfp_n}} \right) \right)^{-1} \right]^{-1} \tag{11}$$

λ_{mfp} mean free path, i.e., lossless path of the CNT. The magnetic inductance (L_M) is depends on the magnetic fields inside and between the shells due to the magnetic field induced by current flowing through a nanotubes.

$$L_K = \frac{L_k}{2N_{total}^{Bundle}}, \text{ where, } L_k = \frac{h}{2e^2 v_f}$$

The kinetic inductance (L_K) exists due to the kinetic energy of the electron.

$$L_M = \frac{\mu_0}{2\pi} \cosh^{-1}\left(\frac{D_n + 2h_t}{D_n}\right) \times N_w \tag{12}$$

where h_t is the height between CNTs and the ground plane and N_w is the number of CNT along with width of bundle. Electrostatic capacitance (C_E) occurs due to the potential difference between the ground plane and the bundle, where as the finite density of electronics states in a quantum wire provides the quantum capacitance (C_Q).

$$C_E = \frac{2\pi \varepsilon_0 \varepsilon_r}{\cosh^{-1}\left[\frac{(D_n + h_t)}{D_n}\right]} \times N_w \tag{13}$$

and

$$C_Q = 2C_q \times N_{total}^{Bundle} \tag{14}$$

where $C_q = \frac{2e^2}{hv_f}$ and D_n is outer shell diameter. The equivalent capacitance of ESC model such as

$$C_{eq} = \left[\frac{1}{C_E} + \frac{1}{C_Q}\right]^{-1} \tag{15}$$

4　Simulation Setup

To compare the power dissipation and propagation delay a DIL system used for different bundled CNT structures as shown in Fig. 6. The DIL system consists of three integrals component as driver, interconnects and load. In the recent VLSI research CMOS inverter is treating as driver due to small power consumption. The power supply used for get ON to the CMOS inverter is $V_{DD} = 1$ V. The interconnect models are the ESC model of SWCNT bundle, Single SWCNT and Double MWCNT. The load in DIL system is used as a capacitive load (C_L). The value of C_L is 10 aF [14].

5　Results and Discussions

Power dissipation and propagation delay are compared using the above-mentioned setup. To analyze the performance among MWCNT bundle and SWCNT bundle interconnect a SPICE simulation is used for the 32 nm technology node given by
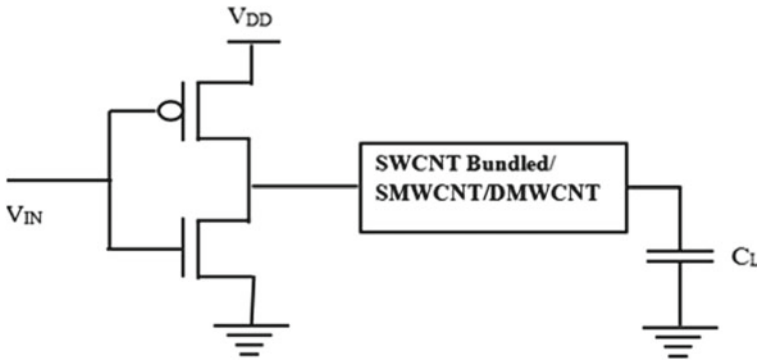
Fig. 6 Simulation setup (DIL system)

International Technology Roadmaps for Semiconductor (ITRS). First of all utilized the above mention dimension for SWCNT bundle accommodation and analyzed the performance of SWCNT bundle. At the same bundle dimension it has try to accommodate single MWCNT in hole bundle but the some of the area infect equal area are unutilized by placing single MWCNT because the bundle have 2.5 time higher length with respect to width. Therefore finally we have taken MWCNT bundle and placed two MWCNT to reduce the unutilized area and compared the performance of CNTs. According to increasing length of the interconnect, the propagation delay and power dissipation also increase due to length dependent resistance that is called scattering resistance. Figure 7 presents the simulation results and comparison of different CNT bundle for propagation delay and power dissipation for global interconnect length. Due to the variation in number of conducting channel the resistive and capacitive parasitic for different CNTs got change and the same way the propagation delay and power dissipation got a regularly increment according to length of CNTs interconnects.
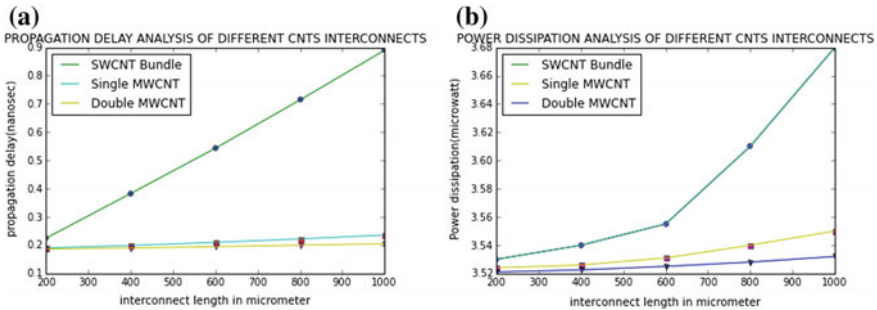
Fig. 7 **a** Propagation delay analysis of different CNT bundle. **b** Power dissipation analysis of different CNT bundle

The simulation results for propagation delay and power dissipation has been shown in Tables 1 and 2 respectively. According to results placed in table it is observed that the power dissipation and propagation delay increase with the global length for all types of CNTs interconnects, but at the same time the percentage reduction in propagation delay for double MWCNT compares to conventional SWCNT bundle and single MWCNT are almost 77% and 13% respectively and the percentage reduction in power dissipation for MWCNT bundle having double MWCNT compares to conventional SWCNT bundle and single MWCNT are almost 4.2% and 0.7% respectively at the global length of interconnects. The area utilization by bundle having double MWCNT is double as compare to bundle having single MWCNT where as lesser as compare to SWCNT bundle. The mix CNTs can give the batter replacement

Table 1 Propagation delay for different CNTs and percentage reduction in propagation delay

Interconnects lengths (μm)	Propagation delay (in ns) of different CNTs			% reduction in propagation delay for the bundle having double MWCNT compared to	
	SWCNT bundle	Bundle having single MWCNT	Bundle having two MWCNT	SWCNT bundle	Bundle having single MWCNT
100	0.193	0.187	0.185	4.14	1.06
200	0.225	0.190	0.187	16.88	1.57
400	0.3442	0.199	0.190	44.76	4.50
600	0.543	0.210	0.195	64.08	7.14
800	0.714	0.222	0.200	71.98	9.90
1000	0.890	0.236	0.205	76.96	13.13

Table 2 Power dissipation for different CNTs and percentage reduction in power dissipation

Interconnects lengths (μm)	Power dissipation (in microwatt) of different CNTs			% reduction in power dissipation for the bundle having double MWCNT compared to	
	SWCNT bundle	Bundle having single MWCNT	Bundle having two MWCNT	SWCNT bundle	Bundle having single MWCNT
100	3.527	3.523	3.518	0.25	0.04
200	3.531	3.524	3.521	0.28	0.08
400	3.540	3.526	3.522	0.50	0.11
600	3.555	3.531	3.525	0.84	0.16
800	3.611	3.540	3.528	2.29	0.33
1000	3.680	3.557	3.532	4.02	0.70

as compare to higher density of MWCNT due to the maximum utilization of bundle area.

6   Conclusions

In this research paper, primarily CNT interconnects have been demonstrated for the global interconnects lengths ranging from 100 to 1000 μm. Based on the number of conducting channels, different number of shells and arrangements of CNT in bundle an (ESC) transmission line model has been developed and used to analyze the effect of power dissipation and propagation delay. With the help of modeling and formulations of design parameters numerical calculations have carried out. DIL system is used to analyze the power dissipation and propagation delay performance. A SPICE simulation is used to compare the interconnect performance between SWCNT bundle and bundled MWCNTs. After the performance analysis it has been observed that the percentage reductions in propagation delay for bundle MWCNT having two MWCNTS as compares to conventional SWCNT bundle and MWCNT bundle are almost 77% and 13% respectively, where power dissipation for bundle MWCNT having two MWCNT has a small reductions compares to conventional SWCNT bundle and MWCNT bundle are almost 4.2% and 0.7% respectively at the global length of VLSI interconnects.

References

1. International technology roadmap for semiconductors. http://itrs.net/reports (2013)
2. Srivastava, N., Banerjee, K.: A comparative scaling analysis of metallic and carbon nanotube interconnections for nanometer scale VLSI technologies. In: Proceedings of the 21st International VLSI Multilevel Interconnect Conference (VMIC), 29 September–2 October, Waikoloa, HI, pp. 393–398 (2004)
3. Das, P.K., Yadav, A., Kumar, K.: Propagation delay analysis of multi-layered GNR and multi-walled CNT through-silicon via at different technology nodes. In: Proceedings of the World Congress on Engineering and Computer Science, WCECS 2017, October 2017, San Francisco, USA, vol. 1 (2017)
4. Wei, B.Q., Vaijai, R., Ajayan, P.M.: Reliability and current carrying capacity of carbon nanotubes. Appl. Phys. Lett. **79**(8), 1172–1174 (2001)
5. Berber, S., Kwon, Y.-K., Tomanek, D.: Unusually high thermal conductivity of carbon nanotubes. Phys. Rev. Lett. **84**(20), 4613–4616 (2000)
6. Collins, P.G., Hersam, M., Arnold, M., Martel, R., Avouris, P.: Current saturation and electrical breakdown in multiwalled carbon nanotubes. Phys. Rev. Lett. **86**(14), 3128–3131 (2001)
7. Li, H., Xu, C., Srivastava, N., Banerjee, K.: Carbon nanomaterials for next-generation interconnects and passives: physics, status and prospects. IEEE Trans. Electron Devices **56**(9), 1799–1821 (2006)
8. Pu, S.N., Yin, W.Y., Mao, J.-F., Liu, Q.-H.: Crosstalk prediction of single- and double-walled carbon-nanotube (SWCNT/DWCNT) bundle interconnects. IEEE Trans. Electron Devices **56**(4), 560–568 (2009)

9. Sathyakam, P.U., Mallick, P.S.: Transient analysis of mixed carbon nanotube bundle interconnects. IET Electron. Lett. **47**(20), 1134–1136 (2011)
10. Majumder, M.K., Pandya, N.D., Kaushik, B.K., Manhas, S.K.: Analysis of MWCNT and bundled SWCNT interconnects: impact on crosstalk and area. IEEE Electron Device Lett. **33**(8), 1180–1182 (2012)
11. Burke, P.J.: Luttinger liquid theory as a model of the gigahertz electrical properties of carbon nanotubes. IEEE Trans. Nanotechnol. **1**(3), 129–144 (2002)
12. Majumder, M.K., Das, P.K., Kaushik, B.K.: Delay and crosstalk reliability issue in mixed MWCNT bundle interconnect. 0026–2714/©2014 elsevier Ltd
13. Rossi, D., Cazeaux, J.M., Metra, C., Lombardi, F.: Modeling crosstalk effects in CNT bus architecture. IEEE Trans. Nanotechnol. **6**(2), 133–145 (2007)
14. Majumder, M.K., Kaushik, B.K., Manhas, S.K.: Analysis of mixed CNT bundle interconnects: impact on delay and power dissipation. In: 5th International conference on computers and devices for communication (CODEC) (2012)

Parasitic Series Resistance for 4H-SiC and Diamond-Based IMPATT Diode at Ku Band

Bhupendra Jothe, Girish Chandra Ghivela and **Joydeep Sengupta**

Abstract In this work, parasitic series resistance (R_s) has been computed for IMPATT using 4H-SiC and diamond at Ku band. R_s is calculated from the conductance- susceptance profile of the IMPATT. R_s of 4H-SiC was found to be less than diamond at the corresponding frequency. As 4H-SiC is having lesser R_s value than diamond, it can deliver more power to the output RF circuit as compared to diamond-based IMPATT diode.

Keywords 4H-SiC · Diamond · Series resistance · IMPATT · Conductance-susceptance

1 Introduction

The oscillation in IMPATT diode is due to its negative resistance. Crucial device parameter that limits the output power to RF circuit and efficiency of an IMPATT diode is its parasitic series resistance (R_s) [1]. In this paper author has concentrated on RF output power. As the negative resistance that develops in an IMPATT diode is very small thus the positive series resistance (R_s) should not cross the value of the negative resistance, otherwise the oscillations will cease or stop. The parasitic positive series resistance of an IMPATT diode come into existence due to the substrate, the metal semiconductor contact, undepleted epitaxial layer and a small part arise from device package. Thus an appropriate technology is required to reduce the value of

B. Jothe (✉)
Electronics and Telecommunication Engineering Department, Government Polytechnic College, Itarsi 461446, India
e-mail: bhupendra.jothe@gmail.com

G. C. Ghivela · J. Sengupta
Electronics and Communication Engineering Department, Visvesvaraya National Institute of Technology, Nagpur 440010, India

this positive resistance, so that the oscillation can sustain. Measuring R_s through network analyzer is difficult task because of circuit modeling complexity [2, 3]. So a computer-based method has been developed here. Electrical series resistance R_s is well recognized as limiting factor for power and efficiency in IMPATT diodes. In 1983 an effective method of measuring R_s had been proposed by Adlerestein et al. [4]. They used observation of the oscillation threshold bias current for a diode in a standard circuit. The material used was GaAs near 40 GHz. Similarly calculation of R_s for Si (Single Drift Region (SDR) Structure) IMPATT was carried out by Mitra et al. by considering unequal ionization rates and drift velocities of the two types of charge carriers in Silicon in1994 [5]. So lot many researchers have been experimentally measured the parasitic series resistance of IMPATT diodes operating at different frequencies. But the experimental measurement of series resistance is very much time-consuming and expensive method. That is why our aim is to develop a simple and generalized computer-based simulation method which can predict close value of parasitic resistance of particular structure of IMPATT diode operating at particular frequency. So we have to develop the numerical model of the parasitic resistance, which should be independent of operating frequency of the device, it means the model should be sustain at any operating frequency band. The numerical model is developed as described under Sect. 2, which is followed by DC and Small-signal analysis of IMPATT devices. Here, we have focus on the parasitic series resistance determination parts as the details of DC and small-signal analysis are described in our previous works [6–8]. The complete method of parasitic resistance determination can be applied to any IMPATT structure. The characteristics of microwave device and the contribution of diode negative resistance from the depletion layer of the diode can be obtained through the use of this method. As per recently reported articles [6–11], 4H-SiC and diamond are having the better performances as compared to other materials. Therefore, 4H-SiC and Diamond(c) based Double Drift Region (DDR) IMPATT diodes are designed here to oscillate in Ku-Band. This paper mainly describes the detailed computer-based method to calculate the parasitic positive series resistance R_s of packaged DDR IMPATT diodes from high frequency small-signal conductance-susceptance characteristics. This method would be suitable for determining R_s for Ku band diodes with appropriate values of ionization rate and drift velocity, at maximum electric field.

2 Numerical Estimation of Series Resistance

The circuit shown in Fig. 1 [12] below represents equivalent circuit for IMPATT diode, where G and B represent the diode conductance and susceptance respectively. We can determine the magnitudes of G and B from simulation by using values of
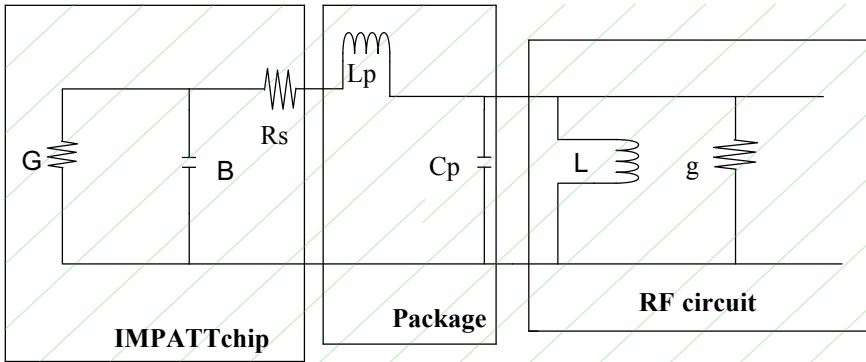
Fig. 1 Equivalent circuit for IMPATT diode

ionization rate and drift velocities of charge carriers for 4H-SiC and Diamond taken from [13, 14]. R_s is the series resistance of the device, g is the load conductance and L is the circuit inductance. C_p and L_p are the package capacitance and package inductance respectively. By using [15] Impedance of the packaged device is given,

$$Z_{dp} = \frac{1}{G + j(B - B_p)} + R_s \tag{1}$$

$$Z_{dp} = \frac{G + R_s G^2 + R_s (B - B_p)^2 - j(B - B_p)}{G^2 + (B - B_p)^2} \tag{2}$$

where $B_p = \frac{1}{\left(\omega L_p - \frac{1}{\omega C_p}\right)}$ = Effective susceptance caused by package parameter.

Impedance of the RF circuit is given by [15],

$$Z_{rf} = \frac{j\omega L\left(\frac{1}{g}\right)}{\frac{1}{g} + j\omega L} = \frac{\omega^2 L^2 g + j\omega L}{1 + \omega^2 L^2 g^2} \tag{3}$$

The oscillation condition is given by,

$$Z_{dp} + Z_{rf} = 0 \tag{4}$$

Both real and imaginary parts of Eq. (4) must be separately equal to zero,

$$\text{Real } (Z_{dp} + Z_{rf}) = 0 \tag{5}$$

$$\text{Imaginary } (Z_{dp} + Z_{rf}) = 0 \tag{6}$$

From above equation we get

$$\frac{G + R_s G^2 + R_s (B - B_p)^2}{G^2 + (B - B_p)^2} + \frac{\omega^2 L^2 g}{1 + \omega^2 L^2 g^2} = 0 \tag{7}$$

As the method proposed for IMPATT oscillator by Scharfetter and Gummel [16], where the authors have shown that with increasing voltage swing, the susceptance of the diode remains almost constant. Thus the frequency of oscillation does not change much with the build-up of amplitude of oscillation. At threshold, $g = 0$ [15]. So Eq. (7) reduces to

$$G_{th} + R_s G_{th}^2 + R_s (B_{th} - B_p)^2 = 0 \tag{8}$$

where G_{th} and B_{th} are conductance and susceptance of the diode at the threshold condition when the total conductance of the packaged diode becomes just negative. The equivalent circuit of the IMPATT diode is shown in Fig. 1; where L_p and C_p are package inductance and capacitance respectively. The expression for series resistance, R_s is [15]

$$R_s = \frac{|G_{th}|}{\left[G_{th}^2 + (B_{th} - B_p)^2\right]} \tag{9}$$

2.1 Material Parameters

The used parameters of the materials are listed in Table 1.

Table 1 Material parameters

Material parameters	4H-SiC	Diamond
A_n	4.57×10^{10}	1.935×10^8
A_p	5.13×10^8	1.935×10^8
b_n	5.24×10^9	7.749×10^6
b_p	1.57×10^9	7.749×10^6
V_{ns}	2.508×10^5	1.981×10^5
V_{ps}	2.5×10^5	1.829×10^5
ε_r	1	5.7
μ_p	120	0.18
μ_n	120	0.22
J_0	4×10^8	5×10^8

*A_n, b_n are ionization coefficient of electron and A_p, b_p for holes

* J_0 is the total current density

*V_{ns}, V_{ps} saturation velocity of electron & hole

*ε_r, is relative dielectric constant

*μ_n, μ_p are the electron and hole mobility

3 Results and Discussion

For calculation of the value of series resistance R_s, typical value of package inductance and capacitance are specified by manufacturer for calculation. In this paper the typical value used are $L_p = 0.080$ nH and $C_p = 0.110$ pF [15]. The change in parasitic series resistance with operating frequency in Ku band for 4H-SiC and diamond are shown in Tables 2 and 3 respectively. The susceptance-conductance profiles for 4H-SiC and diamond are shown in Figs. 2a,b respectively.

4 Conclusion

After observing the above result we can conclude that the series resistance for 4H-SiC is less as compared to the Diamond IMPATT diode, as R_s is the factor which will limits the output power from IMPATT diode to the RF circuit. So for maximum output power the value of R_s should be as small as possible. Therefore 4H-SiC IMPATT diode deliver more power to the output RF circuit as compare to Diamond IMPATT diode.

Table 2 Susceptance and parasitic series resistance values at different frequency for 4H-SiC and diamond

	f (GHz)	12	13	14	15	15.5	16	17	18
For 4H-SiC	$-B_p$ (mho)	8.725×10^{-3}	9.539×10^{-3}	1.037×10^{-2}	1.123×10^{-2}	1.168×10^{-2}	1.213×10^{-2}	1.305×10^{-2}	1.40×10^{-2}
	R_s (ohm)	0.0187	0.0203	0.0213	0.0225	0.023	0.023	0.0251	0.0264
For diamond	$-B_p$ (mho)	8.725×10^{-3}	9.539×10^{-3}	1.037×10^{-2}	1.123×10^{-2}	1.168×10^{-2}	1.213×10^{-2}	1.305×10^{-2}	1.40×10^{-2}
	R_s (ohm)	0.0391	0.0417	0.0440	0.0470	0.048	0.049	0.0522	0.05

Table 3 Conductance and susceptance values at different frequency for 4H-SiC and diamond

	f (GHz)	12	13	14	15	15.5	16	17	18
For 4H-SiC	$-G_{th}$ (mho)	29.683	27.442	25.512	23.832	23.072	22.35	21.054	19.893
	B_{th} (mho)	26.467	24.820	23.362	22.062	21.464	20.89	19.847	18.889
For diamond	$-G_{th}$ (mho)	14.428	13.343	12.408	11.594	11.225	10.87	10.246	9.682
	B_{th} (mho)	12.645	11.875	11.191	10.580	10.299	10.031	9.535	9.086
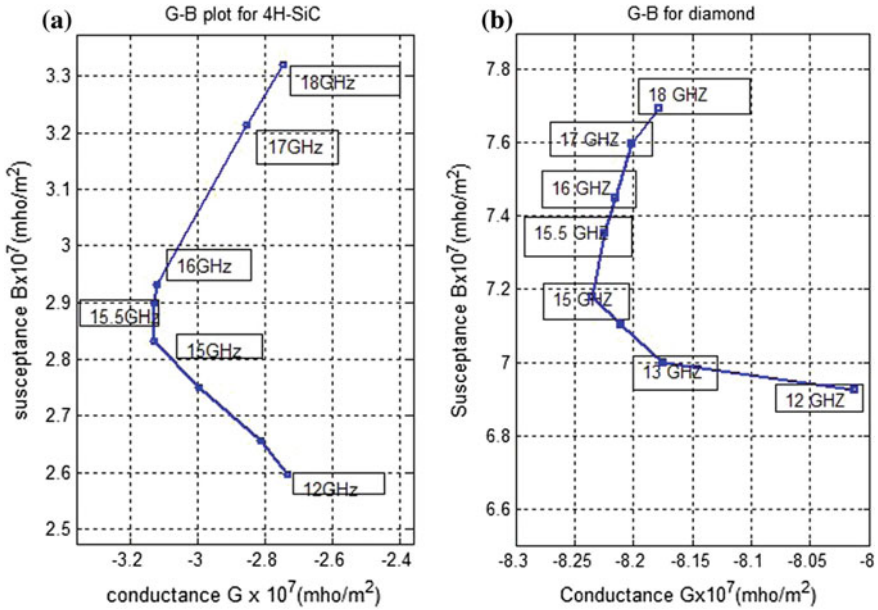
Fig. 2 Susceptance-conductance profile **a** for 4H-SiC and **b** for diamond

References

1. Midford, T.A., Bernick, R.L.: Millimeter wave CW IMPATT diodes and oscillators. IEEE Trans. Microw. Theory Tech. **27**, 483–492 (1979). https://doi.org/10.1109/TMTT.1979.1129653
2. Roy, U.C., Gupta, A.K.: Measurement of electrical series resistance of W-band Si IMPATT diode. In: Proceedings of 2nd Asia Pacific Microwave Conference, Beijing, China (1988)
3. Miswa, T.: Multiple uniform layer approximation in analysis of negative resistance in p-n junction in breakdown. IEEE Trans. Electron Devices **14**, 795 (1967). https://doi.org/10.1109/T-ED.1967.16113
4. Alderstein, M.G., Holway, L.F., Chu, S.L.G.: Measurement of series resistance in IMPATT diodes. IEEE Trans. Electron Devices **30**, 179–182 (1983). https://doi.org/10.1109/T-ED.1983.21092
5. Mitra, M., Das, M., Kar, S., Roy, S.K.: A study of the electrical series resistance of silicon IMPATT diodes. IEEE Trans. Electron Devices **40**, 1890–1893 (1993). https://doi.org/10.1109/16.277354
6. Ghivela, G.C., Sengupta, J., Mitra, M.: Ka band noise comparison for Si, Ge, GaAs, InP, WzGaN, 4H-SiC based IMPATT diode. Int. J. Electron. Lett. **7**, 107–116 (2019). https://doi.org/10.1080/21681724.2018.1460869
7. Ghivela, G.C., Sengupta, J.: Prospects of impact avalanche transit time diode based on chemical vapor deposited diamond substrate. J. Electron. Mater. **48**, 1044–1053 (2019). https://doi.org/10.1007/s11664-018-6821-5
8. Sengupta, J., Ghivela, G.C., Mitra, M.: Dynamic characterization and noise analysis of 4H-SiC IMPATT diode at Ka band. Int. J. Soft Comput. Eng. **4**, 145–149 (2014)
9. Sengupta, J., Ghivela, G.C., Gajbhiye, A., Mitra, M.: Measurement of noise and efficiency of 4H-SiC IMPATT diode at Ka band. Int. J. Electron. Lett. **4**, 134–140 (2016). https://doi.org/10.1080/21681724.2014.966774

10. Ghivela, G.C., Sengupta, J., Mitra, M.: Space charge effect of IMPATT diode using Si, Ge, GaAs, InP, WzGaN, 4H-SiC at Ka band. IETE J. Educ. **58**, 61–66 (2017). https://doi.org/10.1080/09747338.2017.1378132

11. Sengupta, J., Ghivela, G.C., Gajbhiye, A., Jothe, B., Mitra, M.: Temperature dependence of 4H-SiC IMPATT diode at Ka band. Int. J. Electr. Electron. Comput. Syst. **19**, 1–5 (2014)

12. Brooker, G.M.: Long-range imaging radar for autonomous navigation. Ph.D. dissertation, University of Sydney. Aerospace, Mechanical and Mechatronic Engineering, p. 231 (2006)

13. Trew, R.J., Yan, J.B., Mock, P.M.: The potential of diamond and SiC electronic devices for microwave and millimeter wave power applications. Proc. IEEE **79**, 598–620 (1991). https://doi.org/10.1109/5.90128

14. Watanabe, T., Teraji, T., Ito, T., Kamakura, Y., Taniguchi, K.: Monte Carlo simulations of electron transport properties of diamond in high electric fields using full band structure. J. Appl. Phys. **95**, 4866–4874 (2004). https://doi.org/10.1063/1.1682687

15. Pal, T.K.: Series resistance of silicon millimeter wave (Ka-Band) IMPATT diodes. Def. Sci. J. **59**, 1–5 (2009)

16. Scharfetter, D.L., Gummel, H.K.: Large-signal analysis of a silicon read diode oscillator. IEEE Trans. Electron Devices **16**, 64–77 (1969). https://doi.org/10.1109/T-ED.1969.16566

Design and Analysis of Self-biased OTA for Low-Power Applications

G. Manikanta, R. A. Mishra, N. A. Srivastava and R. K. Jaiswal

Abstract This paper presents an operational-transconductance-amplifier (OTA) for ultra-low power applications with high CMRR (common mode rejection ratio) and PSRR (power supply rejection ratio). The proposed OTA is a three-stage design. In order to attain the lower supply voltage and high CMRR, a bulk-driven differential pair with the tail current source has been considered as the first stage. The current mirror biasing technique makes sure that all the transistors operate in subthreshold region. A common source amplifier has been opted with current mirror as a load in second stage. At last, common source inverting amplifier is third stage of the designed OTA. The circuit has been designed and synthesized using cadence virtuoso simulator in 180 nm CMOS technology. It has been found that these stages are helpful in achieving high low-frequency gain. Hence, CMRR and PSRR also increase in significant amount. The results describe that the proposed design offers low-frequency gain of 58 dB with CMRR of 72 dB and PSRR of 56 dB for a supply voltage (V_{DD}) of 0.5 V. The proposed OTA provides the power dissipation of 1.8 μW at $V_{DD} = 0.5$ V. Also, the low-frequency gain of 57 dB, CMRR of 70 dB and PSRR of 55 dB with a power dissipation of 2.5 μW have been measured at $V_{DD} = 0.6$ V.

Keywords Bulk-driven MOS transistor · CMRR · OTA · PSRR

G. Manikanta · R. A. Mishra · N. A. Srivastava (✉) · R. K. Jaiswal
Motilal Nehru National Institute of Technology Allahabad, 211004 Prayagraj, India
e-mail: rel1702@mnnit.ac.in

G. Manikanta
e-mail: gmanikanta10@gmail.com

R. A. Mishra
e-mail: ramishra@mnnit.ac.in

R. K. Jaiswal
e-mail: rel1406@mnnit.ac.in

1 Introduction

In recent time, the electronics industry is running toward the design of low-power and high-performance integrated circuits (ICs) to accommodate the current market demand especially for portable devices. As per recent reports, the supply voltage will be scaled below 0.6 V for high-performance applications till 2025 [1]. These low-power circuits find their integration more rapidly in handheld devices for faster analog/digital computations. However, the scaling of supply voltage and transistor channel length has been continuously deteriorating the performance of analog circuit design parameters like gain, output voltage swing, CMRR, PSRR, etc. Several techniques have been reported to overcome these issues. Multistage cascading is one of the best methods to get high gain, and also bulk-driven transistors have been taken over gate driven transistors to achieve low power consumption. In recent literatures, various low-voltage design techniques had been proposed such as self-cascoding, bulk-biasing, level–shifting and floating-gate-transistors [2, 3].

An OTA is one of the most important design blocks in analog circuit and dominates with its application in low-power automatic gain control (AGC) amplifiers, high-speed data acquisition systems, etc. Various designs of OTA have been reported recently which operate supply voltage less than 1 V [4–6]. Moreover, low-power and high-performance OTA design is still one of the most challenging areas of research among analog industries [7–10].

In this paper, design and analysis of an ultra-low-power OTA has been presented that operates at supply voltage of less than 1 V. In which, a differential pair is used to achieve design requirements like high DC gain, high CMRR, and PSRR. Also, bulk-driven transistor has been considered to achieve the ultra-low-power requirements. Section 2 discusses the analysis of separate biasing OTA and the self-biased OTA design is explained in Sect. 3. The proposed design of self-biased OTA has been discussed in Sect. 4 for overall circuit analysis along with the experimental results. Section 5 concludes with the contribution of this paper.

2 Separate Biased OTA

The separate biased three-stage OTA is shown in Fig. 1 [11]. First stage is pseudo differential pair with common mode feed forward circuit [8]. The pseudo differential pair has been taken to attain the advantage of low power by reducing the required supply voltage. The minimum supply voltage required is $V_{DD_{min}} = V_{GS1} + V_{GS2}$. So, the power consumption will also reduce. The threshold voltage of both PMOS and NMOS is 0.3 V. Common source amplifier is the second stage with current mirror as load and common source amplifier with triode load is the third stage. These horizontal cascading using common source amplifiers will increases the overall gain of the OTA.
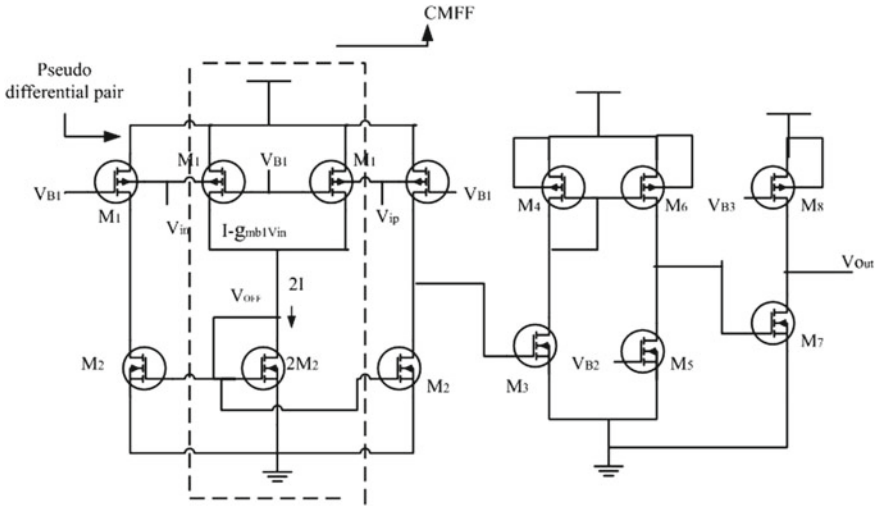
Fig. 1 Low voltage three-stage OTA [11] with V_{B1}, V_{B2}, V_{B3} as the external supply voltages

3 Self-biased OTA

The self-biased OTA is shown in Fig. 2 [11]. The external biasing voltages are removed and the voltages generated inside the circuit are used as biasing for the transistors in order to make a separate biasing OTA as self-biased OTA [7, 11]. The voltage generated at V_{OFF} is connected at V_{B1} to bias all M1 transistors and this V_{B1}
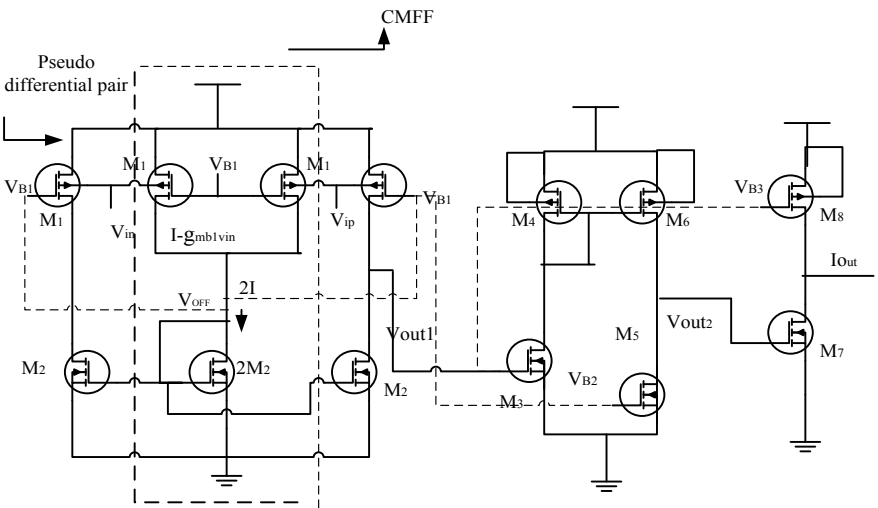
Fig. 2 Self-biased OTA [11]

is connected to V_{B2} to bias the M_5 transistor. The first stage output of is connected to the third stage to bias the M8 transistor. The biasing is indicated with dashed lines in Fig. 2. This type of biasing increases the CMRR, PSRR and also power consumption decreases by 25% approximately.

4 Proposed OTA Design

The proposed OTA is a three-stage circuit. In the first stage, one tail current source is connected to avoid the common mode deterioration in self-biased OTA. In which, common source amplifier is the second stage and the common source inverting amplifier is the last stage. These two stages are useful to get high DC gain.

4.1 Design of Input Stage

The problem in the self-biased OTA is less CMRR, PSRR due to the pseudo differential pair at the input stage. This Problem can be eliminated by connecting a tail current source [12] or a transistor in saturation region or by using current mirror circuit as presented in Fig. 3. The tail current source will allow the common current to flow through it and will result in increased CMRR, PSRR approximately by 50% and the design is practically good enough to use. The proposed input stage also increases

Fig. 3 Input stage of proposed OTA circuit
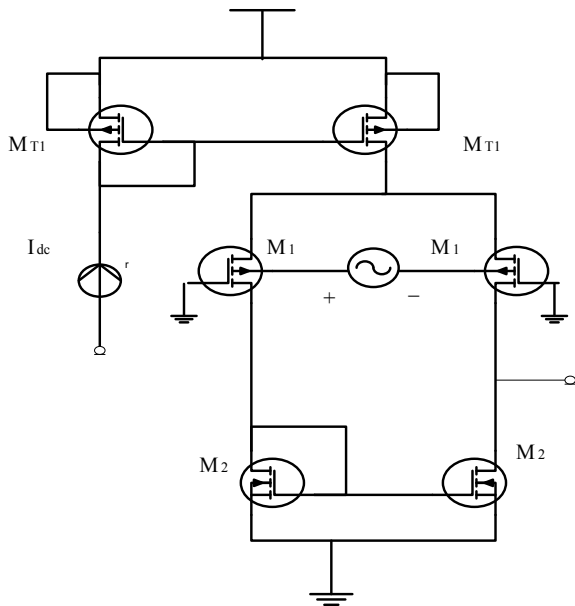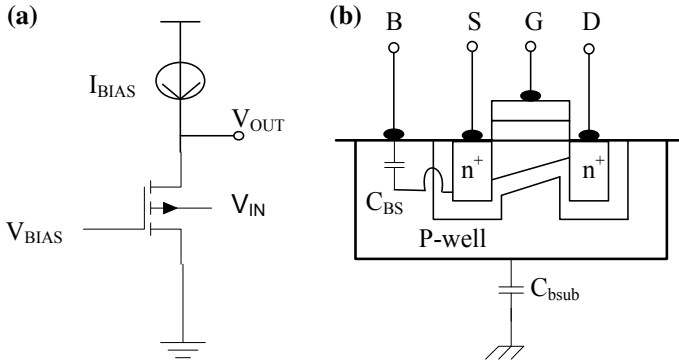
Fig. 4 Bulk-driven NMOS transostor [11]. **a** Circuit operation. **b** Cross-section

the unity gain bandwidth which is helpful in using the design over a wide range of frequencies.

4.2 Bulk-Driven MOS Transistor

The bulk-driven MOS transistor technique avoids the threshold voltage demand and hence, extends the admissible voltage of operation [3, 11]. The bulk-driven MOS will behave like junction field transistor operated in depletion region. Figure 4 shows the cross sectional view of bulk-driven MOS transistor. Although, it has the capability to cut-down the threshold voltage requirement, it also has several drawbacks. First, the relation between the transconductance of the bulk-driven transistor (g_{mb}) and the gate driven MOS transistor g_m is found as, $g_{mb} = ng_m$ (n = 0.2−0.4). Second, gate driven has better noise immunity over bulk-driven MOS and also access to the bulk is available only with the P-MOS.

4.3 Proposed Circuit and Discussion

The proposed circuit is a three-stage OTA which is shown in Fig. 5. The first stage is PMOS bulk-driven differential pair with tail current source which will improves the CMRR and PSRR. Common source amplifier is the second stage in which current mirror acts as a load [12]. The transistor M5 biased with help of current mirror of M2. Third stage is common source inverting amplifier formed by M7 and MT2, which is equivalent to a current source loading for third stage. M8 transistor source follower is biased with help of current mirror M7. The second and third stages are helpful to achieve high DC gain. The current mirror biasing is helpful in making subthreshold region operation for the transistor. So, all transistors in the proposed
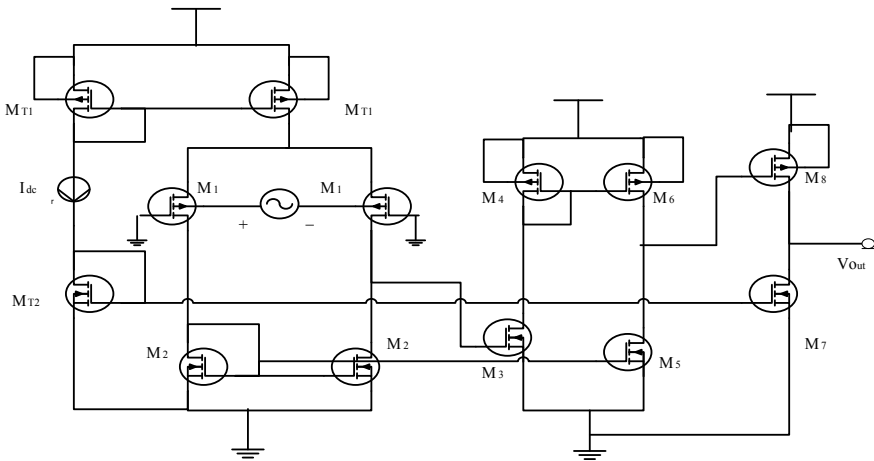
Fig. 5 The proposed three-stage Self-Biased OTA design

circuit will operate in subthreshold region. Since all are in subthreshold region, so less current will flow (practically in nano amperes) through the circuit and power supply is also less than 0.6 V. These will result to low power consumption.

Mathematical Analysis of Proposed OTA Design. Since the first stage is a differential amplifier, its gain is given by

$$A_v = -g_m R_D \tag{1}$$

Now the gain of the first stage is given by

$$A_v = -g_{mb}(r_{o1}||r_{o2}) \tag{2}$$

Common source amplifier is the second stage in which current mirror acts as a load, whose gain is $g_m(r_{o1}||r_{o2})$.

So the gain after the stage second is given by

$$A_{v2} = g_{mb2}(r_{o1}||r_{o2})g_{m3}(r_{o5}||r_{o6}) \tag{3}$$

The common source amplifier is the last stage with gain $g_{m8}(r_{o7}||r_{o8})$

The overall gain of differential mode is given by

$$A_{DM} = g_{mb2} \times g_{m3} \times g_{m8}(r_{o1}||r_{o2})(r_{o5}||r_{o6})(r_{o7}||r_{o8}) \tag{4}$$

The gain of the differential amplifier for common mode is given by $\frac{R_D/2}{g_m/2+R_{SS}}$

In the similar way the gain in common mode after the first stage is computed as

$$\frac{(r_{o1}||r_{o2})}{g_{mb2}/2 + R_T}$$

where R_T is resistance of the tail current source of the differential pair. The gain after second stage is $\frac{(r_{o1}||r_{o2})}{g_{mb2}/2+R_T} g_{m3}(r_{o5}||r_{o6})$

The overall common mode gain is given by

$$A_{CM} = \frac{g_{m3}g_{m8}}{g_{mb2}/2 + R_T}(r_{o1}||r_{o2})(r_{o5}||r_{o6}) \tag{5}$$

So, the CMRR can be found as

$$\text{CMRR} = \frac{A_{DM}}{A_{CM}} = g_{mb2}(g_{mb2}/2 + R_T)(r_{o7}||r_{o8}) \tag{6}$$

Since input is disconnected, $V_X = V_{DD}$ and $V_{O1} = V_{DD}$

Now, supply to output gain given by

$$A_{DD} = \frac{V_{DD}}{V_O} = \frac{V_{DD}}{V_{DD}} = 1 \tag{7}$$

The output voltage after first stage is V_{DD} which is input for the next stage

Output voltage after second stage is given by

$$V_{O2} = g_{m3}(r_{o5}||r_{o6})V_{O1} \tag{8}$$

Output voltage after third stage is given by

$$V_O = g_{m3}g_{m8}(r_{o5}||r_{o6})(r_{o7}||r_{o8})V_{DD} \tag{9}$$

So, supply to output gain equal to

$$A_{DD} = \frac{V_{DD}}{V_O} = g_{m3}g_{m8}(r_{o5}||r_{o6})(r_{o7}||r_{o8}) \tag{10}$$

From the gain of differential mode and gain of supply to output, the PSRR is calculated as

$$\text{PSRR} = g_{mb2}(r_{o1}||r_{o2}) \tag{11}$$

Measured Results. The proposed self-biased OTA shown in Fig. 5 has been designed and simulated on cadence tool in order to justify the novel features of the studied circuit. The OTA is designed on 180 nm technology and the parameters are listed in Table 1. The OTA was designed with a open loop gain of 58 dB and unity gain frequency of 5.8 MHz operating at 0.5 V supply. The OTA drives 1pF load capacitance. The analysis of open loop gain and the phase margin of proposed design are drawn

Table 1 Transistor dimension and parameters for $V_{dd} = 0.5$ V

	W (µm)/L (nm)	g_{ds} (µA/V)	g_m (µA/V)	I_D (nA)
M_1	10/180	0.288	9.51	449.2
M_2	10/180	0.139	11.9	449.2
M_3	10/180	2.760	9.32	357.4
M_4	20/180	0.238	7.86	357.4
M_5	10/180	2.670	9.34	357.4
M_6	20/180	0.238	7.86	357.4
M_7	15/180	0.292	24.4	989.5
M_8	23/180	0.129	21.82	989.5
M_{T1}	10/180	2.60	16.63	898.4
M_{T2}	10/180	0.287	24.65	1000

in Fig. 6 at supply voltage of 0.5 V. Also, the maximum output voltage swing of the studied design is shown in Fig. 7. Peak-to-peak output voltage is the maximum output voltage swing. The maximum output voltage swing is measured as 0.45 V.

Figure 8 shows the common mode gain of the proposed OTA at $V_{DD} = 0.5$ V. The common mode gain is measured as -14.6 dB. Figure 9 shows the gain from supply to output of the OTA at $V_{DD} = 0.5$ V. The gain from supply to the output of the OTA is measured as 2 dB. Further the dc current of the studied circuit is 3.603 µA at $V_{DD} = 0.5$ V. Hence, the proposed OTA consumes power of 1.8 µW. The proposed circuit provides better results at supply voltage of 0.6 V. The simulation results are listed in Table 2. The comparison of simulation results of different OTA topologies are mentioned in Table 3. One can clearly find that the proposed circuit is operating at low supply voltage and also consuming ultra-low-power hence the design useful for both low-voltage- and also low-power applications.

Fig. 6 Measured open loop frequency response of proposed OTA
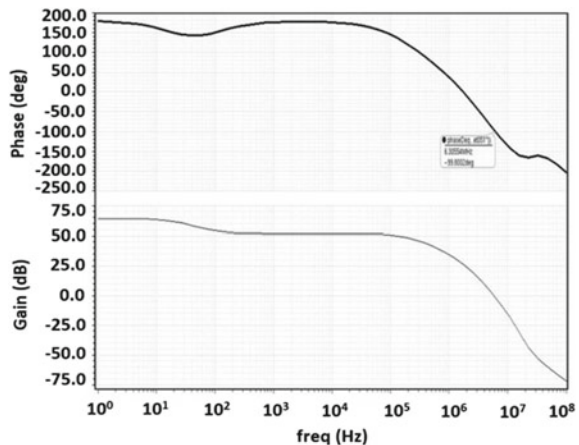
Fig. 7 Maximum output voltage swing of proposed OTA
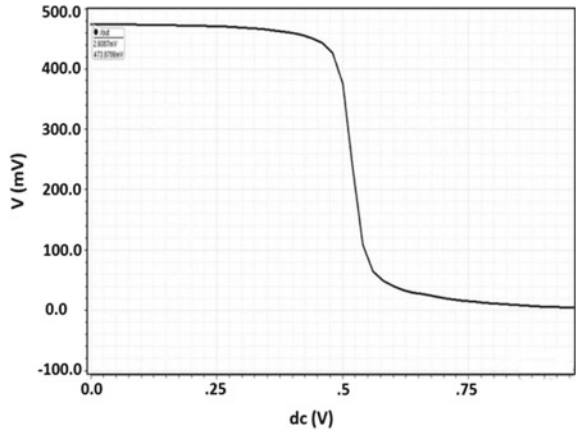
Fig. 8 Common mode gain (A_{CM}) of proposed OTA
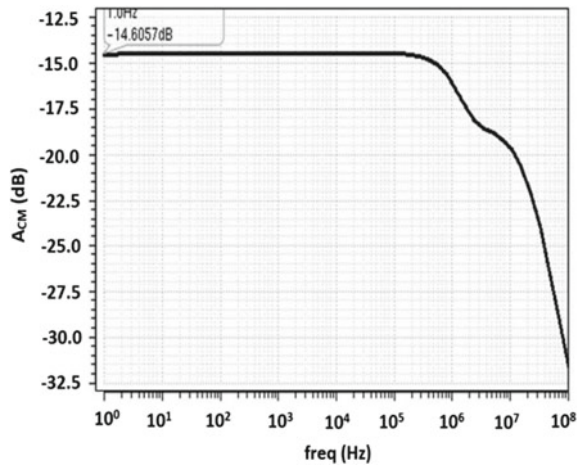
Fig. 9 Gain from supply to the output (A_{DD}) of proposed OTA
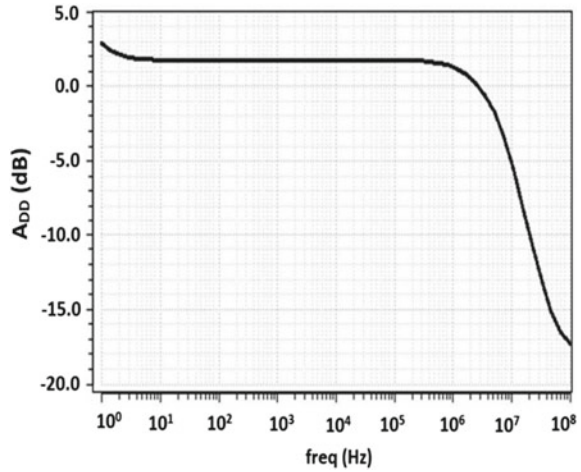
Table 2 Proposed OTA simulation results

Parameter	$V_{DD} = 0.5$ V	$V_{DD} = 0.6$ V
DC gain (dB)	58	57
Unity gain frequency (MHz)	5.8	5.9
Slew rate (v/μs)	2.5	2.4
Phase margin (°)	92	90
Output voltage swing (V)	0.45	0.45
Common mode gain (dB)	−14	−13
Differential mode gain (dB)	58	57
Gain from supply to output (dB)	2	2
CMRR (dB)	72	70
PSRR (dB)	56	55
Power (μW)	1.8	2.5

Table 3 Comparison of different OTA topologies

Parameter	Separate-bised OTA	Self-biased OTA	Proposed OTA
Slew rate (v/μs)	2.4	2.4	2.5
Unity gain frequency	–	690 kHz	5.8 MHz
Phase margin (°)	60	40	92
Output voltage swing (V)	0.45	0.45	0.45
CMRR (dB)	21	40	72
PSRR (dB)	−6	42	56
Power consumption (μW)	80	60	1.8

5 Conclusion

In this work, an OTA has been designed to tackle the challenges of ultra-low-voltage of operation in the today's IC technology. In which, the input stage utilizes differential pair with tail current source and bulk-driven MOS transistor to avail ultra-low-voltage of operation and also low-power consumption. The current mirror biasing technology is used here to eliminate external biasing, and also to bias the transistors to operate them in subthreshold region which is helpful in low power consumption. The OTA is

designed on 180 nm CMOS technology and also compared with the available state-of-art. The proposed OTA provides DC gain of 58 dB at $V_{DD} = 0.5$ V and 57 dB at $V_{DD} = 0.6$ V. The CMRR is measured as 72 dB at $V_{DD} = 0.5$ and 71 dB at $V_{DD} = 0.6$. The PSRR is measured as 56 dB at $V_{DD} = 0.5$ and 54 dB at $V_{DD} = 0.6$ V. Also, the power consumption of the proposed OTA is 1.8 µW at $V_{DD} = 0.5$ V and 2.4 µW at $V_{DD} = 0.6$ V. These attractive features of the proposed OTA make it a suitable candidate for low-power and high-performance applications.

References

1. International Technology Roadmap for Semiconductors (2011). http://www.itrs2.net/2011-itrs.html
2. Yan, S., Sanchez-Sinencio, E.: Low voltage analog circuit design techniques: a tutorial. IEICE Trans. Analog Integr. Circuits Syst. **E00-A**(2), 1–17 (2000)
3. Rajput, S.S., Jamuar, S.S.: Low voltage analog circuit design techniques. IEEE Circuits Syst. Mag. **2**(1), 24–42 (2002). 1531-636X/4/$10.00©2002IEEE
4. Blalock, B.J., Allen, P.E., Rincon-Mora, G.A.: Designing 1-V op amps using standard digital CMOS technology. IEEE Trans. Circuits Syst.-II: Analog Digit. Signal Process. **45**(7), 769–780 (1998). doi: 1057-7130(98)05054-X
5. Lin, T.-H., Wu, C.-K., Tsai, M.-C.: A 0.8-V 0.25mW current mirror OTA with 160-MHz GBW in 0.18-µm CMOS. IEEE Trans. Circuits Syst.-II: Exp. Briefs. **54**(2), 131–135 (2007). https://doi.org/10.1109/tcsii.2006.886465
6. Wang, J., Lee, T.-Y., Kim, D.-G., Matsuoka, T., Taniguchi K.: Design of a 0.5 V op-amp based on CMOS inverter using floating voltage sources. IEICE. Trans. Electron. **E-91-C**(8), 1375–1378 (2008). https://doi.org/10.1093/ietele/e91-c.8.1375
7. Rezzi, F., Baschirotto, A., Castello, R.: A 3 V 12–55 MHz BiCMOS pseudo-differential continuous–time filter. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **42**(11), 896–903 (1995)
8. Mohieldin, A.N., Sanchez-Sinencio, E., Silva-Martinez, J.: A fully balanced pseudo-differential OTA with common mode feedforward and inherent common-mode feedback detector. IEEE J. Solid-State Circuits **38**(4), 663–668 (2003). https://doi.org/10.1109/JSSC.2003.809520
9. Ferreiraand, L., Sonkusale, S.: A 60-dB gain OTA operating at 0.25 V power supply in 130 nm digital CMOS process. IEEE Trans. Circuits Syst.-I: Reg. Pap. **61**(6), 1609–1617 (2014). https://doi.org/10.1109/tcsi.2013.2289413
10. Ragheb, A.N., Kim, H.W.: Ultra-low power OTA based on bias recycling and subthreshold operation with phase margin enhancement. Microelectron. J. **60**, 94–101 (2017). https://doi.org/10.1016/j.mejo.2016.12.007
11. Abdelfattah, O., Roberts, G.W., Shih, I., Shih, Y.-C.: An ultra-low-voltage CMOS process-insensitive self-biased OTA with rail-to-rail input range. IEEE Trans. Circuits Syst. I: Regul. Pap. **62**(10), (2015). https://doi.org/10.1109/tcsi.2015.2469011
12. Razavi, B.: Design of analog CMOS Integrated Circuits, 1st edn. McGraw- Hill, New York (2001)

Work Function Estimation of Copper-Doped ZnO Thin Film

Lucky Agarwal, Brijesh Kumar Singh, Shweta Tripathi and P. Chakrabarti

Abstract In this article we report Copper-doped ZnO (CZO) based heterojunction diodes. The CZO thin film (p-type) has been grown over a n and p-Si substrate using sol-gelspin coating method. X-ray diffraction study spectacles that the deposited thin film show preferential grains orientation along the direction (101). The optical bandgap obtained using ellipsometer is 3.14 eV. The p-type conductivity of the CZO films has been established by hot point probe method. The electrical properties of prepared layers have been determined from the current vs voltage graph derived from semiconductor device analyzer. Finally, the work function of CZO layer has been estimated from the electrical parameter obtained from I–V calculations.

Keywords CZO · Work function · P-nature

1 Introduction

Thin film technology has attracted considerable attention in view of its tremendous application and simple fabrication technique. Thin film based devices are expected to have an innovator impression on the semiconductor industry. Among various II–VI group elements, Zinc oxide (ZnO) which is a wide bandgap material with large exciton binding energy is expected to emerge as a potential material for development of thin film devices [1–3]. ZnO will be ideally perfect for highly transparent conducting layer. Apart from that, it would be a good choice for near-UV detection,

L. Agarwal (✉) · B. K. Singh
Department of Electronics & Communication Engineering, Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India
e-mail: rel1408@mnnit.ac.in

S. Tripathi
Department of Electronics & Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India

P. Chakrabarti
Department of Electronics & Communication Engineering, Indian Institute of Engineering Sciences and Technolgy, Shibpur, West Bengal, India

biomedical applications, and nanosensors etc. [4]. All these properties craft ZnO thin film unique for applications in optoelectronics, photonics, and piezoelectric devices. Additionally, it is widely available, thermally stable, amenable to wet chemically etching, bio-compatible, good thermal conductor, highly resistant to radiations and easily processable.

The utmost important key to understand the electronic properties of ZnO is to control the doping that is also important to achieve p-type nature in ZnO. However, as grown ZnO always shows n-type conductivity due to self-compensating defects, like oxygen vacancy, zinc interstitials, and due to hydrogen incorporation [5]. To achieve n-type conductivity one just need to substitute the Zn site by a trivalent cation, which can be easily achieved by doping IIIrd group element [6]. However, to achieve p-type conductivity one must ensure to provide low acceptor level. Low solubility of the dopant as well as strong lattice relaxations are the only possibilities for deep energy level of the dopants [7]. Although many researchers have reported about p-type ZnO but their reliability, raise a big concern. Growth of p-type ZnO has been thought of in various ways including substitution of elements from Group IA or II B of periodic table on a Zn site [8, 9]. Cu is one of the important dopant to make ZnO p-type and is post-transition element that lies to the right of transition metals and to the left of the metalloids [10]. It is also interesting to note that Cu dopant can accumulate in the grain boundaries to form $ZnO–Cu_2O$ binary system due to which nonlinear I–V characteristic has been exhibited [11]. In this connection, using to the several attractive properties of CZO, we have previously analyzed influence of Cu concentration in ZnO in terms of nature, structural, optical, and electrical properties [12]. Further, using as the as obtained p-type CZO thin film, we have reported the fabrication of Pd-based schottky diode [13]. In the current report, p-type CZO based heterojunctions diodes were fabricated and work function of the CZO thin film was evaluated.

2 Experimental Details

The *p*-type and n-type Si (100) substrates of resistivity 5 and 0.4 Ω cm have been Piranahh ($H_2SO_4 + H_2O_2$) cleaned prior to the deposition of the thin ZnO layer to remove organic and ionic impurities respectively. The cleaned samples have been dipped in a buffered HF solution to remove the native oxide from the surface of the substrate.

The CZO thin films were deposited using sol–gel spin coating method on the cleaned Si-samples using precursor zinc acetate dehydrate ($Zn(CH_3COO)2.2H_2O$), isopropanol (solvent) and stabilizing agent diethanolamine (DEA). The molar proportion of zinc acetate dehydrates and stabilizer DEA was kept at 1:1. For Cu doping we add 5 molar percentage of copper acetate to the starting solution. The prepared solution was stirred at 60 °C for 1 h and then the prepared sol was reserved for at least 24 h at room temperature for aging. Further, CZO thin films have been deposited over n-type and p-type Si substrate by spin coating method. The films have been prepared

Fig. 1 Schematic outlook of the CZO—n/p-Si heterojunction device
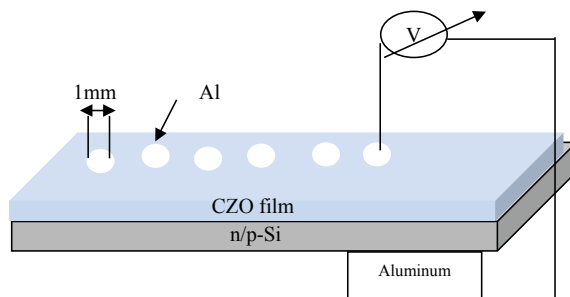
at spinning speed of 2000 rpm for 20 s, and prebaked at 100 °C for 10 min in oven to evaporate the solvent and organic residuals. In the last the films were annealed under air ambient in a muffle furnace.

The thickness of resulted CZO film as measured from ellipsometer is 200 nm on n and p-Si substrate. To offer low resistance electrical contacts to n-type and p-type Si sample a 250 nm thick aluminum contacts using shadow mask has been grown over the heterojunction between CZO and Si and a thin layer of aluminum is deposited on the back of silicon wafer. The resultant structure has been shown in Fig. 1. The electrical plot between current and voltage has been plotted using a semiconductor device analyzer. All the measurement was done under dark condition.

3 Result and Discussion

To investigate the crystalline properties of CZO samples, XRD measurements were carried out. Figure 2 shows the X-ray diffraction spectra of ZnO sample doped with 5 mol% Cu. The Bragg reflection which correspond to (002) peaks of CZO signifies that the orientation of prepared CZO samples is perfectly along the c-axis. XRD indicate that the deposited thin film is polycrystalline with hexagonal wurtzite

Fig. 2 XRD spectra of CZO thin film
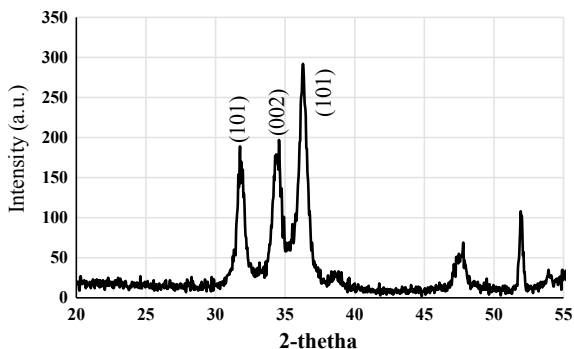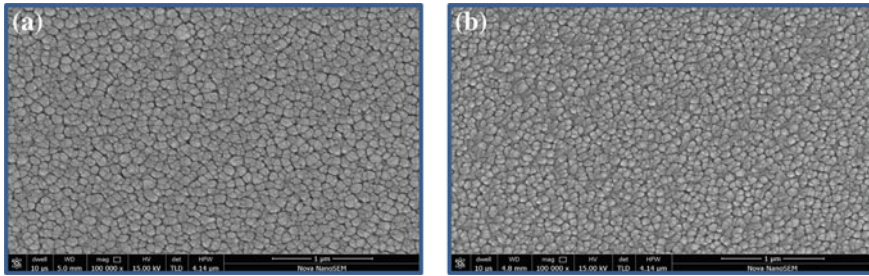
Fig. 3 SEM image of CZO thin film on **a** n and **b** p-Si substrate

structure. XRD spectra as shown in Fig. 2 represent the multiple peak among which the dominant peak (101) is located at $2\theta = 36.09°$. Another major orientation present at (100) and (002) are also observed in Fig. 2. Figure 3 shows the SEM image of CZO thin film on n and p-Si. The SEM analysis shows that the deposited CZO thin films are continuous without any pinholes.

The optical properties of CZO thin film were determined from the absorbance measurement in the range of 300–800 nm. The optical bandgap (E_g) of the CZO thin film has been determined considering the Tauc [14] relation between absorption coefficient and photon energy (hv) for direct transition between the edges of the valence and the conduction band. An extrapolation of the linear region of a plot of $(\alpha hv)^2$ on the y-Axis versus photon energy (hv) on the x-axis gives the value of the E_g as shown in Fig. 4. The E_g of CZO thin film has been found to be 3.42 eV.

To determine the nature of majority carriers in the deposited CZO thin films a very simple and efficient hot point probe method based on Seebeck effect has been performed. From the results, it can be clearly seen that CZO thin film shows p- type nature. To confirm the consistency of p-type nature of CZO thin films, the experiment was reiterated several time under same experimental condition and subsequently it has been found that outcome remains the same. The typical resistivity of the doped film was found to be nearly 0.1–0.65 Ω cm.

Fig. 4 Tauc Plot of CZO thin film for bandgap calculation
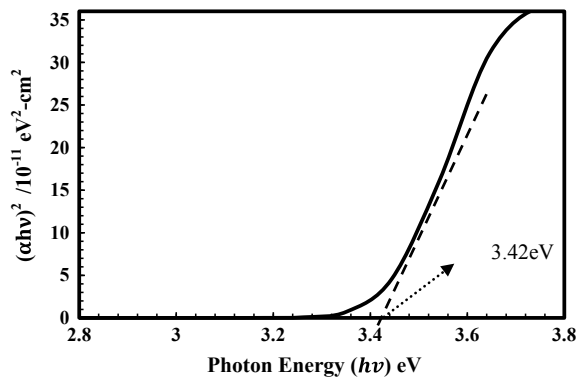
Fig. 5 Current (I)-Voltage (V) characteristic of the fabricated heterojunction
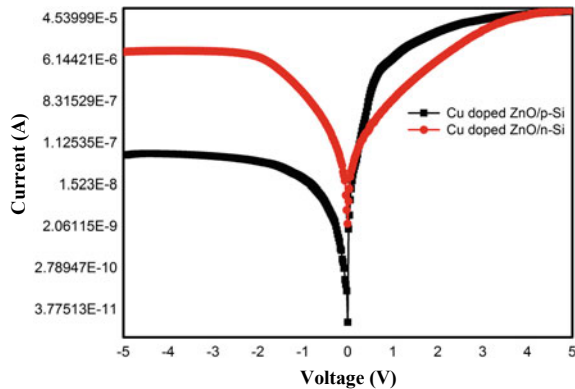
To draw the I–V characteristic of the deposited heterojunctions, the device has been placed under the microprobe station that has been connected to semiconductor device analyzer. The p-ZnO/n-Si and p-ZnO/p-Si show good rectifying characteristics. The forward biasing condition of fabricated heterojunction found to be very much similar to Schottky junction. The reverse saturation current (I_o) of the diode has been determined by extrapolating the reverse current of the log I–V plot. Since result shows high conductivity of the film therefore it has been concluded that the Fermi level might be coinciding with the valence band of doped ZnO. Following this result the electron affinity and work function of Zinc oxide film are assumed to be nearly equal. The semi-logarithmic I–V characteristics of Al/CZO/p-Si and Al/CZO/n-Si heterojunction have been own in Fig. 5. The relation between current and voltage can be given as [15]

$$I = I_0\left(e^{\frac{qV}{\eta kT}} - 1\right) \qquad (1)$$

where q is electronic charge on electron, η is ideality factor, T is the temperature in Kelvin, K is Boltzmann constant, V is applied voltage and I_o is reverse saturation current expressed as follow

$$I_0 = AA^*T^2 e^{\frac{-q\Phi_B}{kT}} \qquad (2)$$

where A is heterojunction diode area, A^* is the Richardson constant for ZnO. Φ_B in the above equation is the barrier height at zero bias which is evaluated by using the relation

$$\Phi_B = -\frac{kT}{q}\ln\left(\frac{I_0}{AA^*T^2}\right) \qquad (3)$$

From Fig. 4 and using Eq. (3) Φ_{bp1} of CZO/n-Si heterojunction was calculated as 0.57 eV. Similarly, Φ_{bp2} of CZO/p-Si heterojunction has been evaluated as 0.61 eV.

Based on the Schottky Mott model, the work function of ZnO has been calculated following the relation

$$\Phi_{ZnO} = E_g + X_{ZnO} - \Phi_{bp} \tag{4}$$

where E_g and X_{ZnO} are the bandgap and electron affinity of ZnO having values 3.42 and 4.35 eV. The work function for CZO/p-Si heterojunction and CZO/n-Si heterojunction has been found to be 7.2 eV and 7.16 eV respectively. The work function has been found to be very close to defect free ZnO single crystal [16]. Before the ZnO deposition, the Si wafer had been cleaned and due to this practice an inflexible thin (2–3 nm) insulating silicon oxide layer has been formed in the silicon surface. Due to the formation of such an insulating layer between silicon and ZnO barrier height get modified and has been evaluated by following Bardeen model which is given by [17]

$$\Phi_{bp} = C\left(E_g - \Phi_m + X_{si}\right) + (1 - C)(\Phi_0) \tag{5}$$

where

$$C = \frac{\varepsilon_i}{\varepsilon_i + q^2\delta D_s}$$

where δ is the thickness of oxide layer, D_s is the density of state, ε_i is total permittivity, Φ_o is the neutral Fermi level. Using the data obtained from the work reported by Turner and Rhoderick [18] showing metal/n-Si Schottky barrier having $qD_s = 2 \times 10^{12}/cm^3$ for n-Si wafer, $\delta = 1.5$ nm and $_i = 3.9$, $\Phi_o = 0.27$ eV and the work function of ZnO has assumed to be around 7.1 eV as calculated in this experiment. Φ_{bp} has been evaluated using Eq. (4) and has been obtained as 0.6 eV. Likewise extracting the data from the work done by Smith and Rhoderick [19] i.e. $qD_s = 3 \times 10^{12}/cm^3$ for p-Si wafer, and $\Phi_o = 0.33$ eV, Φ_{bp} has been evaluated using Eq. (5) and obtained as 0.54 eV. Hence, due to the presence of very thin oxidation layer between silicon and ZnO, Φ_{bp} has found to be decreased and Φ_{bp} shows increasing behavior as compared with the Schottky Mott model approach. At higher temperature depositing CZO causes, a formation of oxide layer simultaneously reduction in density of state due to which there is a decrease in Φ_{bp} for p-Si device and increase in Φ_{bp} for n-Si devices. It may be because at higher temperature oxygen diffusion toward Si interface increase which makes ZnO less stoichiometric due to larger oxygen scarcity. This oxygen induced defects in ZnO lead to reduction of the work function of ZnO which in due course decreases the barrier height as shown in Fig. 6.

Fig. 6 Variation of barrier height with substrate temperature
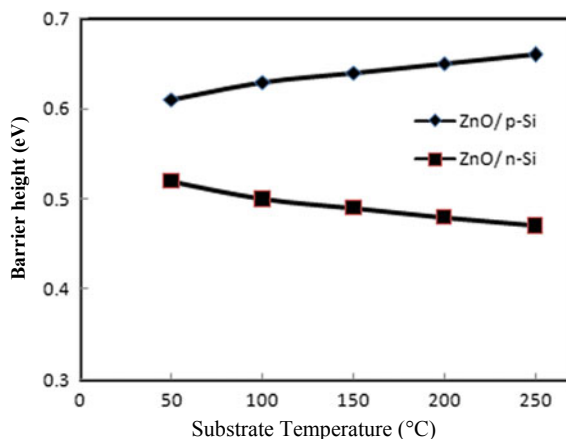
4 Conclusion

In this report p-ZnO-based heterojunctions were fabricated using spin coating to evaluate the work function of CZO thin film. The derived value of the work function based on the Schottky barrier model exhibits value between 7.26 and 7.16 eV for the CZO films. The effect of very thin oxide layer at the interface of heterojunction on the barrier heights has also been construed.

References

1. Hickernell, F.S., Trans, I.E.E.E.: Sonics Ultrason. **32**, 621 (1986)
2. Mang, A., Reimann, K.: Solid State Commun. **94**(4), 251–254 (1995)
3. Wang, Z.L.: J. Phys.: Condens. Matter **16**, R829–R858 (2004)
4. Gulia, S., Kakkar, R.: Adv. Mat. Lett **4**, 876 (2013)
5. Xu, J., Han, J., Zhang, Y., Sun, Y., Xie, B.: Sens. Actuators B Chem. **132**, 334 (2008)
6. S.B. Ogale, Springer Science & Business Media, (2006)
7. Roberts, B.K., Pakhomov, A.B., Shutthanandas, V.S., Krishnan, K.M.: J. Appl. Phys. **97**, 10D310 (2015)
8. Ip, K., Frazier, R.M., Heo, Y.W., Norton, D.P., Abernathy, C.R., Pearton, S.J.: J. Vac. Sci. Technol. **21**, 1476–1481 (2003)
9. Suja, M., Bashar, S.B., Morshed, M.M., Liu, J.: ACS Appl. Mater. Interfaces **7**(16), 8894–8899 (2015)
10. Park, C.H., Zhang, S.B., Wei, S.-H.: Phys. Rev. B **66**, 073202 (2002)
11. Bragg, W.H., Bragg, W.L.: The structure of some crystal as indicated by their diffraction of X-rays. Proc. R Soc. Lond. Ser. A Contain. Pap. Math. Phys. Charact. **89**, 248–277 (1913)

12. Agarwal, L., Tripathi, S., Chakrabarti, P.: Effects of copper concentration on structural and optical properties of copper doped ZnO prepared by sol-gel spin coating method. Mater. Focus **7**(1), 18–25 (2018)
13. Agarwal, L., Singh, B.K., Tripathi, S., Chakrabarti, P.: Fabrication and characterization of Pd/Cu doped ZnO/Si and Ni/Cu doped ZnO/Si Schottky diodes. Thin Solid Films **612**, 259–266 (2016)
14. Zaki, M.F.: Gamma-induced modification on optical bandgap of CR-39 SSNTD. J. Phys. D Appl. Phys. **38**, 558–562 (2008)
15. Pankove, J.I.: Optical Processes in Semiconductors, vol. 88. Prentice-Hall, New Jersey (1971)
16. Sze, S.M.: Physics of Semiconductor Devices, 2nd edn., pp. 256–270. Wiley, New York (1981)
17. Gopel, W., Brillson, L.J., Brucker, C.F.: J. Vac. Sci. Technol. **17**, 894 (1980)
18. Turner, M.J., Rhoderick, E.H.: Solid-State Electron. **11**, 291 (1968)
19. Smith, B.L., Rhoderick, E.H.: Solid-State Electron. **14**, 71 (1971)

Refractive Index and Dielectric Constant Evaluation of RF Sputtered Few Layer MoS$_2$ Thin Film

Richa Singh and Shweta Tripathi

Abstract In this paper thin film of MoS$_2$ has been deposited on pSi substrate at different time by RF sputtering method, X-ray diffraction pattern shows crystalline growth. Different optical parameters like refractive index, extinction coefficient and dielectric constant have been analyzed. It has been observed that the estimated optical parameters show the thickness depended behavior. Surface analysis has been done using FESEM that agrees with obtained optical parameters. All the optical parameters have been calculated using VASE Ellipsometer.

Keywords MoS$_2$ (molybdenum disulfide) · Sputtering · Ellipsometer

1 Introduction

2D (two dimensional) materials have gain considerable attention in recent year, researchers have been exploring 2D material like graphene, h-BN from past few decades, due to bandgap limitation of graphene and h-BN, these materials are not suitable for fabricating semiconductor devices, further in this context, a new family of 2D material known as Transition metal dichalcogenide (TMD) has been researched at lot, due to its immense potential in transparent electronics devices [1, 2]. In particular, MoS$_2$ is the most promising material of TMD group, which has direct band gap of 1.8 eV for monolayer and 1.2 eV for bulk material, extraordinary mechanical properties and optical properties. The most common polytype of MoS$_2$ is 2H-MoS$_2$, it has trigonal prismatic structure around one atom of Mo and two atom of sulfur (S), which is highly stable and almost inert in atmosphere [3, 4]. There are various techniques of depositing MoS$_2$ film, mechanical exfoliation is widely used deposition method in which a thin layer of MoS$_2$ is transferred on to the desired substrate using scotch tape, but controlling thickness and large area deposition are the main challenge in this method, which leads to development of other synthesis technique, physical

R. Singh (✉) · S. Tripathi
Department of Electronics & Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India
e-mail: singhricha51@gmail.com

vapor deposition (PVD) is most recent developed technique in this context, which allows large area coverage and thickness control up to some extent. Among various PVD (physical vapor deposition) techniques, sputtering is most commercially used method that provides highly quality film growth in a controlled environment [5, 6].

Semiconductor devices based on monolayer and few layer shows excellent optical properties, further, in order to understand optoelectronics devices based on MoS_2, optical properties like refractive index (n) and exciton coefficient (k) of MoS_2 should be analyzed [7]. Variable angle spectroscopic Ellipsometry (VASE) tool is used to calculate optical parameters, it provides wavelength depended n and k values by simply analyzing the change in of polarization of light, which is passed through the sample and developing an optical model of the sample [8]. In this work, thin film of MoS_2 up to few layer is deposited on pSilicon (100), via RF sputtering technique for different sputtering time (5, 10, and 15 min) and then optical parameters refractive index n, extinction coefficient k and dielectric constant are analyzed.

2   Experimental Details

Initially, three pSi substrates oriented along <100> cleaned, piranha cleaning is used to clean these substrates, and the first step is to rinse the substrate into DI water, followed by acetone repeat this process two times. After that samples are poured into the heated solution of H_2SO_4: H_2O_2 (3:1) for 20 min, then the substrate are dipped into acetone followed by DI water three times. Next the cleaned substrate are dried under nitrogen environment, further we place the samples in sputtering chamber, MoS_2 target (99.99%, sigma Aldrich.) is used under high vacuum pressure up to 10^{-6} mbar. Argon (Ar) gas is used to create plasma and ensure uniform deposition. Now three samples are deposited for 5, 10 and 15 min duration at the power of 50 W. Deposited samples are annealed at 300 °C for 2 h. Further Hot probe test is used on the samples to confirm the nature of the deposited film, all the films deposited on the samples shows n-type behavior.

3   Result and Discussion

3.1   Structural Analysis

Structural analysis of MoS_2 thin film deposited on crystalline pSi substrate with <100> orientation has been performed using X-ray Diffraction pattern (Smart Lab by Rigaku). The presence of few peaks in XRD pattern shown in Fig. 1 indicates crystalline nature of the deposited film. It can be seen from the graph that two peaks corresponding to (101) and (106) confirms the MoS_2 deposition. The observed inter planer distance (d) is 0.22 Å, which has been calculated by the equation given below

Fig. 1 XRD spectrum of thin film of MoS$_2$ deposited on pSi

$$n\lambda = 2dsin\theta \tag{1}$$
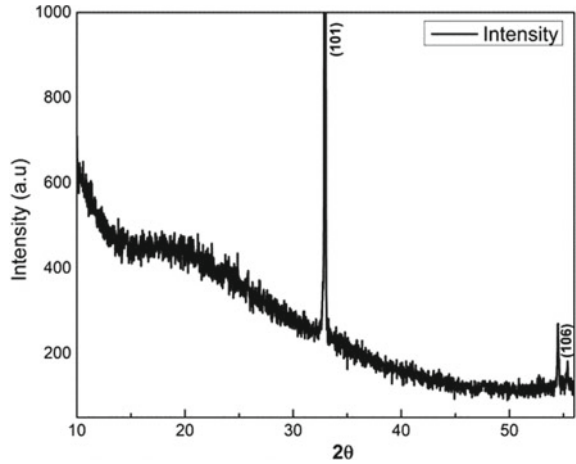
where λ is wavelength incident wave, n is positive integer.

Crystalline size (D) along the plane (101) is 34.57 nm, it has been calculated from the Scherrer formula

$$D = \frac{K\lambda}{\beta cos\theta} \tag{2}$$

where β is full width half maximum in rad, λ wavelength of X-ray, k is constant and Θ is the Bragg's angel.

3.2 Morphological Analysis

FESEM (Field Emission Scanning Electron Microscopy) image of deposited MoS$_2$ film on psilicon for different sputtering time is shown in Fig. 2. Morphological variations can be clearly seen with diverse sputtering time, as sputtering time change from 5 to 10 min more MoS$_2$ flakes appears, further deposition causes accumulation of MoS$_2$ cluster which forms vertically standing 3D layers (Fig. 2c). The vertically standing MoS$_2$ layers are chemically very active which enables it to use as a catalyst also it can be utilized in application of energy storing devices.
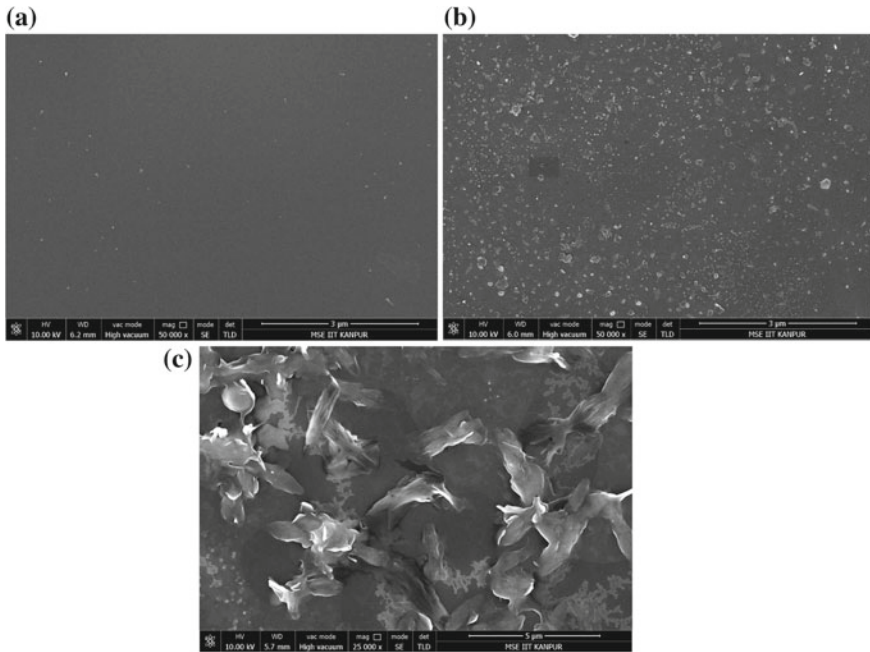
Fig. 2 FESEM image of thin film MoS2 deposited on pSilicon **a** 5 min **b** 10 min **c** 15 min

3.3 Morphological Analysis

Ellipsometer is predominantly a tool that detects the change in state of polarization of the reflected and incident layer. Based upon the difference in between the state of polarization various important parameters of the film can be determined like the thickness of the film, dielectric constant, refractive index, etc. It examines the amplitude (Ψ, psi) and phase difference (Δ, delta) of the reflected light that are related by the factor ρ, where $\rho = \tan(\Psi)\exp(i\Delta)$ [7]. In the present analysis of few layer MoS_2 film, variable angle Ellipsometry has been used to observe the optical characteristics. For obtaining the Ellipsometry results two layer model which consists of Si substrate and MoS_2 layer on it is used. Cauchy model has been used to calculate the optical parameter of MoS_2 film on the Si substrate [9]. Refractive index (n) of the material is very important property of the material, which can be utilized in fabrication of detectors, solar cell and sensors, etc. Refractive index is given by equation [10].

$$n = \sqrt{\frac{4R}{1 - R^2} - K^2} + \left(\frac{1 + R}{1 - R}\right) \tag{3}$$

where R is reflectivity of the deposited material and k is extinction coefficient. Figure 3 shows variation of refractive index of thin film deposited on pSi at different sputtering times (i.e. 5, 10 and 15 min). It can be seen that value of refractive index increases linearly with sputtering time. It means that the speed of light decreases as the number of layers of MoS_2 increases on the substrate in the visible range of electromagnetic spectrum.

k is known as extinction coefficient that tells the degree of absorption or absorption coefficient (α) [8].

$$\alpha = \frac{4\pi k}{\lambda} \qquad (4)$$

where k is extinction coefficient, λ is wavelength. It is shown in Fig. 4, as the deposition time increases, extinction coefficient increases. It is due to that fact that thickness of the film increases with sputtering time, thicker film will absorb more

Fig. 3 Variation of refractive index of MoS_2 film on pSi substrate at different sputtering time
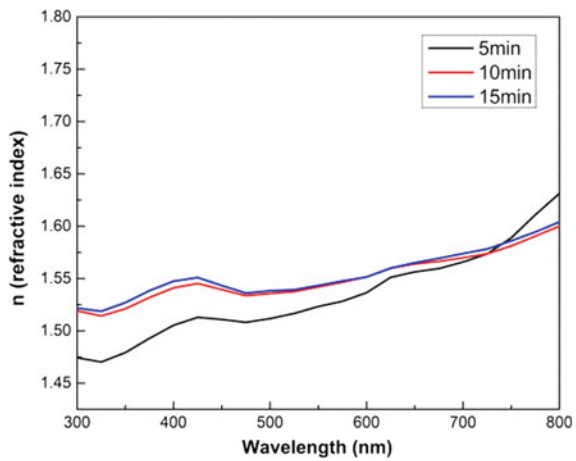
Fig. 4 variation of extinction coefficient of MoS_2 film deposited on pSi at different (5, 10, 15 min)
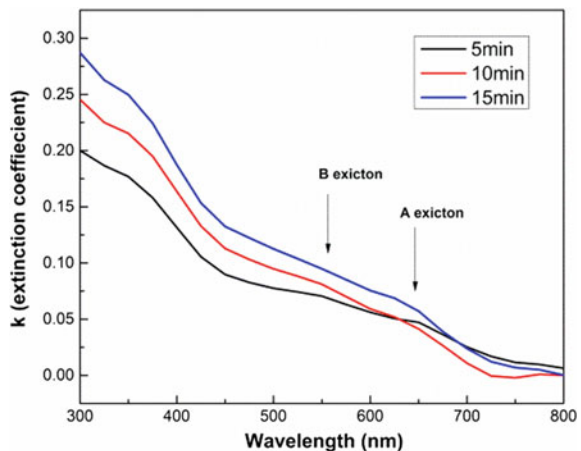
light. In addition to this A and B exciton can be seen at 649 and 548 nm, which falls into the brillouin zone [11].

Dielectric constant given by the equation [12]

$$\varepsilon = \varepsilon_{rel} + \varepsilon_{img} \tag{5}$$

where real and imaginary part is given by the equation [12]

$$\varepsilon_{rel} = n^2 - k^2 \tag{6}$$

$$\varepsilon_{img} = 2nk \tag{7}$$

Figures 5 and 6 the real value of dielectric constant that signifies the energy that can be stored by material and the imaginary part represents loss of energy. Figure 4 shows variation of dielectric constant for MoS$_2$ film at 5, 10, and 15 min with wavelength.

Fig. 5 Real parts of dielectric constant of MoS$_2$ thin layer on pSi substrate
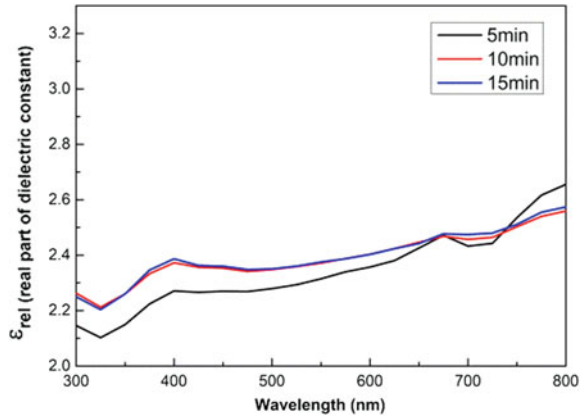
Fig. 6 Imaginary part of dielectric constant of MoS$_2$ film deposited on pSi substrate
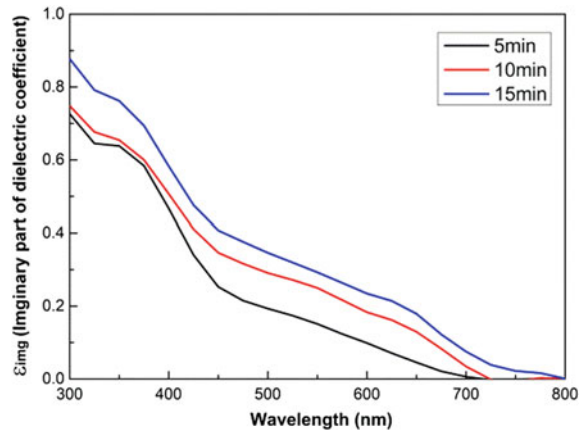
The real and imaginary value of dielectric constant shows opposite pattern, as the deposition time increases, real dielectric constant increases, which shows that energy storing capability of thin film MoS_2 increases with thickness as can be depicted from FESEM image Fig. 2c, it is due the fact that as the thickness of the film increases more number of charge will be trapped, giving high value of capacitance [13]. It can be seen from the Fig. 4 that the deposited film sustains the energy storing feature in visible region of electromagnetic spectrum (350–700 nm). On the other hand, imaginary parts of dielectric constant have been shown in Fig. 5, decreases in between the range of (350–700 nm) depicting that the losses are suppressed in visible range. It can be noted here, that the increasing thickness of deposited film demonstrates more energy losses.

4 Conclusion

In this paper, few layer of MoS_2 is deposited on single crystal pSi substrate for different growth time by RF sputtering method. The obtained XRD pattern confirms that the thin film of MoS_2 has been deposited on pSi substrate with good crystallinity and preferred peak location of (101). The deposited film shows that speed of light slows down with increasing thickness of film (5–15 min), In addition to this film absorb more light with increase in sputtering, it due to increase in thickness with sputtering time. With increase in growth time, thickness of the film increases which absorb more light. Further, energy storing capacity of film increases with thickness. So deposited film can be used for fabrication of sensors and energy storing devices like capacitors.

References

1. Benavente, E., Santa Ana, M.A., Mendizabal, F., Gonzalez, G.: Intercalation chemistry of molybdenum disulfide. Coord. Chem. Rev. **224**, 87–109 (2002)
2. Lee, C., Yan, H., Brus, L.E., Heinz, T.F., Hone, J., Ryu, S.: Anomalous lattice vibrations of single-and few-layer MoS2. ACS Nano **4**(5), 2695–2700 (2010)
3. Song, I., Park, C., Choi, H.C.: Synthesis and properties of molybdenum disulphide: from bulk to atomic layers. RSC Adv. **5**(10), 7495–7514 (2015)
4. Li, Y., Wang, H., Xie, L., Liang, Y., Hong, G., Dai, H.: MoS_2 nanoparticles grown on graphene: an advanced catalyst for the hydrogen evolution reaction. J. Am. Chem. Soc. **133**(19), 7296–7299 (2011)
5. Li, X., Zhu, H.: Two-dimensional MoS_2: properties, preparation, and applications. J. Materiomics **1**(1), 33–44 (2015)
6. Kim, H.S., Kumar, M.D., Kim, J., Lim, D.: Vertical growth of MoS_2 layers by sputtering method for efficient photoelectric application. Sens. Actuators A **269**, 355–362 (2018)
7. Bruzzese, D.G.: Use of spectroscopic ellipsometry and modeling in determining composition and thickness of barium strontium titanate thin films (2010)

8. Yim, C., O'Brien, M., McEvoy, N., Winters, S., Mirza, I., Lunney, J.G., Duesberg, G.S.: Investigation of the optical properties of MoS_2 thin films using spectroscopic ellipsometry. Appl. Phys. Lett. **104**(10), 103114 (2014)

9. Funke, S., Miller, B., Parzinger, E., Thiesen, P., Holleitner, A.W., Wurstbauer, U.: Imaging spectroscopic ellipsometry of MoS2. J. Phys. Condens. Matter **28**(38), 385301 (2016)

10. Di Giulio, M., Micocci, G., Rella, R., Siciliano, P., Tepore, A.: Optical absorption of tellurium suboxide thin films. Phys. Status Solidi (a) **136**, K101 (1993)

11. Xu, J., He, T., Chai, L., Qiao, L., Wang, P., Liu, W.: Growth and characteristics of self-assembled MoS 2/Mo-SC nanoperiod multilayers for enhanced tribological performance. Sci. Rep. **6**, 25378 (2016)

12. Kang, H.S., Ahn, B.D., Kim, J.H., Kim, G.H., Lim, S.H., Chang, H.W., Lee, S.Y.: Structural, electrical, and optical properties of p-type ZnO thin films with Ag dopant. Appl. Phys. Lett. **88**(20), 202108 (2006)

13. Kaindl, R., Bayer, B. C., Resel, R., Müller, T., Skakalova, V., Habler, G., Abart, R., Cherevan, A.S., Eder, D., Blatter, M., Fischer, F.: Growth, structure and stability of sputter-deposited MoS_2 thin films. Beilstein J. Nanotechnol. **8**, 1115 (2017)

Design and Optimization of MEMS Piezoelectric Cantilever for Vibration Energy Harvesting Application

Namrata Gupta, Abhishek Ray, Alok Naugarhiya and Abhinav Gupta

Abstract This paper presents a reverse trapezoidal unimorph piezoelectric cantilever structure for frequency tuning and power optimization. A proof mass is incorporated at the upper side of the free end of the cantilever. The proposed cantilever structure with nonconventional geometry finds its application in vibration energy harvesting. The design is simulated with COMSOL Multiphysics 5.2 to convert mechanical energy into electrical energy. The proposed harvester is able to generate $1300\,\mu$m displacement, 0.46V induced voltage, and $9\,\mu$W induced electrical peak power at the low frequency of about 160 Hz at applied 1 g acceleration and $12\,k\Omega$ optimal load resistance.

Keywords MEMS · Piezoelectric material · Cantilever · Sensors

1 Introduction

Microelectromechanical system (MEMS) technology has played a key role in microelectronic devices in order to develop miniaturized components and low-power operating systems. MEMS is a technology where microsensors and mechanical parts with processing circuit are integrated on a small piece of silicon wafer. These devices are capable to sense, control, and operate in the microscale and also generate output within the same range. The conventional battery is a traditional energy source for compact or wearable devices and remote system applications like wireless sensor network (WSN). But due to its limited lifetime and bulkier size, researchers are more interested to develop MEMS energy harvesters.

The energy harvesters scavenge energy from some renewable energy sources like solar, wind, thermal, etc. Apart from these resources, motion, vibration, or other

N. Gupta (✉) · A. Ray · A. Naugarhiya
National Institute of Technology, Raipur 492010, India
e-mail: ngupta.phd2017.etc@nitrr.ac.in

A. Gupta
Rajkiya Engineering College, Sonbhadra 231206, India

mechanical energy is also one of the interesting ambient energy sources for energy harvesting. The mechanical vibration sources with low frequency are easily available in the surrounding environment like household machines, human walk, fluid flow, mines, etc. The vibration energy harvesting system converts mechanical energy into electrical energy that can be used in low-power electronics system [6]. Mechanical to electrical energy conversion revolves around these three principles which are piezoelectricity, electromagnetic, and electrostatic. The piezoelectric transducers are more useful in energy harvesting applications due to its higher power density without using external supply at a wide range of frequencies [1, 5].

Researchers are interested to design MEMS sensors which have a compact size, low-power consumption, and compatible with CMOS IC technology as compared to conventional structures. Some research groups are using silicon compound and piezo resistor to design their structure but that is also not completely compatible with CMOS IC fabrication technology. The power density of piezoelectric energy harvesters using MEMS cantilever can be improved by either changing their configuration or by changing the geometry of the cantilever structure. Florenta et al. [3] have presented a nonconventional structure which is a trapezoidal cantilever with a tip mass made by silicon generating $2\,\mu W$ power. Asanet al. [4] have compared the performance of triangular, trapezoidal, and rectangular cantilever structure. The authors state that among them, the triangular structure gives the highest power $0.2\,\mu W$. Recently, Nan Chen et al. [2] have reported the multi-beam and trapezoidal structure for wideband vibration energy harvesting using PZT material. In this paper, trapezoidal shape generates 2.3–$2.5\,\mu W$ power with the dimension in mm range.

In this paper, we have proposed a trapezoidal cantilever which is reversed in structure (less width of the cantilever is fixed) using PZT material with tungsten proof mass on the free side of the cantilever. This structure has advantages like uniform strain, lower resonance frequency, higher peak output voltage, and wide bandwidth. The voltage, displacement, and power for the proposed structure have been calculated. Rest of the paper is outlined as, design of cantilever is described in Sect. 2, mathematical modeling and functioning are explained in Sect. 3. Section 4 represents the simulated results. At last, conclusion and acknowledgment is mentioned in Sects. 5 and 6, respectively.

2   Design of Cantilever Beam

The stiffness and sensitivity of the cantilever beam are basically depended upon the structure and property of the piezoelectric material used. The cantilever is fixed at one side and free at the other side to provide higher output. Piezoelectric cantilever structure consists of a beam, anchor, and proof mass.

Basically, it is a sandwich-type structure where the top layer is made by tungsten serves as an electrode for this type of harvester, the bottom layer of the beam is made by silicon, and the mid layer is formed by the active piezoelectric layer. Here, we use Lead Zirconate Titanate (PZT-5H) as piezoelectric active material. Figure 1 shows
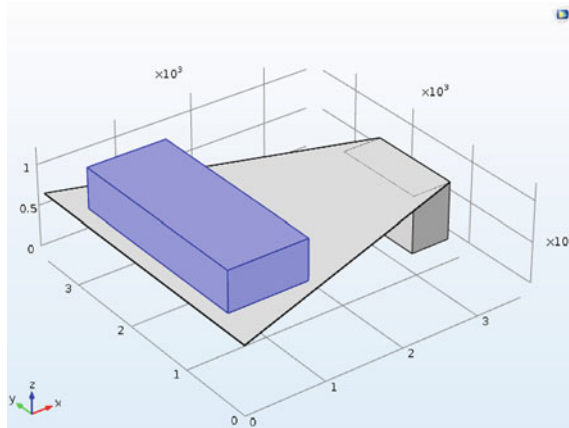
Fig. 1 Proposed cantilever based sensor

the 3D view of proposed nonconventional cantilever structure which is unimorph and reverse trapezoidal in the geometry. PZT has higher conversion efficiency as compared to another piezoelectric material. Table 1 shows the material property of the chosen material to design the proposed structure. We assume the thickness of silicon is much larger than the beam to reduce the complexity of the model. The anchor of a structure is made by SiO_2 and silicon material as a substrate. The dimensions of the structure are shown in Table 2. Cantilevers have different types of structures. Here, we use nonconventional structure to generate maximum power, making wider bandwidth and getting low resonance frequency. Meshing is done on any geometry

Table 1 Property of chosen material for designing the layers

Parameter	Si	PZT	Tungsten (W)	SiO_2
Young's modulus (GPa)	17	308	41.1	7
Density (kg/m^3)	2329	7500	19,350	2200
Poisson's ratio	0.28	0.31	0.28	0.17

Table 2 Dimension of cantilever structure

Material	Length (μm)	Width (μm)	Thickness (μm)
Si (Anchor)	500	1300	700
SiO2 (anchor)	500	1300	1
Metal	3600	p = 1300, q = 3600	1
Piezoelectric	3600	p = 1300, q = 3600	0.5
Silicon	3600	p = 1300, q = 3600	5
Metal (Mass)	1000	2500	500
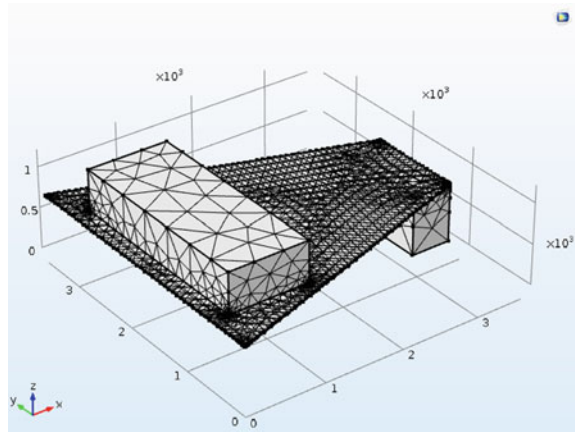
or structure so that the applied pressure, stress, or load is equally distributed on the structure so it will provide a better result. Here, we use normal fine meshing which is shown in Fig. 2.

3 Analytical Modeling and Functioning

To find out the parameters of piezoelectric material, there are no conventional methods or algorithms. There are two approaches, the first approach uses the theoretical relationship between differential surface stress and tip displacement of the cantilever structure during the applied pressure on the top of the structure. The relationship between the piezoelectric tip displacement and the induced output voltage can be calculated. In the second approach, finite element method (FEM) simulations by COMSOL Multiphysics which provide surface stress on the top of cantilever so that the tip deflection of the cantilever beam occurs due to external pressure or load [8]. There are two coupling modes of operation of the piezoelectric material known as 31 mode and 33 mode. In 33 mode, the external force is being applied perpendicular to the polling direction while in 31 mode, external force is applied on the same side as the polling direction. Energy harvester provides higher output when its resonance frequency matches with the system vibration frequency (which is input for the harvester) and this frequency can be tuned by placing the proof mass or tip mass on the structures. The fundamental frequency is given by Eq. 1

$$\omega_{sc} = \sqrt{\frac{k_{eq}}{m_{eq}}} \tag{1}$$

Fig. 3 Top view of trapezoidal structure without proof mass
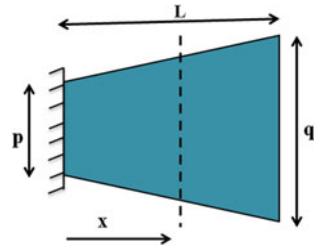
where ω_{sc} is the fundamental natural frequency, m_{eq} represents the total mass of cantilever beam which includes cantilever structure mass with proof mass, and k_{eq} is spring constant. Figure 3 shows the top view of trapezoidal cantilever structure without proof mass. The width of the beam at the free end is denoted by q and width of the beam at the fixed end is p. The distance from the fixed end of cantilever represented by x and the beam width is given by q(x). L represents the total length of the cantilever structure. So the section change rate is presented by $\delta = \frac{p-q}{L}$ [7].

Capacitance and resistance of the device can be calculated by Eqs. 2 and 3, respectively [7].

$$C = \frac{\epsilon_p(2p - \delta L)L}{2t} \tag{2}$$

$$R = \frac{2t}{\epsilon_p(2p - \delta L)L\omega_{sc}} \tag{3}$$

$$P(\delta) = \frac{V^2}{R} \tag{4}$$

$$P(\delta) = \frac{t\omega_{sc}^2\theta^2}{2\epsilon_p(2p - \delta L)L\omega_{sc}} \times$$
$$\frac{\mu m_{eq} a}{\sqrt{[-m_{eq}\omega_{sc}^2 + k_{eff} + \frac{t\theta^2}{\epsilon_p(2p-\delta L)L}]^2 + [\frac{\epsilon_p(2p-\delta L)L\omega_{sc}}{2t} + \frac{t\theta^2}{\epsilon_p(2p-\delta L)L}]^2}} \tag{5}$$

where t is termed as thickness of material, ϵ_p is dielectric constant of material, θ is electromechanical coupling factor, R is load resistance, C is capacitance, k_{eff} is total stiffness of structure, μ is correction factor, a is acceleration amplitude, V is peak output voltage, and P(δ) is power as function of change in section rate. The output peak power distribution over cantilever can be calculated by Eq. 5 [7].
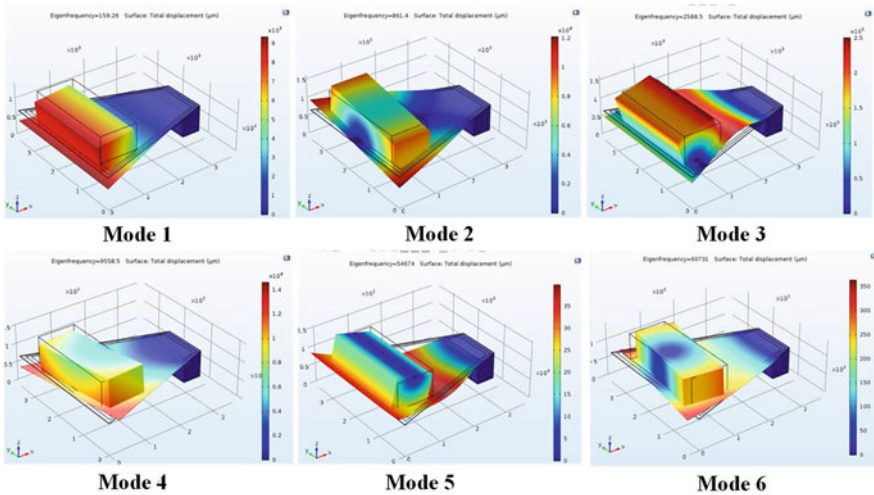
Fig. 4 Six mode of structure

4 Simulation Results

The shape of proof mass or load at the tip is more important for displacement, stress, and induced potential for any sensing device at the resonance frequency. Finite element analysis program is done with the help of COMSOL Multiphysics software. In Multiphysics software, we use MEMS module which is capable of MEMS designing, simulation, and modeling. In this software, we have used physics such as solid mechanics, piezoelectric, and electrical circuit. We have measured the displacement and stress at different eigenfrequencies and induced voltage, peak power distribution at the resonance frequency. An external load resistance $12\,\text{k}\Omega$ and $1\,\text{g}$ acceleration have been taken for the simulation. Figure 4 shows six modes of displacement of the structure. From the simulation results, we found that out of six modes of different eigenfrequencies, mode 1 has the maximum displacement, power, and induced voltage. Figure 5 shows the maximum tip deflection with a value $1300\,\mu\text{m}$ at $160\,\text{Hz}$ frequency. Figure 6a shows the voltage distribution over the cantilever and Fig. 6b shows the peak power distribution over the structure. The proposed harvester has provided $0.46\,\text{V}$ voltage and peak output power $9\,\mu\text{W}$ at $160\,\text{Hz}$ frequency. The outputs power is quite high with the low resonance frequency.

5 Conclusion

We have designed and simulated trapezoidal-shaped cantilever for scavenging energy at a frequency of $160\,\text{Hz}$ to feed low-power electronics devices by the principle of piezoelectricity. For an input acceleration of $1\,\text{g}$, the harvester scavenges $0.46\,\text{V}$
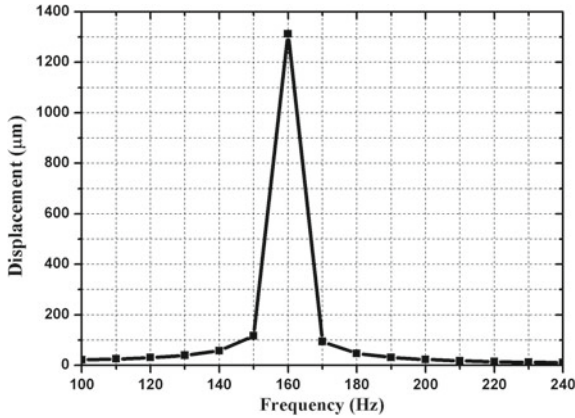
Fig. 5 Tip displacement versus frequency
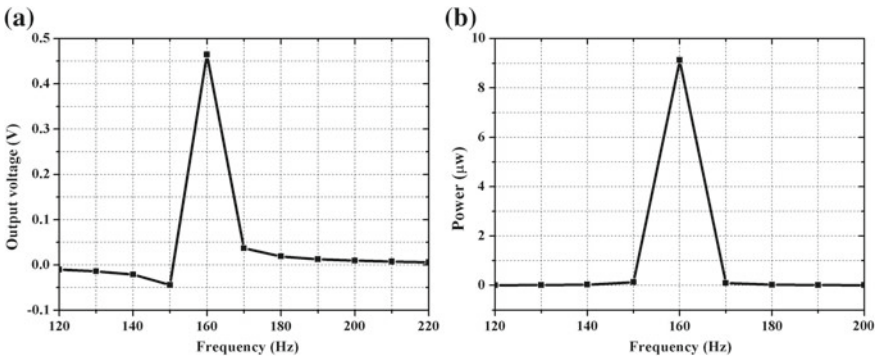
Fig. 6 a Voltage distribution versus frequency. **b** Peak Power distribution versus frequency

which is better than other reported trapezoidal-shaped harvesters. Moreover, the peak power of $9\,\mu$W is obtained when connected to a resistive load of $12\,k\Omega$. Based on several literature, it is found that the input acceleration and excitation frequency for the harvester should be known before designing the structures. Currently, many researches are truly focussing on increasing the bandwidth and ultra-low-frequency structures.

References

1. Ben Ayed, S., Abdelkefi, A., Najar, F., Hajj, M.R.: Design and performance of variable-shaped piezoelectric energy harvesters. J. Intell. Mater. Syst. Struct. **25**(2), 174–186 (2014)
2. Chen, N., Bedekar, V.: Modeling, simulation and optimization of piezoelectric bimorph transducer for broadband vibration energy harvesting in multi-beam and trapezoidal approach. J. Mater. Sci. Res. **7**(2), 26 (2018)
3. Costache, F., Pawlik, B., Rieck, A.: Development of a compact, low-frequency vibration, piezoelectric mems energy harvester. In: Multidisciplinary Digital Publishing Institute Proceedings, vol. 1, p. 588 (2017)
4. Muthalif, A.G., Nordin, N.D.: Optimal piezoelectric beam shape for single and broadband vibration energy harvesting: modeling, simulation and experimental results. Mech. Syst. Signal Process. **54**, 417–426 (2015)
5. Priya, S., Song, H.C., Zhou, Y., Varghese, R., Chopra, A., Kim, S.G., Kanno, I., Wu, L., Ha, D.S., Ryu, J., et al.: A review on piezoelectric energy harvesting: materials, methods, and circuits. Energy Harvest. Syst. **4**(1), 3–39 (2017)
6. Soliman, M., El-Saadany, E., Abdel-Rahman, E., Mansour, R.: Design and modeling of a wideband mems-based energy harvester with experimental verification. In: Microsystems and Nanoelectronics Research Conference. MNRC 2008, 1st edn. pp. 193–196. IEEE (2008)
7. Zhang, G., Gao, S., Liu, H., Niu, S.: A low frequency piezoelectric energy harvester with trapezoidal cantilever beam: theory and experiment. Microsyst. Technol. **23**(8), 3457–3466 (2017)
8. Zhou, W., Khaliq, A., Tang, Y., Ji, H., Selmic, R.R.: Simulation and design of piezoelectric microcantilever chemical sensors. Sens. Actuators A **125**(1), 69–75 (2005)

Analyzing the Sensitivity of Heterostructure of BP-Graphene/TMDC Layer Coated SPR Biosensor

Sarika Pal, Y. K. Prajapati and J. P. Saini

Abstract In this paper, the heterostructure of Au-BP-Graphene/TMDC layers coated SPR sensor is proposed for sensing of biomolecules. The proposed sensor is compared with conventional SPR and graphene based SPR biosensor in terms of sensitivity. Angular interrogation method is used for analyzing the sensitivity of proposed SPR biosensor at 633 nm operation wavelength. The sensitivity obtained for conventional SPR, Au-Graphene SPR and Au-BP-Graphene based SPR biosensors are 75.43°/RIU, 76.29°/RIU and 108.85°/RIU respectively. Highest sensitivity of 153.21°/RIU is obtained for two layer of WS_2 for heterostructure of Au-BP/TMDC-based SPR biosensor. The sensitivity increases for higher number of BP layers due to extraordinary sensing ability of BP. It is also observed that sensitivity increases with sensing layer RI due to better binding of biomolecules on sensor surface.

Keywords Black phosphorus (BP) · Graphene · Molybdenum disulfide (MoS_2) · Tungsten disulfide (WS_2) · Molybdenum diselenide ($MoSe_2$) · Tungsten diselenide (WSe_2) · Surface plasmon resonance (SPR)

1 Introduction

Surface plasmon resonance (SPR) has been prime research area especially in the field of biosensing for drug diagnostic [1], disease diagnostic [2], biomolecular sensing [3] and food safety, etc. [4]. SPR is one of the leading optical sensing techniques

S. Pal (✉)
Department of Electronics Engineering, NIT Uttarakhand, Srinagar, Garhwal 246174, Uttarakhand, India
e-mail: sarikapal@nituk.ac.in; narensarru@gmail.com

Y. K. Prajapati
Department of Electronics and Communication Engineering, MNNIT, Allahabad 211004, India
e-mail: yogendrapra@gmail.com

J. P. Saini
Netaji Subhas Institute of Technology (NSIT), New Delhi 110078, India
e-mail: jps_uptu@rediffmail.com

663

due to its unique features of real time, label-free and fast detection procedures [5]. Total attenuated reflection geometry (ATR), Kretschman configuration among various other configuration like Raether, Otto, etc., is preferred configuration due to its simple and realizable geometry [6, 7]. SPs propagating at metal/dielectric are charge density oscillations. Only p-polarized component of incident light can excite SPs and Resonance condition occurs on phase matching condition of horizontal component of incident wave vector with SPs wave vector indicating complete transfer of energy from evanescent wave to surface plasmon wave (SPW) [8]. The dip in reflectance is observed at SPR condition. SPR biosensor works on the principle of refractive index change in sensing medium on adsorbtion of biomolecules at sensor surface, which leads to redistribution of SPR electromagnetic field, i.e., shift in dip position of SPR curve [3]. Researchers used various methods for improving the performance of SPR sensor by using different material, metal, lower RI prism and high dielectric layer over metal layer, etc. [9]. Researchers utilized unique electrical and optical properties of graphene like high conductivity, high carrier mobility (10^5 cm^2 V^{-1} s^{-1}), high transparency, zero bandgap, lower resistivity, and high surface to volume ratio in SPR biosensor along with available advanced fabrication technique for developing SPR sensor with enhanced sensitivity [10]. High surface to volume ratio of 2D sp^2 bonded carbon atoms in graphene leads to efficient adsorbtion of biomolecules due to stronger van der Waals π interactions between its hexagonal cells and carbon based ring structure in biomolecule [11]. Recently, 2D-TMDC materials are widely used in photonics, electronics, and biomedical applications with rapid development in efficient fabrication techniques [12]. Monolayer of TMDC materials formed by sandwiching transition metal layer (Group IV elements—Mo, W, Ta, Nb) in between two layers of chalcogen material (Group VII elements—S, Se, Te) provides more optical confinement due to layering structure [12]. Till now, mostly graphene and TMDCs investigated as layered material graphene's zero band gap and higher band gap with relatively smaller mobility's in TMDC materials limits their performance [13]. So, BP a narrow band gap material possessing band gap in between graphene and TMDCs can be used in combination with these materials for enhancing light matter interactions in nanophotonic application [13]. BP a 2D material (sp^3 bonded phosphorus atoms in puckered honeycomb lattice structure) with remarkable electrical, optical and phonon properties like tunable direct band gap, very high carrier mobility, superior molecular adsorbtion energy, very high sensing ability and higher surface to volume ratio in comparison to graphene and MoS$_2$ widely used in nanoelectronics, optoelectronics and sensing applications [14, 15]. BP shows light adsorbtion in comparison to graphene and MoS$_2$ which is useful for binding of biomolecules on sensor surface [15]. Very recently S. Pal et al. improved the sensitivity of SPR biosensor 1.42 and 1.40 times of conventional and graphene-based SPR biosensor respectively by using BP [16]. Ouyang et al. analyzed the sensitivity and FWHM of SPR biosensor by using silicon nanosheets and TMDC layers (MoS$_2$, MoSe$_2$, WS$_2$, WSe$_2$) at different operating wavelengths [17]. The maximum sensitivity of 155.68°/RIU is achieved on using 35 nm thick gold layer, 7 nm Si layer with monolayer WS$_2$ at 600 nm wavelength. For further enhancing the sensitivity of the proposed SPR biosensor by utilizing light absorption properties of BP, we propose a heterostructure

of BP-Graphene/TMDC layers (MoS$_2$, MoSe$_2$, WS$_2$, WSe$_2$) coated over gold layer of SPR biosensor.

2 Proposed SPR Biosensor Design Considerations

This section describes the design considerations and N-layer modeling for reflectivity calculation for the proposed SPR biosensor.

2.1 Proposed Design for SPR Biosensor

The SPR biosensor in Fig. 1 consisting of SF-10 prism-Au layer-BP-Graphene/TMDC layers (MoS$_2$, MoSe$_2$, WS$_2$, WSe$_2$) operating at 633 nm wavelength. First SF-10 prism with RI 1.7230 is chosen on which Au layer with optimized thickness of 47 nm and RI $n_{Au} = 0.1726 + 3.4218 * i$ is deposited for SPs generation. Au is preferred over Ag for its chemical inertness and stable binding of biomolecules on sensor surface. BP layer with thickness 5 nm and RI $n_{BP} = 5.3 + 0.01 * i$ is deposited over Au for improvement in sensitivity of proposed SPR biosensor. For further improvement in sensitivity Graphene/TMDC layers of MoS$_2$, MoSe$_2$, WS$_2$, WSe$_2$ are deposited over BP to protect it from oxidation and for signal enhancement due to high charge transfer efficiency from TMDC layer to Au layer. The thicknesses and RIs for Graphene/TMDC layers are given in Table 1 shown below: The last layer is sensing layer having RI variation from 1.334 to 1.40, RI changes on adsorbtion of biomolecules on Graphene/TMDC layers. The TMDC layer in proposed biosensor increases its sensitivity by capturing biomolecules through van der Waals interactions [18].
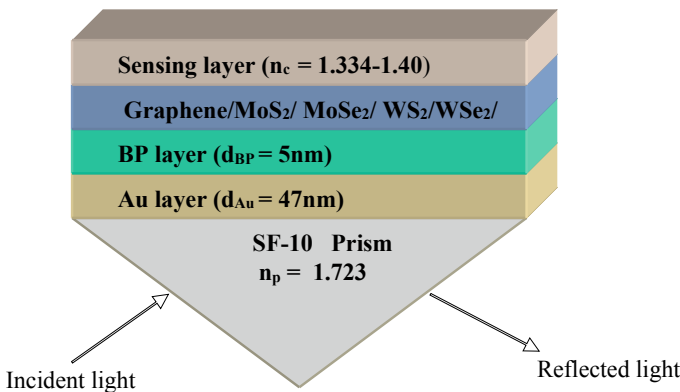
Fig. 1 Schematic diagram of SPR biosensor

Table 1 Thicknesses and refractive indices for graphene/TMDC layers [18]

S. No.	Type of 2D material	Single layer thickness (nm)	Refractive index (RI)
1.	Black phosphorus (BP)	0.53	$3.5 + 0.1 * i$
2.	Graphene	0.34	$3 + 1.49106 * i$
3.	MoS_2	0.65	$5.08 + 1.1723 * i$
4.	$MoSe_2$	0.7	$4.62 + 1.0063 * i$
5.	WS_2	0.8	$4.9 + 0.3124 * i$
6.	WSe_2	0.7	$4.55 + 0.4332 * i$

2.2 N-Layer Modeling

This subsection demonstrates the N-Layer modeling for reflectivity calculation of the proposed SPR biosensor. Transfer matrix method for N-layer modeling is used for calculating the reflectivity as it is efficient method for reflectivity calculation which do not consider any approximation [18]. The thickness, dielectric constants and refractive indices for the kth layer are denoted as d_k, ε_k, and n_k respectively along z-axis. On applying boundary conditions the tangential field at $Z = Z1 = 0$ are presented in terms of tangential field at $Z = Z_{N-1}$ as follows:

$$\begin{bmatrix} U_1 \\ V_1 \end{bmatrix} = M \begin{bmatrix} U_{N-1} \\ V_{N-1} \end{bmatrix} \tag{1}$$

where U_{N-1} and V_{N-1} are the corresponding electric and magnetic fields for the boundary at Nth layer and U_1 and V_1 is tangential components of respective fields respectively at the boundary of the first layer. The M_{ij} is characteristics matrix of the combined structure of the sensor, it can be given as

$$M_{ij} = \left(\prod_{k=2}^{N-1} M_k \right)_{ij} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{2}$$

with,

$$M_k = \begin{bmatrix} \cos \beta_k & (-i \sin \beta_k)/q_k \\ -i q_k \sin \beta_k & \cos \beta_k \end{bmatrix} \tag{3}$$

where

$$q_k = \left(\frac{\mu_k}{\varepsilon_k}\right)^{1/2} \cos\theta_k = \frac{\left(\varepsilon_k - n_1^2 \sin^2\theta_1\right)^{1/2}}{\varepsilon_k} \tag{4}$$

$$\beta_k = \frac{2\pi}{\lambda} n_k \cos\theta_k (z_k - z_{k-1}) = \frac{2\pi d_k}{\lambda}\left(\varepsilon_k - n_1^2 \sin^2\theta_1\right)^{1/2} \tag{5}$$

After going through straightforward steps, we obtain the reflection coefficient for p-polarized light which is given below

$$r_p = \frac{(M_{11} + M_{12}q_N)q_1 - (M_{21} + M_{22}q_N)}{(M_{11} + M_{12}q_N)q_1 + (M_{21} + M_{22}q_N)} \tag{6}$$

The reflectivity (R_p) for multilayer configuration is given as

$$R_p = \left|r_p\right|^2 \tag{7}$$

Sensitivity (S) of the SPR sensor is ratio of change in resonance angle ($\Delta\theta_{res}$) with respect to change in RI of the sensing layer (Δn_s). Δn_s considered here for the proposed SPR biosensor is 0.07.

$$S = \Delta\theta_{res}/\Delta n_s \tag{8}$$

3   Simulation and Results

In this paper, first gold thickness is optimized at 47 nm for proposed SPR biosensor at which minimum reflectance is obtained. Then SPR curve for conventional SPR, Au-Graphene SPR and Au-BP-Graphene SPR biosensor is plotted as shown in Fig. 2. The resonance angle obtained before and after adsorption of biomolecules for conventional SPR, Au-Graphene SPR and Au-BP-Graphene SPR sensor are (56.513°, 61.792°), (56.723°, 62.067°) and (62.811°, 70.425°) respectively. Thus the change in resonance angle obtained for conventional SPR, Au-Graphene SPR sensor and Au-BP-Graphene-based SPR biosensor are 5.28°, 5.34° and 7.614° respectively.

On calculating sensitivity from Eq. (8) are 75.43°/RIU, 76.29°/RIU, and 108.85°/RIU respectively for conventional SPR, Au-Graphene SPR and Au-BP-Graphene based SPR biosensor. Higher sensitivity obtained for Au-BP-Graphene based SPR biosensor is due to stronger binding between hexagonal cells of graphene and carbon based ring structure in biomolecules and better sensing ability of BP [18]. Figure 3 shown below indicates the SPR curve plotted for heterostructure of Au-BP-Graphene/TMDC layers, i.e., proposed SPR biosensor. The resonance angle for proposed SPR biosensor with single layer graphene, TMDC layers (MoS_2, $MoSe_2$, WS_2, WSe_2) before and after adsorbtion of biomolecules are (62.811°, 64.641°, 64.444°, 65.193°, 64.458°) and (70.425°, 73.047°, 72.781°, 73.984°, 72.858°)
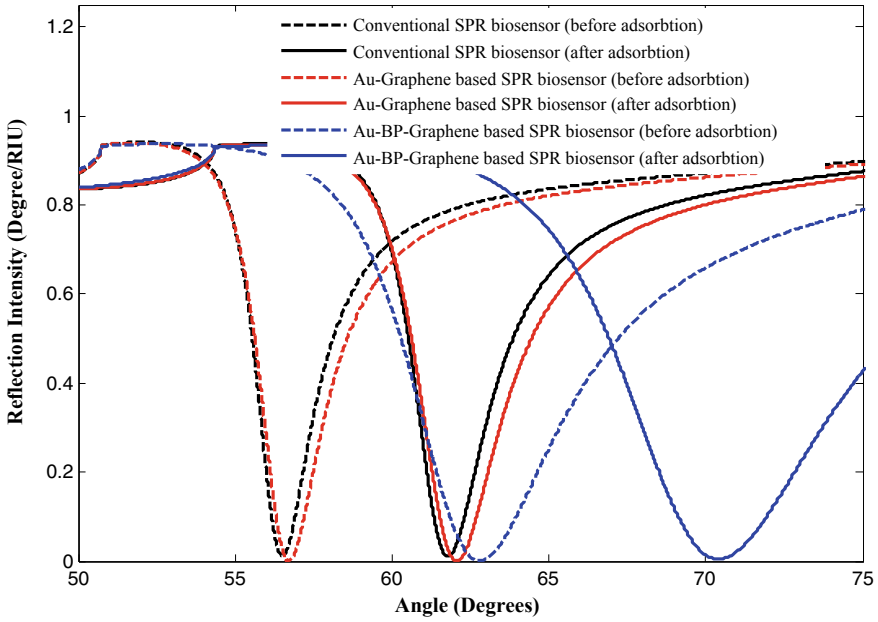
Fig. 2 SPR curve: conventional SPR, Au-graphene SPR and Au-BP-graphene based SPR biosensor before and adsorbtion of biomolecules
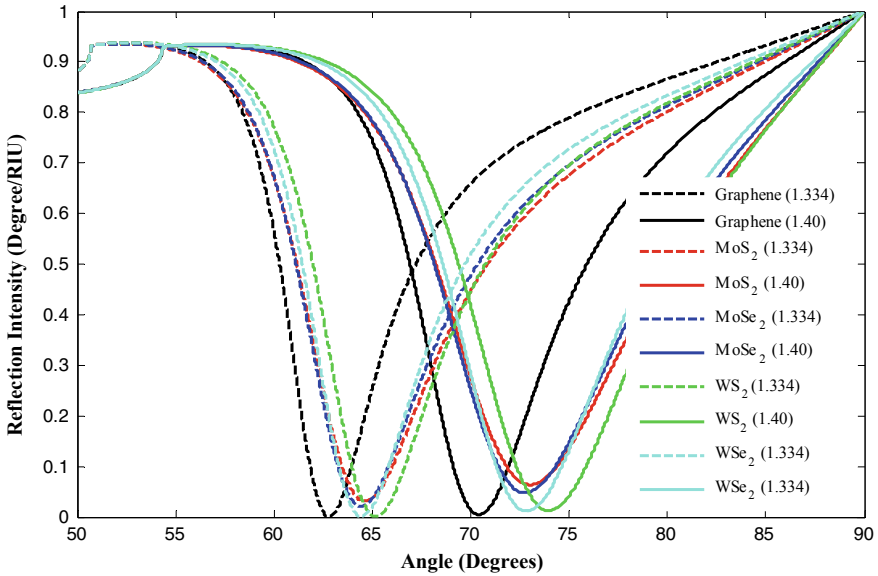
Fig. 3 SPR curve for proposed SPR biosensor with single layer of graphene, MoS2, MoSe2, WS2, and WSe2 layer

respectively. The shift in resonance angle for proposed SPR biosensor with single layer graphene, TMDC layers (MoS_2, $MoSe_2$, WS_2, WSe_2) before and after adsorbtion of biomolecules are 7.61°, 8.41°, 8.34°, 8.79°, and 8.40°, respectively. The sensitivities obtained for proposed SPR biosensor with single layer graphene, TMDC layers (MoS_2, $MoSe_2$, WS_2, WSe_2) are 108.71°/RIU, 120.14°/RIU, 119.14°/RIU, 125.57°/RIU, and 120°/RIU. Thus, higher sensitivities are obtained for BP/TMDC layers interface as TMDC provides more confinement of charge carriers at BP/TMDC interface and enhanced light matter interaction.

Thereafter, Sensitivity is optimized for different number of graphene, MoS_2, $MoSe_2$, WS_2, and WSe_2 layers for the proposed SPR biosensor. Figure 4 indicates variation of sensitivity of proposed SPR biosensor for different number of graphene, MoS_2, $MoSe_2$, WS_2, and WSe_2 layers. The maximum sensitivities obtained for proposed SPR biosensor for different number of Graphene/MoS_2 layer/$MoSe_2$ layer/WS_2 layer/WSe_2 layer are summarized in Table 2. The highest sensitivity of 153.21°/RIU is obtained for two layers of WS_2 for proposed SPR biosensor which is due to higher penetration of evanescent field into sensing layer because of its high value of real part of RI. It is observed that sensitivity increases with sensing layer refractive index and highest sensitivity is obtained for WS_2 layer (Fig. 5).

Thereafter, sensitivity for proposed SPR biosensor for different number of BP layers is obtained in Fig. 6. It is observed that sensitivity increases with increase in number of BP layers and highest sensitivity of about 177.54°/RIU is obtained for 13 layers of BP for BP/WS_2 TMDC interface. Higher sensitivity for more number of BP layers is due to enhanced light matter interaction at interface.

Fig. 4 Sensitivity versus number of graphene/MoS_2/$MoSe_2$/WS_2 and WSe_2 layer for proposed SPR biosensor
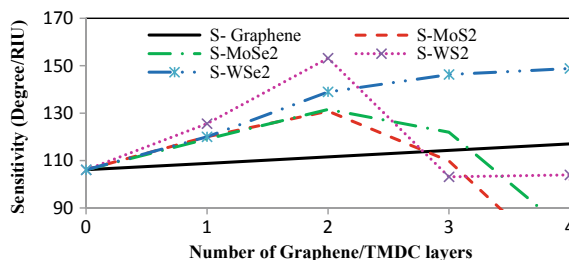
Table 2 Sensitivities obtained for different number of graphene/TMDC layers

Graphene/type of TMDC layer	Single layer thickness (nm)	No. of optimized graphene/TMDC layers	Sensitivity (°/RIU)
Graphene	0.34	4	116.98
MoS_2	0.65	2	130.84
$MoSe_2$	0.7	2	131.58
WS_2	0.8	2	153.21
WSe_2	0.7	4	148.86

Fig. 5 Sensitivity versus sensing layer RI for proposed sensor using graphene/MoS$_2$/MoSe$_2$/WS$_2$/WSe$_2$ layer
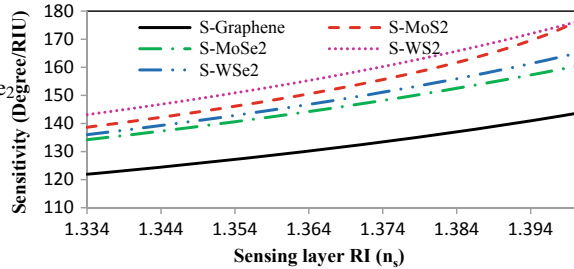
Fig. 6 Sensitivity versus BP layers for proposed SPR sensor using different combination of graphene/TMDC layers
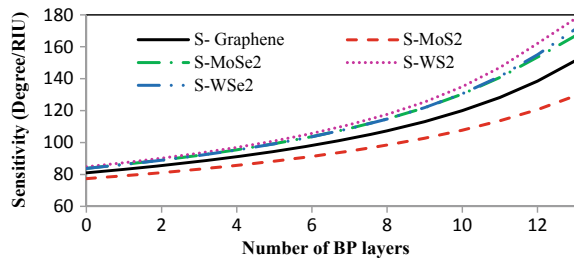
4   Conclusion

This paper analyzes the sensitivity of heterostructure of Au-BP-Graphene/TMDC interface based SPR biosensor for detection of biomolecules through angular interrogation method at 633 nm operating wavelength. The highest sensitivity of 153.21°/RIU for proposed SPR biosensor is obtained for two layers of WS$_2$. It is concluded that sensitivity increases on increasing BP layers due to enhanced light matter interaction at BP/TMDC interface. The sensitivity of proposed SPR sensor is better than conventional SPR- and graphene-based SPR biosensor proposed in literature. The proposed SPR biosensor can be extensively used in biosensing applications for realizing it using existing fabrication technology.

References

1. Lukosz, W.: Integrated optical chemical and direct biochemical sensors. Int. J. Sens. Actuators B **29**, 3750–1975
2. Zeng, S., Baillargeat, D., Ho, H.P., Yong, K.T.: Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. Chem. Soc. Rev. **43**, 3426–3452 (2014)
3. Pal, S., Prajapati, Y.K., Saini, J.P., Singh, V.: Sensitivity enhancement of metamaterial based SPR biosensor for near infrared. Opt. Appl. **46**(1), 131–134 (2016)
4. Ladd, J., Taylor, A., Jiang, S.: SPR biosensors for food safety. In: Homola, J. (eds.) Surface Plasmon Resonance Based Sensors. Springer Series on Chemical Sensors and Biosensors, p. 4. Springer, Berlin, Heidelberg (2006)

5. Pal, S., Prajapati, Y.K., Saini, J.P., Singh, V.: Resolution enhancement of optical SPR sensor using metamaterial. Int. J. Photonic Sens. **5**(4), 330–338 (2015)

6. Kretschmann, E., Reather, H.: Radiative decay of non-radiative surface plasmons excited by light. Int. J. Zeitschrift für Naturforschung **23**, 2135–2136 (1968)

7. Otto, A.: Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. Z. Phys. **216**, 398–410 (1968)

8. Kowalczyk, S.W., Tuijtel, M.W., Donkers, S.P., Dekker, C.: Unrevealing single stranded DNA in solid state nanopore. Nano Lett. **10**, 1414–1420 (2010)

9. Shalabney, A., Abdulhalim, I.: Sensitivity-enhancement methods for surface plasmon sensors. Laser Photon. Rev. **5**, 571–606 (2011)

10. Karimi, H., Yusof, R., Rahmani, R., Hosseinpour, H., Ahmadi, M.T.: Development of solution-gated graphene transistor model for biosensors. Nanoscale Res. Lett. **9**(1), 71 (2014). https://doi.org/10.1186/1556-276X-9-71

11. Song, B., Li, D., Qi, W., Elstner, M., Fan, C., Fang, H.: Graphene on Au(111): a highly conductive material with excellent adsorption properties for high-resolution bio/nanodetection and identification. ChemPhysChem **11**, 585–589 (2010)

12. Wang, Q.H., Kalantar-Zadeh, K., Kis, A., Coleman, J.N., Strano, M.S.: Electronics and opto-electronics of two-dimensional transition metal dichalcogenides. Nat. Nanotechnol. **7**, 699–712 (2012)

13. Xia, F., Wang, H., Jia, Y.: Rediscovering black phosphorus as an anisotropic layered material for optoelectronics and electronics. Nat. Commun. **5**, 4458 (2014). https://doi.org/10.1038/ncomms5458

14. Cai, Y., Zhang, G., Zhang, Y.W.: Layer-dependent band alignment and work function of few-layer phosphorene. Sci. Rep. **4**, 6677–6682 (2014)

15. Cho, S.Y., Lee, Y., Koh, H.J., Jung, H., Kim, J.S., Yoo, H.W., Kim, J., Jung, H.T.: Superior chemical sensing performance of black phosphorus: comparison with MoS_2 and graphene. Adv. Mater. **28**(32), 7020–7028 (2016)

16. Pal, S., Verma, A., Prajapati, Y.K., Saini, J.P.: Influence of black phosphorous on performance of surface plasmon resonance biosensor. Int. J. Opt. Quantum Electron. (OQEL) **49**(403), 1–13 (2017) (Springer Publication)

17. Ouyang, Q., Zeng, S., Li, J., Hong, L., Xu, G., Dinh, X.Q., Qian, J., He, S., Qu, J., Coquet, P., Yong, K.T.: Sensitivity enhancement of transition metal dichalcogenides/silicon nanostructure-based surface plasmon resonance biosensor. Sci. Rep. **6** (2016)

18. Pal, S., Verma, A., Saini, J.P., Prajapati, Y.K.: Sensitivity Enhancement Using Silicon-Black Phosphorus-TDMC Coated Surface Plasmon Resonance Biosensor, IET Optoelectronics, vol. 13, pp. 1–7 (2019)

Short Channel Effects (SCEs) Based Comparative Study of Double-Gate (DG) and Gate-All-Around (GAA) FinFET Structures for Nanoscale Applications

Vadthiya Narendar, Richa Parihar and Ashutosh Kumar Pandey

Abstract The FinFET device architecture is a leading contender in the field of electronic integrated circuits (ICs). A comparative study of double-gate (DG) and gate-all-around (GAA) FinFET structures on the ground of performance parameters like on-state current (I_{ON}), off-current (I_{OFF}), subthreshold swing (SS), drain induced barrier lowering (DIBL) and threshold voltage (V_t) have been done. It has been found that GAA FinFETs are able to control the short channel effects (SCEs) more accurately over DG FinFETs when fin width to gate length ratio is properly optimized. Further, the comparison between rectangular and cylindrical channel GAA FinFETs has also done and found that cylindrical GAA FinFETs gives a better performance with respect to aforementioned parameters. All the numerical simulated results were performed on TCAD supported the stated findings.

Keywords Cylindrical gate-all-around (GAA) FinFETs · Drain induced barrier lowering (DIBL) · Double-gate (DG) FinFETs · Off-current (I_{OFF}) · On-current (I_{ON}) · Subthreshold swing (SS)

V. Narendar (✉)
Electronics and Communication Engineering Department, National Institute of Technology Warangal, Telangana, India
e-mail: narendarv@nitw.ac.in

R. Parihar
Electronics and Communication Engineering Department, Institute of Engineering and Technology, Lucknow, Uttar Pradesh, India
e-mail: richa1parihar@gmail.com

A. K. Pandey
Electronics and Communication Engineering Department,
Motilal Nehru National Institute of Technology, Prayagraj, Uttar Pradesh, India
e-mail: chandan270697@gmail.com

1 Introduction

Since six decades, the miniaturization of microelectronic devices have become a prime driving force with same pace of technology to provide better performance along with reduced power consumption and area. Initially, the Planar bulk MOSFETs were used and served its purpose for three decades, but as the technology advances based on scaling their performance hampered greatly by short channel effects (SCEs) [1]. As an alternative solution, the silicon on insulator (SOI) MOSFETs were introduced and known for their higher current drive per unit silicon area and better short channel immunity [2]. Some other features of SOI MOSFETs [3, 4] includes: (1) improved isolation because of which latch up problems can be totally eliminated, (2) they show reduced parasitic capacitances which results in faster switching and reduced power dissipation, (3) they possess enhanced radiation hardness; and (4) higher validation tolerance which makes them suitable for space application, (5) Lastly because of lower leakage, the source and drain junctions of these devices can be operated at higher temperature. Another device named double-gate (DG) MOSFET which is a modified structure SOI MOSFET [4]. DG MOSFETs are promising candidates in sub-micrometer regime and offers various advantages such as, it has two channel, a better gate control over the channel region and more immune to SCEs. Also the device permits lower doping because of the fact that the channel is controlled by both front and back gates. Both the gates of DG MOSFET can be operated independently or electrically tied together. Independent DG MOSFET device undergoes a problem of misalignment between its two gates. So, an alternative device structure which overcome this problem is a FinFET, which is a multigate device structure and shows improved electrostatic control over the gate [5–8]. Other multigate FinFET devices are tri-gate, Pi (π), omega (Ω) gate and gate all around (GAA) which are greatly reduces the SCEs like DIBL, threshold voltage roll-off and leakage current [1]. Colinge first proposed the GAA MOS transistor [9–11]. The gate leakage current in predominate while scaling the device into sub-nanometer regime and it can be tackled by various alternative solutions [1, 12–14].

In this paper, initially the DG and rectangular GAA FinFET structures are simulated with identical dimensions. The modified rectangular GAA FinFET structure is presented later and which is known as the cylindrical GAA FinFET. In the rectangular GAA MOS structure there are two the additional channel in comparison with DG structure because of which GAA exhibit excellent control over the gate and this leads to reduced SCEs. It also shows better transconductance behavior. After analysing GAA structures, it was found that by optimizing fin width to gate length ratio, the short channel effects can be minimized [15]. The multigate devices mainly trigate FinFET and rectangular GAA were having the 3-D current flow because their cross sections do not have two dominant current flow planes [16]. The 3-D current flow anticipates some effects very easily which were known as corner effects. As these device structures exhibits the edges and the corners so whenever the gate voltage is applied on the device there is a creation of large electric field at the corners than at the edges. Due to this, there exists more electron density at the corners that results low

threshold voltage and thereby leads to the degradation of subthreshold swing (SS). The corner components of current causes severe changes in the device performance like lower threshold voltage, higher leakage current and for some structures they cause huge growth in the subthreshold characteristics. Therefore to overcome these effects a new modified structure known as Cylindrical Gate All Around (Cy-GAA) has been presented which not only reduces these effects but also diminish SCEs [17]. In this paper the comparison has been made between the DG and GAA MOSFETs for 20 nm technology by considering SCEs and corner effects. The channel remains undoped as undoped channels possess more mobility and less scattering of carriers.

This paper is organized as follows: Sect. 2 explains the device structure. The comparative study of DG and GAA FinFETs and the simulation results of cylindrical GAA FinFET structure with variation of physical dimensions of the device were presented in Sect. 3 and finally Sect. 4 concludes the paper.

2 Device Structure And Simulation Setup

Figure 1 shows the structure of DG and GAA FinFETs which were made on SOI substrate and the simulations were carried out in Sentaurus TCAD. The gate length (Lg), fin width (Wfin) and fin height (Hfin) were 20 nm, 20 nm, and 40 nm, respectively. For better control of SCEs, the Hfin must be twice of the Wfin. This condition
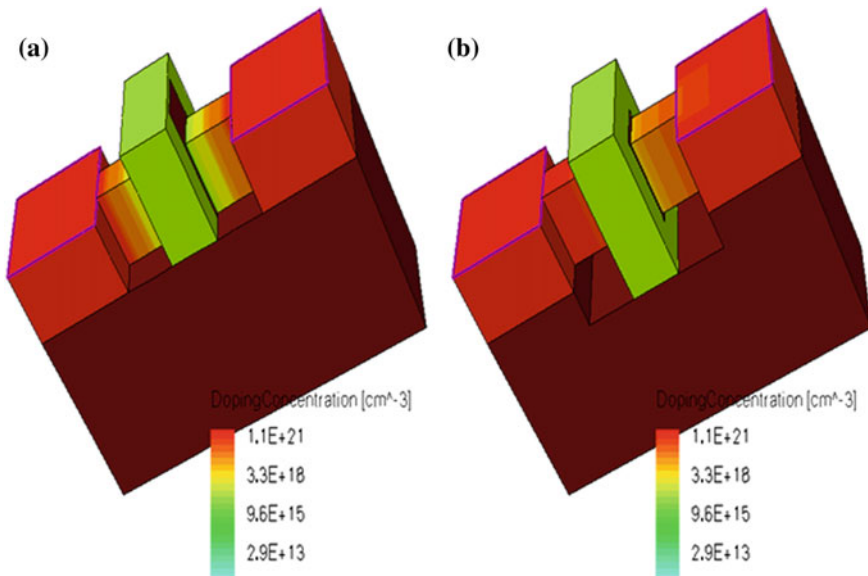
Fig. 1 Structure of FinFET devices. **a** DG, **b** GAA. The physical parameters are: gate length (Lg=20 nm), gate-oxide thickness (tox= 2 nm), and fin height (Hfin=40 nm), fin width (Wfin=20 nm) for both the devices

generally applies to most of the device dimensions for better potential distribution and improved electrostatic integrity. The gate-oxide thickness (tox) is considered as 2 nm. Channel of both the devices are lightly p-doped (1×10^{15} cm^{-3}) as low doped devices offer less random doping fluctuations compared to doped devices as mentioned earlier. The source and drain extension n-doping of 1×10^{20} cm^{-3} is slightly less than that of the source and drain n-doping of 1×10^{21} cm^{-3}. It has been done to avoid the depletion layer penetration in the source and drain extension regions. In numerical simulations for recombination and mobility Shockley–Read–Hall model is included, to solves the quantum potential equations self-consistently with the Poisson the density gradient transport model is incorporated, doping dependent mobility and high-field saturation models are included.

3   Results And Discussion

3.1   Comparison of GAA FinFETs with DG FinFETs

Figure 2 shows the transfer characteristics of DG and GAA FinFETs at drain voltage Vds = 0.1 V and 1 V at device dimensions as gate length (Lg = 20 nm), gate-oxide thickness (tox = 2 nm), fin width (Wfin=20 nm) and fin height (Hfin= 40 nm). It is clear evident from the Fig. 3 that, GAA structure exhibits higher on-current and lower off-current due to more number of gates and better control of channel region over the other counterpart. The same characteristics are used for calculation of DIBL and

Fig. 2 Transfer characteristics of DG and GAA FinFET
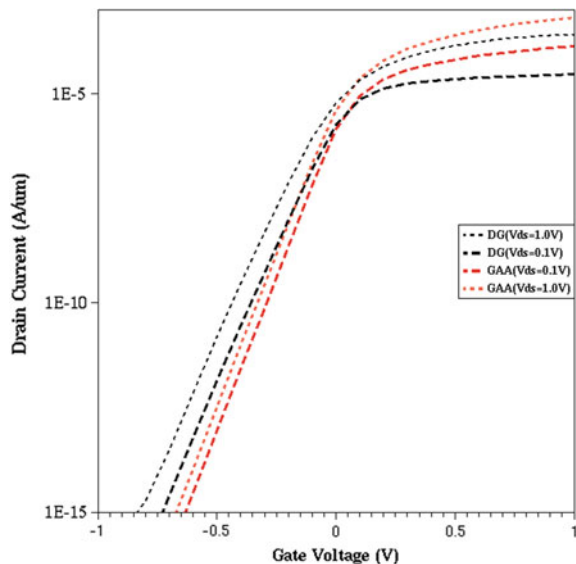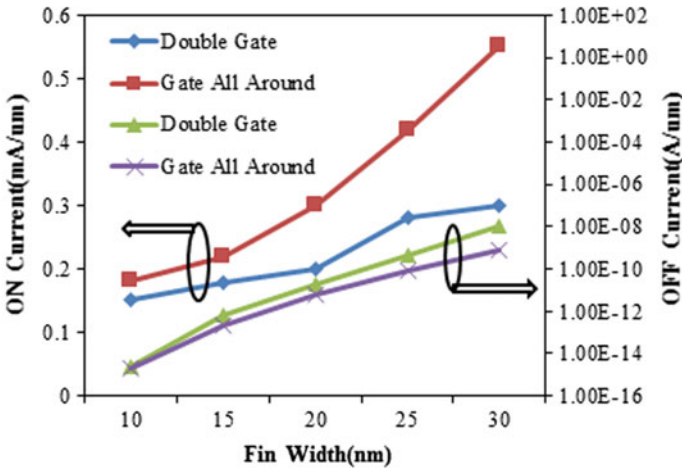
Fig. 3 On-current (I_{on}) and off-current (I_{off}) variation with respect to fin width variation (Wfin) of DG and GAA FinFET devices

threshold voltage roll-off of these two devices. As the threshold voltage roll-off was caused by charge sharing effect. With the increase in the fin width the on-current of GAA increases more swiftly because of the strong volume inversion. The on-current and off-current of FinFET device depends on the width of the channel region. The on-current and off-current of DG and GAA FinFET is shown in Fig. 3. As width of the fin region decreases the leakage path from center of the channel decreases which leads to reduced leakage current from the device. Another important short channel parameter is the subthreshold swing (SS) which need to be estimated to measure the leakage current of the devices. Figure 4 reveals the subthreshold swing (SS) of DG and GAA FinFET for the oxide thickness of 2 and 3 nm with various fin widths. It has been noticed that, GAA structure shows the better performance at tox = 2 nm with respect to all fin widths over the DG structure and at Wfin = 10 nm, GAA FinFET reveals minimum SS = 64.5 mV/decade.

3.2 Cylindrical and Rectangular GAA FinFETs

From the comparison of DG and rectangular GAA structure it has been noticed that rectangular GAA structure supress the SCEs to larger extent and exhibits the better performance. In rectangular GAA structure at all the corners of the fin there is a presence of electric field which hampers the performance and it is known as corner effect. This corner effect can be eliminated by modifying the rectangular structure into a cylindrical structure. Figure 5 shows the structures of cylindrical and rectangular channel GAA FinFETs. Here Lg = 20 nm, tox = 2 nm, Hfin = 30 nm, and Wfin = 30 nm for both the devices. Figure 6 demonstrates the transfer characteristics of

Fig. 4 SS variation with
respect to fin width (Wfin) of
DG and GAA FinFET
devices
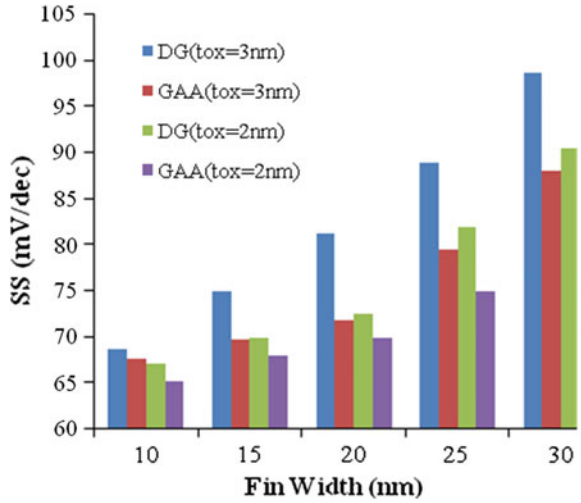
cylindrical and rectangular channel GAA FinFETs at drain voltage Vds = 0.1 and
1.0 V.

Figure 7 shows SS and DIBL of cylindrical GAA FinFET with various cylinder
diameters. With an increase in the diameter the SS and DIBL was increased and
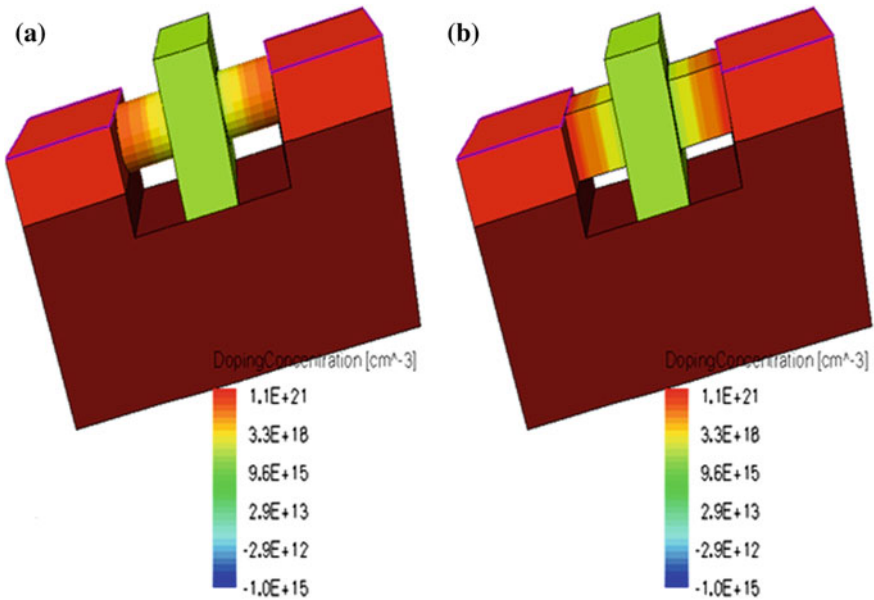
Fig. 5 Structure of FinFET devices. **a** Cylindrical GAA, **b** Rectangular GAA with Lg = 20 nm,
tox = 2 nm, Hfin = 30 nm, and Wfin = 30 nm for both the devices

Fig. 6 Transfer
characteristics of cylindrical
and rectangular GAA
FinFET devices
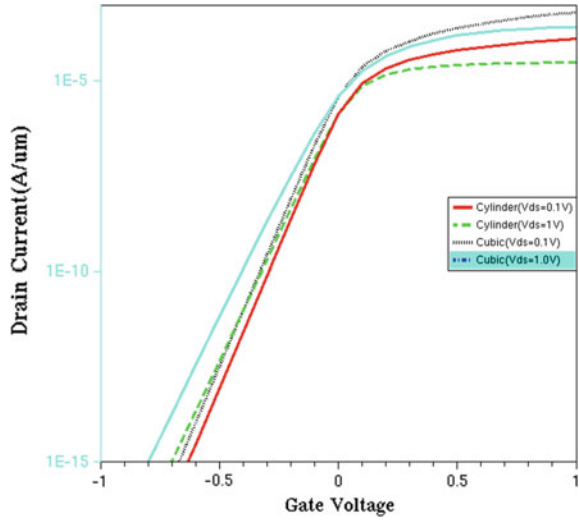
Fig. 7 SS and DIBL
comparison with respect to
diameter of cylinder of a
GAA MOSFET
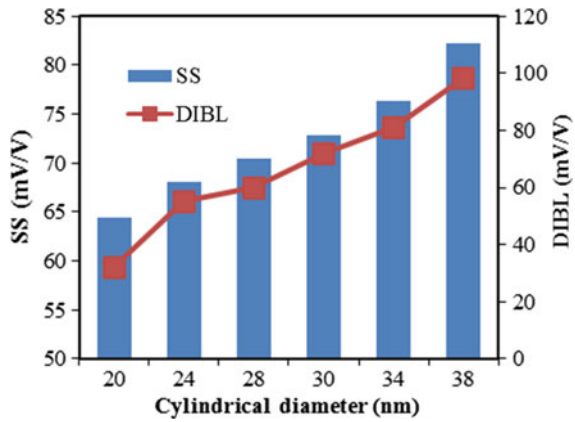
these two parameters are even within the acceptable range when the diameters of
the cylinder decreases which is a requirement for the advancing technology. For
obtaining optimized results in context with SCEs the diameter of cylinder remains
at 20 nm as shown in Fig. 8 that the I_{OFF} current reduces drastically as the diameter
decreases. The I_{ON} current increases with increase in diameter as predicted because
of the enhancement in the volume of the structure.
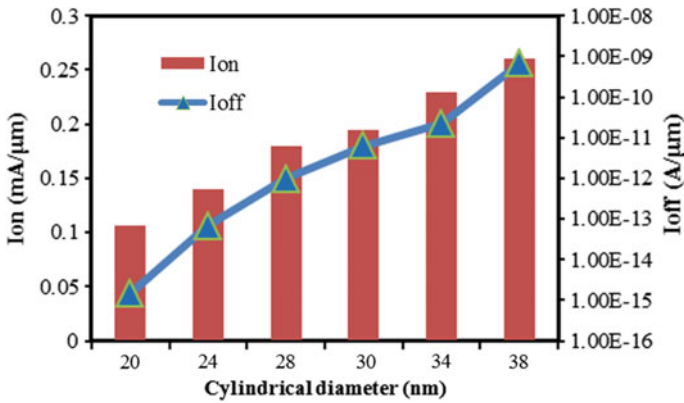
Fig. 8 Comparison of I_{ON} and I_{OFF} current with cylinder diameter of Cylindrical GAA MOSFET

4 Conclusions

Comparison between DG and GAA FinFET device structure has been done for the 20 nm gate length. The results confirmed better SCEs because of more gate control in GAA over DG FinFET counterpart. Off-current decrease in GAA FinFETs more rapidly and also the results shows increased on-current with increase in diameter of the fin. GAA FinFET provides strong volume inversion and that is why SS and DIBL show acceptable values. Simulations were also done for Cylindrical GAA FinFET and better outcome were shown as the corner effects were greatly reduced in these structures.

References

1. Roldán, J.B., Godoy, A., Gámiz, F., Balaguer, M.: Modeling the centroid and the inversion charge in cylindrical surrounding gate MOSFETs, including quantum effects. IEEE Trans. Electron Devices **55**, 411–416 (2008)
2. Havaldar, D.S., Katti, G., DasGupta, N., DasGupta, A.: Subthreshold current model of FinFETs based on analytical solution of 3-D poisson's equation. IEEE Trans. Electron Devices **53**(4), 737–742 (2006)
3. Veeraraghavan, S., Fossum, J.G.: Short-channel effects in SO1 MOSFET's. IEEE Trans. Electron Devices **36**(3), 522–528 (1989)
4. Katti, G., DasGupta, N., DasGupta, A.: Threshold voltage model for mesa-isolated small geometry fully depleted SOI MOSFETs based on analytical solution of 3-D poisson's equation. IEEE Trans. Electron Devices **51**(7), 1169–1177 (2004)
5. Pei, G., Kedzierski, J., Oldiges, P., Ieong, M., Kan, E.C.-C.: FinFET design considerations based on 3-D simulation and analytical modeling. IEEE Trans. Electron Devices **49**(8), 1411–1419 (2002)
6. Frank, D.J., Laux, S.E., Fischetti, M.V.: Monte Carlo simulation of 30 nm dual-gate MOSFET: how short can Si go?. IEDM Tech. Dig. 553–556 (1992)

7. Yan, R.-H., Ourmazd, A., Lee, K.F.: Scaling the Si MOSFET: from bulk to SOI to bulk. IEEE Trans. Electron Devices **39**, 1704–1710 (1992)
8. Suzuki, K., Tanaka, T., Tasaka, Y., Horie, H., Arimoto, Y.: Scaling theory for double-gate SOI MOSFETs. IEEE Trans. Electron Devices **40**, 2326–2329 (1993)
9. Colinge, J.-P.: Multiple-gate SOI MOSFETs. Solid State Electron **48**(6), 897–905 (2004)
10. Colinge, J.-P.: Novel gate concepts for MOS devices. In: Proceedings of the ESSDERC, Leuven, Belgium, pp. 45–49 (2004)
11. Colinge, J.-P., Gao, M.H., Romano-Rodriguez, A., Maes, H., Claeys, C.: Silicon-on-insulator gate-all-around devices. In: IEDM Technical Digest, San Francisco, CA, pp. 595–598 (1990)
12. Narendar, V., Mishra, R.A.: Analytical modeling and simulation of multigate FinFET devices and the impact of high-k dielectrics on short channel effects (SCEs). Superlattices Microstruct. **85**, 357–369 (2015)
13. Narendar, V., Girdhardas, K.A.: Surface potential modeling of graded-channel gate-stack (GCGS) high-K dielectric dual-material double-gate (DMDG) MOSFET and analog/RF performance study. J. Silicon (2018). https://doi.org/10.1007/s12633-018-9826-z)
14. Narendar, V.: Performance enhancement of FinFET devices with gate-stack (GS) high-K dielectrics for nanoscale applications. J. Silicon (2018). https://doi.org/10.1007/s12633-018-9774-7
15. Song, J.Y., Choi, W.Y., Park, J.H., Lee, J.D., Park, B.-G.: Design optimization of gate-all-around (GAA) MOSFETs. IEEE Trans. Electron Devices **5**(3), 186–191 (2006)
16. Pott, V., Moselund, K.E., Bouvet, D., De Michielis, L., Ionescu, A.M.: Fabrication and characterization of gate-all-around silicon nanowires on bulk silicon. IEEE Trans. Electron Devices **7**(6), 733–744 (2008)
17. Abd-Elhamid, H., Iñiguez, B., Roig, J.: Analytical model of the threshold voltage and subthreshold swing of undoped cylindrical gate all around MOSFETs. IEEE Trans. Electron Devices **52**(3), 572–579 (2007)

Cross-Coupled Bulk-Degenerated OTA Using Source Follower Auxiliary Pair to Improve Linearity

Tanmay Dubey [ORCID]**, Ravi Shankar and Vijaya Bhadauria**

Abstract In this paper, linearity is improved by using a source follower auxiliary pair in a source-degenerated cross-coupled bulk-degenerated OTA. This technique uses auxiliary source follower differential amplifier to drive the bulk terminals of the main differential pair. The proposed OTA as well as the conventional source-degenerated OTA are simulated in Cadence Virtuoso Tool using UMC $0.18\,\mu m$ CMOS technology having a power supply voltage of 1.8 V. The proposed OTA shows 22 dB improvement in HD_3 and THD as compared to conventional source-degenerated OTA at the cost of lower transconductance and increased power consumption.

Keywords Bulk degeneration $\cdot$ Cross-coupled OTA $\cdot$ Linearity $\cdot$ THD

1 Introduction

In the field of analog circuits and systems, operational transconductance amplifier (OTA) plays the role of a fundamental building block and is famous in the field of analog electronics for various applications like multipliers, continuous-time-filters, mixers, etc. [1–5]. OTA has a small range of input voltage that can produce linearized output current. Most of the present day applications require very high linear circuits having low power consumption capabilities. Various efforts have been reported using attenuation [6–8], nonlinear terms cancellation [9–12] and source degeneration technique [13–19] to improve linear performance of OTA. A new technique named bulk degeneration along with source degeneration has been reported in [20–22] which

T. Dubey (✉) · V. Bhadauria
Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India
e-mail: dubey.tanmay@gmail.com

V. Bhadauria
e-mail: vijaya@mnnit.ac.in

R. Shankar
Department of Revenue, Ministry of Finance, Hyderabad, India
e-mail: ravishankar.msit@gmail.com

683

further improves linearity performance of OTA. The bulk degeneration technique working principle is based on the utilization of bulk terminals as an additional control terminal and bulk terminals are connected to the inverted replica of the main differential pair or pseudo differential pair. In bulk degeneration technique, conventional methods to connect the bulk (connecting the bulk to the source or V_{DD} in PMOS) is not used instead bulk is connected to a varying differential voltage.

In this paper, a single ended cross-coupled source and bulk-degenerated OTA using auxiliary source follower is presented. Bulk degeneration technique works better with source degeneration technique in comparison of either of them working alone. Hence, in this work, a source-degenerated differential pair is used along with bulk degeneration to improve the linearity. Main differential pair bulk degeneration is obtained by a source follower auxiliary differential pair. The source of the auxiliary differential pair is connected with the bulk of the main differential pair in cross-coupled configuration to provide negative feedback hence increasing the linearity. Rest of the paper is organized as follows. Section 2 deals with the design and analysis of conventional source-degenerated OTA and proposed Cross-Coupled Source Follower Bulk-degenerated (CCSFBD) OTA. Simulation results are discussed in Sect. 3 and finally conclusions are drawn in Sect. 4.

2 The Proposed Cross-Coupled Source Follower Bulk-Degenerated (CCSFBD) OTA

A source-degenerated OTA circuit is shown in Fig. 1. Source degeneration technique is one of the popular techniques widely adopted because of its simplicity. In this technique, a degeneration resistor is connected to the source of differential pair input. The source degeneration resistor is the key factor in this technique because it provides negative feedback and the linearity performance highly depends on it.

The circuit diagram of the proposed OTA is shown in Fig. 2. The proposed OTA is based on the technique of source and bulk degeneration. A degeneration resistor R_{DEG} is connected with the source of main PMOS differential pair M_1–M_2. The bulk terminals of main differential pair M_1–M_2 are connected with the source of auxiliary PMOS differential pair M_4 and M_3 respectively. The bulks of auxiliary differential pair M_3–M_4 are connected to their own source and have no body effect.

Input signal is applied in such a way that $V_{in+} = -(V_{in-})$. If V_{in+} increases, it leads to increase in magnitude of drain current $I_{D,M2}$ and at the same time V_{in-} decreases and it leads to decrease in source voltage $V_{S,M3}$. Since $V_{S,M3}$ is connected to the bulk of M_2, bulk-source voltage of M_2 ($V_{BS,M2}$) decreases and this results in increase in threshold voltage (V_T) of M_2 by a small fraction. Increase in V_T leads to decrease in drain current of M_2 and the change in drain current in saturation region of main differential pair is linearized by bulk degeneration. Similarly, same working principle can be applied if V_{in+} decreases and V_{in-} increases. The advantage
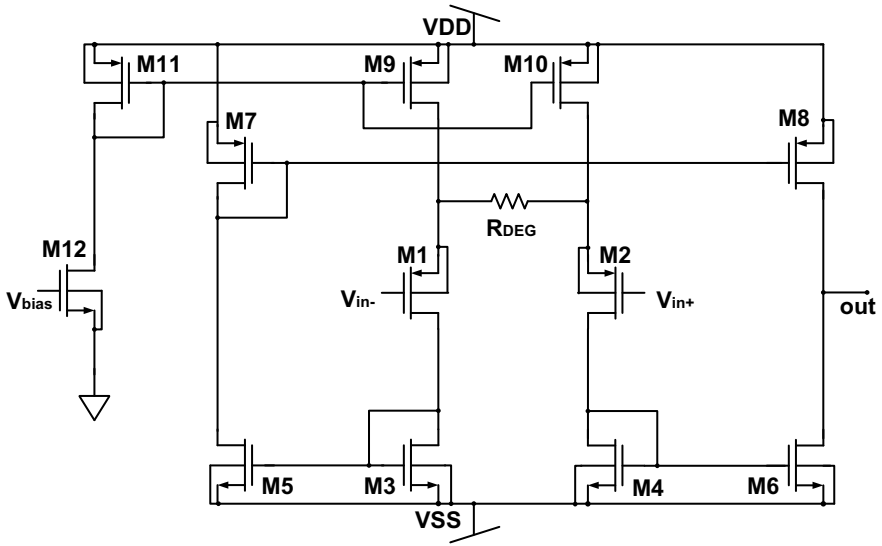
Fig. 1 Source-degenerated OTA
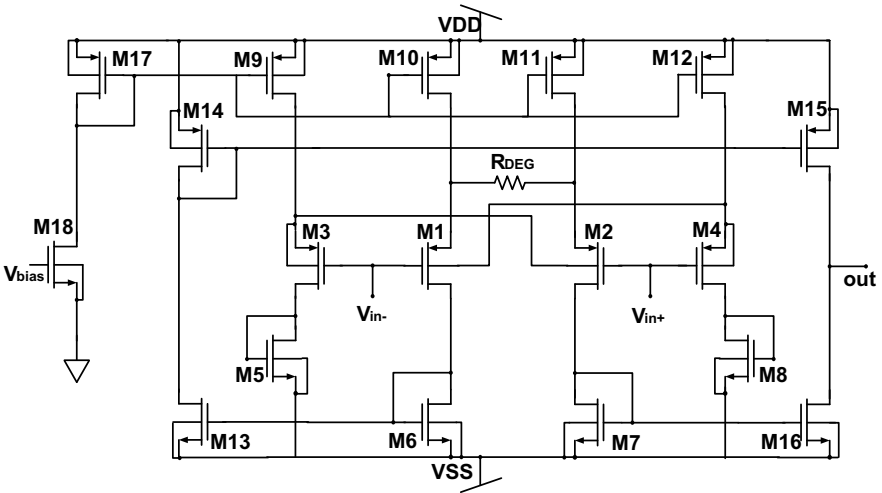
Fig. 2 The proposed CCSFBD OTA

of using source follower circuit helps in improving the output current in terms of varying differential input voltage. The operating principle of this methodology deals with complex mathematical equations. Drain current of transistors M_1 and M_2 in saturation is given by

$$I_{D,M1} = K\left(V_{in-} - V_{S,M1} - V_{T,M1}\right)^2 \tag{1}$$

$$I_{D,M2} = K\left(V_{in+} - V_{S,M2} - V_{T,M2}\right)^2 \tag{2}$$

The output current I_{out} is expressed as

$$I_{out} = I_{D,M1} - I_{D,M2} \tag{3}$$

Using (1) and (2), (3) can be reformed as

$$I_{out} = K\left[(V_{in} - I_{out}R_{DEG}) - \left(V_{T,M1} - V_{T,M2}\right)\right]*$$
$$\sqrt{\frac{4I_B}{k} - \left[(V_{in} - I_{out}R_{DEG}) - \left(V_{T,M1} - V_{T,M2}\right)\right]^2} \tag{4}$$

where R_{DEG} is the degeneration resistance, I_B is the bias current, $V_{T,M1}$ and $V_{T,M2}$ are the threshold voltage of transistors M_1 and M_2, respectively. Considering the body effect, $V_{T,M1}$ and $V_{T,M2}$ are expressed as

$$V_{T,M1} = V_{T0} - \gamma\left(\sqrt{|2\phi_F| + V_{BS,M1}} - \sqrt{|2\phi_F|}\right) \tag{5}$$

$$V_{T,M2} = V_{T0} - \gamma\left(\sqrt{|2\phi_F| + V_{BS,M2}} - \sqrt{|2\phi_F|}\right) \tag{6}$$

Combining (4)–(6), the output current I_{out} is expressed as

$$I_{out} = K\left[(V_{in} - I_{out}R_{DEG}) - \gamma\left\{\sqrt{|(2\phi_F - V_{BS,M2})|} - \sqrt{|(2\phi_F - V_{BS,M1})|}\right\}\right]*$$
$$\sqrt{\frac{4I_B}{k} - \left[(V_{in} - I_{out}R_{DEG}) - \gamma\left\{\sqrt{|(2\phi_F - V_{BS,M2})|} - \sqrt{|(2\phi_F - V_{BS,M1})|}\right\}\right]^2} \tag{7}$$

From (7), it can be observed that the square term is reduced by a positive factor. Hence, the nonlinearity caused by the square term of V_{in} is also reduce by a positive factor resulting the overall expression of I_{out} more linear.

3　Simulation Result and Discussion

The proposed CCSFBD OTA and conventional source-degenerated OTA simulation is performed in Cadence Virtuoso using 180 nm CMOS process technology. The circuit has been powered with low power supply of 1.8 V. Bias current is taken 150 µA for both circuits. To keep the transistors at proper bias, common mode voltage, V_{CM} is taken 0.3 V. Source-degenerated resistor, R_{DEG} is taken 50 KΩ and is connected between the source terminals of main differential pair. Conventional source-degenerated OTA, shown in Fig. 1, is used for simulation purpose and the results of proposed OTA are compared with conventional source-degenerated OTA. Simulation is performed to calculate HD_3, THD, linear output current range and transconductance. Simulation results of HD_3, THD and linear range are shown in Figs. 3, 4 and 5 respectively.

HD_3 result of proposed CCSFBD OTA is better than conventional source-degenerated OTA by −22 dB up to a peak to peak input range of 1.0 V. HD_3 of proposed OTA remains below −65 dB up to 1.0 V_{pp} input voltage. Maximum improvement in HD3 of proposed circuit over conventional circuit is obtained at 0.7 V_{pp} input voltage. THD result of proposed CCSFBD OTA is also better than conventional bulk-degenerated OTA. CCSFBD OTA has THD below −62 dB up to input voltage 1.0 V_{pp} and improved THD is obtained throughout the simulation performed. The proposed circuit performs well at higher input voltage of 1 V_{pp}.

The transconductance variation of conventional source-degenerated OTA and proposed CCSFBD OTA with respect to the input voltage is compared in Fig. 6. The proposed CCSFBD OTA performs very good in terms of linearity. The transconduc-
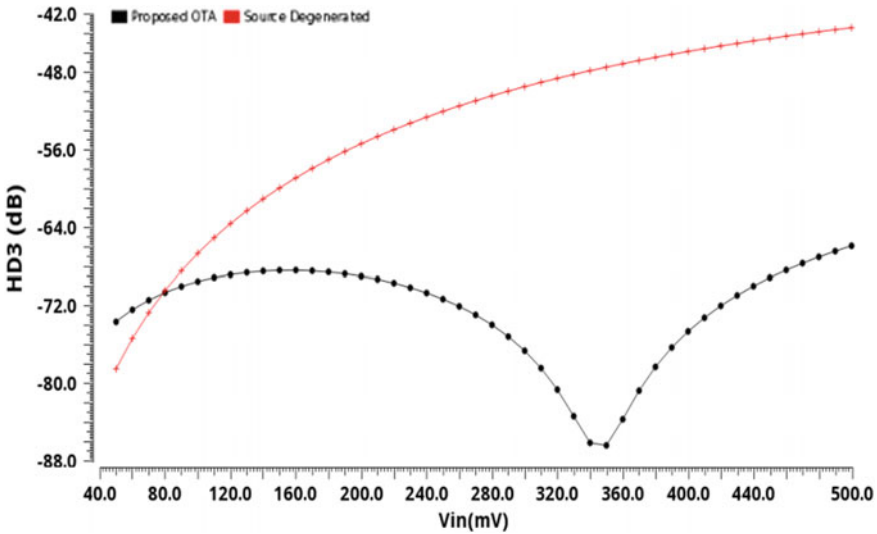
Fig. 3 HD_3 result of OTAs for input frequency 10 MHz
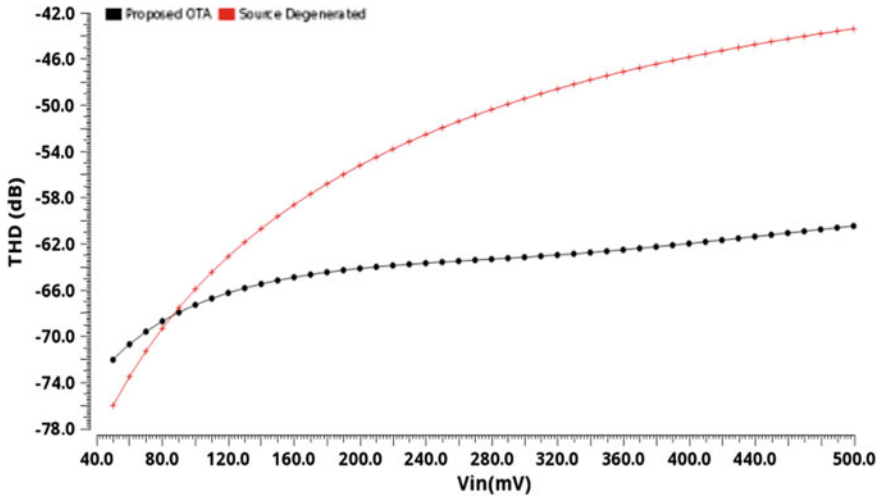
Fig. 4 THD result of OTAs for input frequency 10 MHz
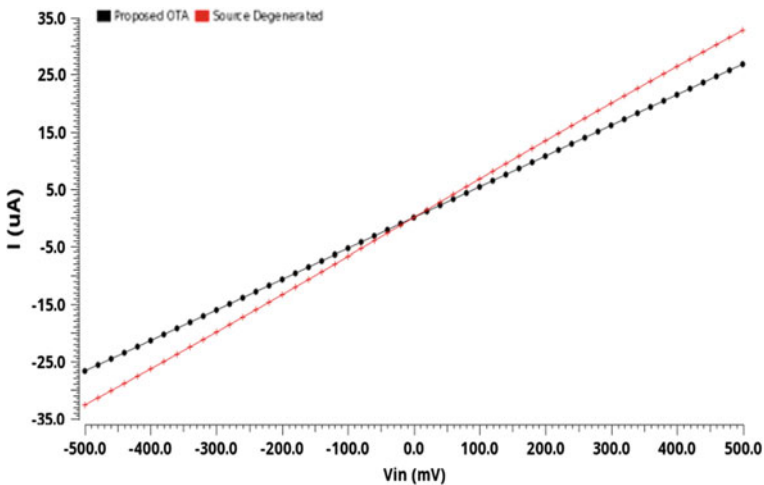
Fig. 5 Output current linear range of OTAs

tance of CCSFBD OTA remains constant for 1 V_{pp} variation of input signal while conventional source-degenerated OTA shows reasonable transconductance variation in this range of input voltage. However, this linearity improvement is achieved at the cost of slight increase in power; low G_m/I_{total} and lower transconductance value of 53 µA/V. Performance summary of both the OTAs are shown in Table 1.
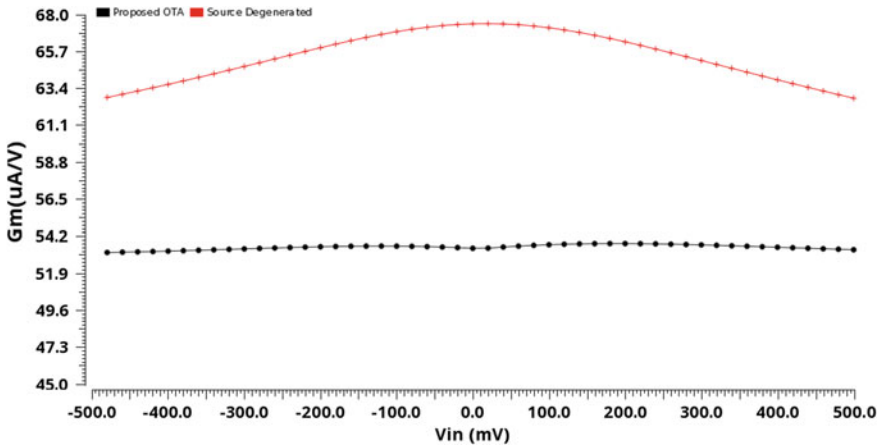
Fig. 6 Transconductance variation of OTAs

Table 1 Performance summary of OTAs

Parameter	Source degenerated OTA	CCSFBD OTA
G_m (μm/V)	67.44	53.47
HD_3 for 1 V_{pp} @ 10 MHz (dB)	-44	-66
THD for 1 V_{pp} @ 10 MHz (dB)	-42	-64
G_m/I_{total}	0.15 V^{-1}	0.09 V^{-1}
P_{Diss} (mW)	0.815	1.089

4   Conclusion

In this paper, a cross-coupled bulk-degenerated OTA is presented. The proposed OTA offers improved linearity over conventional source-degenerated OTA at the cost of reduced transconductance and slight increase in power dissipation. The THD and HD_3 performance is improved by -22 dB for 1 V_{pp} input range. THD performance is throughout better for applied input range.

References

1. Behzad, R.: Design of Analog CMOS Integrated Circuits. Tata McGraw-Hill, New York (2001)
2. Sedra, A.S., Smith, K.C.: Microelectronic Circuits, 6th edn. Oxford University Press, Oxford (1998)
3. Dubey, T., Pandey, R.: Low-voltage highly linear floating gate MOSFET based source degenerated OTA and its applications. Inf MIDEM **48**, 19–28 (2018)
4. Garradhi, K., Hassen, N., Ettaghzouti, T., Besbes, K.: Realization of current-mode biquadratic filter employing multiple output OTAs and MO-CCII. AEU Int. J. Electron. Commun. **83**, 168–179 (2018). https://doi.org/10.1016/j.aeue.2017.08.027
5. Sun, C.Y., Lee, S.Y.: A fifth-order butterworth OTA-C LPF with multiple-output differential-input OTA for ECG applications. IEEE Trans. Circuits Syst. II Express Briefs **65**, 421–425 (2018). https://doi.org/10.1109/TCSII.2017.2695366
6. Bhadauria, V., Kant, K., Banerjee, S.: Linearity enhancement of 0.18 μm transconductor using active attenuation technique. Can. J. Electr. Electron. Eng. **2**, 598–601 (2011)
7. Kulshreshtha, T., Bhadauria, V.: A highly linear CMOS pseudo differential transconductor using active attenuator. In: ICPCES 2010—International Conference on Power, Control and Embedded Systems, pp. 2–5 (2010). https://doi.org/10.1109/icpces.2010.5698684
8. El, Mourabit A., Lu, G., Pittet, P.: Wide-linear-range subthreshold OTA for low-power, low-voltage, and low-frequency applications. IEEE Trans. Circuits Syst. Regul. Pap. **52**, 1481–1488 (2005)
9. Lo, T.Y., Hung, C.C.: A 1-V 50-MHz pseudodifferential OTA with compensation of the mobility reduction. IEEE Trans. Circuits Syst. II Express Briefs **54**, 1047–1051 (2007). https://doi.org/10.1109/TCSII.2007.907559
10. Kwon, K., Kim, H.T., Lee, K.: A 50-300-MHz highly linear and low-noise CMOS Gm-C filter adopting multiple gated transistors for digital TV tuner ICs. IEEE Trans. Microw. Theory Tech. **57**, 306–313 (2009). https://doi.org/10.1109/TMTT.2008.2009904
11. Le-Thai, H.L.-T.H., Nguyen, H.-H.N.H.-H., Nguyen, H.-N.N.H.-N., et al.: An IF bandpass filter based on a low distortion transconductor. IEEE J. Solid-State Circuits **45**, 2250–2261 (2010). https://doi.org/10.1109/JSSC.2010.2063991
12. Khumsat, P.: Linearisation technique for low-voltage tuneable Nauta's transconductor in Gm–C filter design. IET Circuits Devices Syst. **12**, 347–361 (2018). https://doi.org/10.1049/iet-cds.2017.0177
13. Worapishet, A., Naphaphan, C.: Current-feedback source-degenerated CMOS transconductor with very high linearity. Electron. Lett. **39**, 17 (2003). https://doi.org/10.1049/el:20030050
14. López-martín, A.J., Ramirez-angulo, J., Durbha, C., et al.: A CMOS transconductor with multidecade tuning using balanced current scaling in moderate In-version. IEEE J. Solid-State Circuits **40**, 1078–1083 (2005)
15. Petraglia, A.: Linearly tunable CMOS OTA with constant dynamic range using source-degenerated current mirrors. IEEE Trans. Circuits Syst. II Express Briefs **53**, 797–801 (2006). https://doi.org/10.1109/TCSII.2006.881162
16. Kuo, K.C., Leuciuc, A.: A linear MOS transconductor using source degeneration and adaptive biasing. IEEE Trans. Circuits Syst. II Analog Digit. Signal Process. **48**, 937–943 (2001). https://doi.org/10.1109/82.974782
17. Rezaei, F.: Linearity enhancement in the entire tuning range of CMOS OTA using a new tune compensated source degeneration technique. Microelectron. J. **66**, 128–135 (2017). https://doi.org/10.1016/j.mejo.2017.06.008
18. Elamien, M.B., Mahmoud, S.A.: On the design of highly linear CMOS digitally programmable operational transconductance amplifiers for low and high-frequency applications. Analog Integr. Circuits Signal Process. **2** (2018) https://doi.org/10.1007/s10470-018-1128-2
19. Elamien, M.B., Mahmoud, S.A.: Analysis and design of a highly linear CMOS OTA for portable biomedical applications in 90 nm CMOS. Microelectron. J. **70**, 72–80 (2017). https://doi.org/10.1016/j.mejo.2017.10.009

20. Monsurrò, P., Pennisi, S., Scotti, G., Trifiletti, A.: Linearization technique for source-degenerated CMOS differential transconductors. IEEE Trans. Circuits Syst. II Express Briefs **54**, 848–852 (2007). https://doi.org/10.1109/TCSII.2007.906203
21. Gak, J., Miguez, M.R., Arnaud, A.: Nanopower OTAs with improved linearity and low input offset using bulk degeneration. IEEE Trans. Circuits Syst. I Regul. Pap. **61**, 689–698 (2014). https://doi.org/10.1109/TCSI.2013.2284002
22. Sharma, U., Dubey, T., Bhadauria, V.: Linearity enhancement using bulk-degeneration for source degenerated OTAs. In: 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC). IEEE, pp. 1–5 (2018)

Highly Linear Source-Degenerated OTA with Floating Gate Auxiliary Differential Pair

Tanmay Dubey , **Anurag Kumar and Vijaya Bhadauria**

Abstract This paper presents a linear operational transconductance amplifier (OTA) using two signal attenuation techniques (floating gate MOSFET along with nonlinear attenuator) and source degeneration technique to achieve the large input differential voltage range for which linear voltage-to-current conversion is obtained. The proposed linear OTA is simulated on Cadence Virtuoso environment with UMC $0.18\,\mu m$ CMOS process technology having power supply of 1.8 V. The proposed OTA gives transconductance (G_m) of 70 μA/V when biased with 100 μA current and the third-order harmonic distortion (HD_3) is found to be -79 dB for a 400 mV_{pp} input signal at 10 MHz.

Keywords Floating gate · Linearity · Nonlinear attenuator · OTA · Source-degeneration

1 Introduction

OTA is one of the important building blocks for many analog circuits like continuous time filter and tunable signal generators [1, 2]. These devices require the OTA to be more capable to convert input voltage to current linearly over significant input range. However due to the nonlinearity possess by its basic component MOSFET, OTA also shows some kind of nonlinear behavior while converting the voltage into the current. The differential nature of the OTA eliminates even harmonics but odd harmonics still presents in the form of the nonlinearity affecting the overall circuit performance. So

T. Dubey (✉) · V. Bhadauria
Electronics and Communication Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Allahabad, India
e-mail: dubey.tanmay@gmail.com

V. Bhadauria
e-mail: vijaya@mnnit.ac.in

A. Kumar
Bharat Sanchar Nigam Limited, Ballygunge, Kolkata, India
e-mail: anuemail@yahoo.com

for enhancing the linearity of OTA, several technique have been proposed like input signal attenuation technique (through active attenuator [3, 4], bulk driven [5–7] and floating gate [8, 9]), nonlinear term cancellation [10, 11], mobility compensation techniques [12], square root circuit technique [13] and source degeneration [14, 15].

In modern technology, device size is reducing very fast so for satisfactory operation, it requires good linearity at very low power. Sometimes, single technique is not able to provide required linearity and in that case two or more techniques are combined to achieve highly linear operation. Source degeneration technique is one of the famous techniques used for OTA linearity improvement by using a passive resistor or MOS transistor operating in triode region. Due to its simplicity, it is used in combination with other linearity techniques by many researchers to obtain desired linearity.

In this paper, source degeneration technique is used with two signal attenuation techniques: (1) nonlinear attenuator and (2) floating gate MOSFETs, to design a highly linear OTA. Section 2 of this paper describes the basic OTA structures and the proposed OTA is presented in Sect. 3. The simulation results are presented in Sect. 4 and the paper concludes in Sect. 5.

2 Basic OTA Structures

The input stage of a basic OTA is a differential pair and the input differential voltage is applied at the gate of the MOSFETs of differential pair and the output current is the difference of the drain currents of differential pair. By considering saturation region operation, the single ended differential output current of basic differential pair is given by [16]

$$I_{out} = G_m * V_{indiff} \sqrt{1 - \frac{\beta V^2_{indiff}}{8I_b}} \tag{1}$$

where V_{indiff} is the input differential voltage, I_b is the bias current, β is the transconductance parameter and G_m is the overall transconductance of the OTA which is expressed as

$$G_m = \sqrt{2\beta I_b} \tag{2}$$

By expanding (1) in a Taylor series, the third harmonic distortion (HD_3) for the OTA is given by [3],

$$HD_3 = \frac{\beta V^2{}_{indiff}}{64I_b} \tag{3}$$

From (3), it is evident that output current of an OTA contains a third harmonic term. Hence, the main objective for any linearization technique is to reduce the effect of third order harmonic distortion by reducing its harmonics component.

Source degeneration technique is one of the popular linearization techniques. In this technique, the harmonic distortion is reduced by the means of a resistor (R_s) or MOS transistor operating in linear region as shown in Fig. 1 where transconductance is reduced by a factor of ($1 + G_m * R_s$) and HD$_3$ is reduced by a factor of ($1 + G_m * R_s$)^2. So for the source-degenerated OTA, the overall transconductance (G_{mdeg}) and third order harmonics distortion (HD_{3deg}) is given as

$$G_{m\,\mathrm{deg}} = \frac{\sqrt{2\beta I_b}}{(1 + N_{deg})}; \ HD_{3\,\mathrm{deg}} = \frac{\beta V^2{}_{indiff}}{64(1 + N_{deg})^2 I_b} \tag{4}$$

where

$$N_{deg} = G_m * R_s \tag{5}$$

But by using this technique, only a moderate input linear range is achieved with poor common mode rejection ratio. So to improve the input linear range, combination of two different techniques is used as reported in [3]. In this technique, an attenuator is used followed by a source-degenerated OTA as shown in Fig. 2.
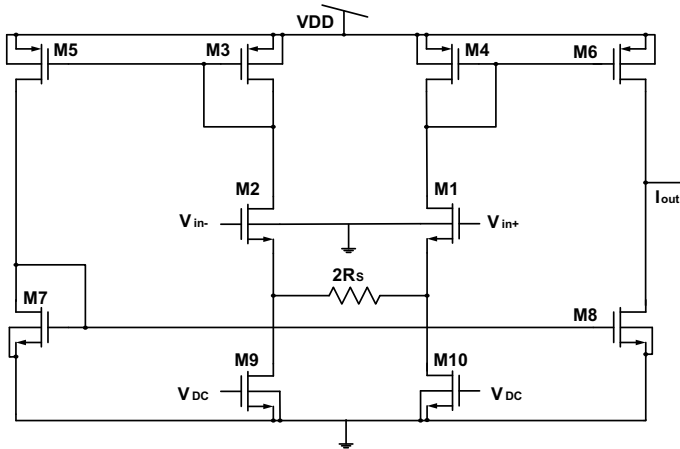
Fig. 1 Source degenerated OTA
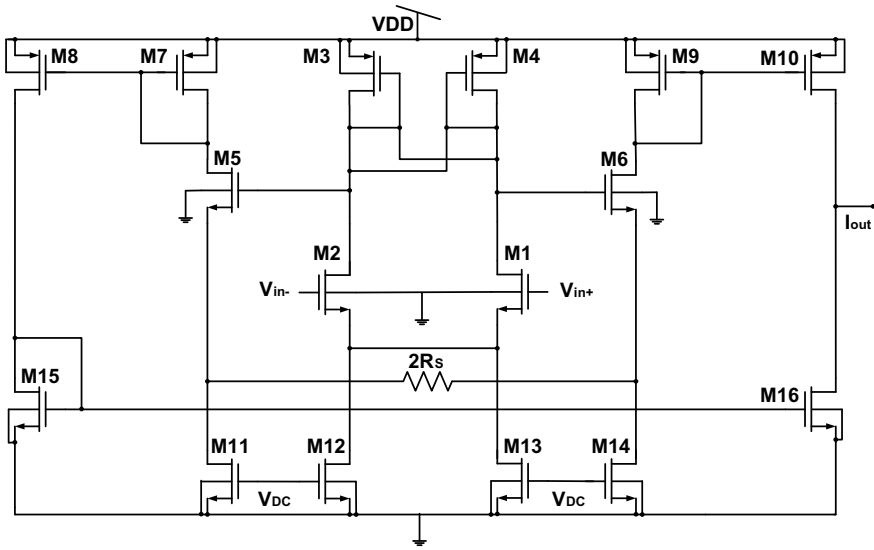
Fig. 2 Attenuated source degenerated OTA

In this OTA, the signal is first attenuated by an auxiliary differential pair and the output of this auxiliary differential pair is given as an input to source-degenerated differential pair. The overall transconductance (G_{madeg}) and third order harmonic distortion (HD_{3madeg}) of this OTA can be derived as reported in [3]

$$G_{madeg} = \frac{\sqrt{2\beta I_b}}{(1 + N_{deg})\sqrt{m_a}}; \quad HD_{3madeg} = \frac{\beta V^2_{indiff}}{64(1 + N_{deg})^2 m_a I_b} \tag{5}$$

where m_a is the attenuation factor and can be expressed as

$$m_a = \frac{\beta_{pMOS}}{\beta_{nMOS}} = \left(\frac{\mu_{pMOS}}{\mu_{nMOS}}\right)\left\{\frac{(W/L)_{pMOS3,4}}{(W/L)_{nMOS1,2}}\right\} \tag{6}$$

So from the above expression, HD$_3$ is reduced by factor $(1 + N_{deg})^2 m_a$ and G_m is reduced by a factor $(1 + N_{deg})m_a^{1/2}$. The increase in the value of m_a or R_s will reduce the value of harmonic distortion and simultaneously increase the linear range.

Source degeneration can also be used with floating gate technique as shown in Fig. 3 where input transistors are floating gate transistors. The input differential voltage is applied to one of the gates of these floating gate transistors and second gate is connected to a proper biasing voltage to reduce the threshold voltage of the floating gate transistor as reported in [17].
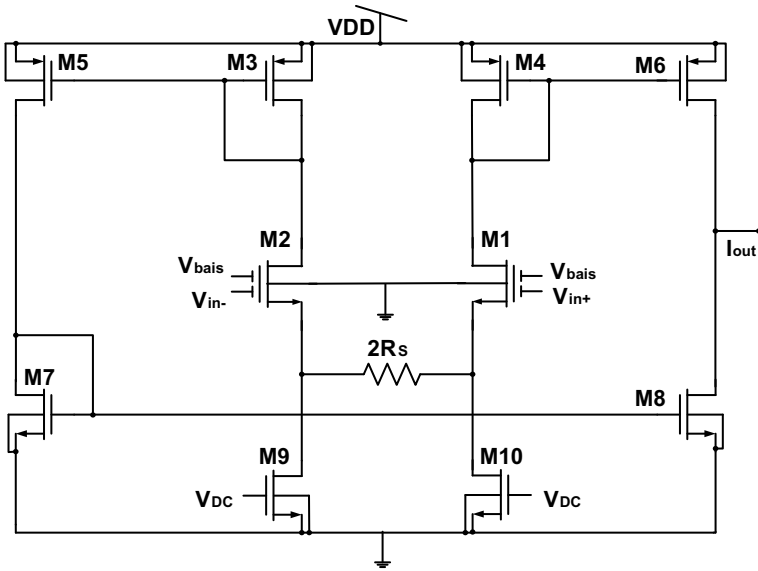
Fig. 3 Source degenerated floating gate OTA

$$V_{Teffe} = \frac{V_T - K_b V_{bias}}{K_1} \tag{7}$$

where $K_1 = C_1/C_T$ and $K_b = C_b/C_T$ are the ratio of the capacitance connected to the input voltage and bias voltage to the total capacitance seen at the floating gate, respectively. The effective transconductance of floating gate transistor is

$$g_{m,FG} = K_1 g_m \tag{8}$$

The total transconductance (G_{mFGdeg}) and third harmonic distortion (HD_{3FGdeg}) of the OTA shown in Fig. 3 can be expressed as

$$G_{mFG\,deg} = \frac{\sqrt{2\beta I_b}}{(1 + N_{deg})} * K_1; \; HD_{3FGdeg} = \frac{\beta V^2_{indiff}}{64(1 + N_{deg})^2 I_b} K_1^2 \tag{9}$$

From (9), it can be observe that the overall transconductance and the third harmonic distortion for the OTA can be reduced by reducing K_1.

3   Proposed OTA

In the proposed OTA, floating gate technique is used with nonlinear attenuator and source degeneration to increase the overall linearity of the OTA. The circuit of the proposed OTA is shown in Fig. 4. In this OTA, the signal is attenuated with the help of auxiliary differential pair whose input transistors are floating gate transistors. The output of this floating gate auxiliary pair is applied to the main differential pair with source degeneration.

Now by using (4), (5) and (9), the overall transconductance and third order harmonic distortion of proposed OTA can expressed as

$$G_{mprop} = \frac{\sqrt{2\beta I_b}}{(1 + N_{deg})\sqrt{m_a}} K_1; \; HD_{3prop} = \frac{\beta V^2_{indiff}}{64(1 + N_{deg})^2 m_a I_b} K_1^2 \qquad (10)$$

From the above expressions for the proposed circuit, it is clear that the overall transconductance is reduced by a factor of $((1 + N_{deg})m_a^{1/2} C_T)/C_1$ and the third order harmonic component is reduced by a factor $((1 + N_{deg})^2 m_a)(C_T/C_1)^2$. Therefore, it can be stated that the third harmonic distortion is reduced by combining floating gate with nonlinear attenuator and the source degeneration, resulting the increase in the overall linearity of proposed OTA.
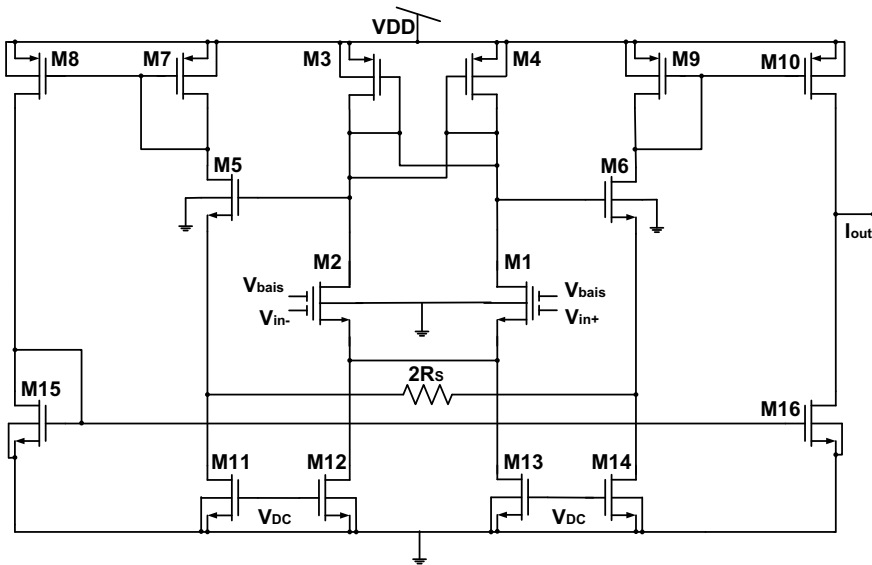
Fig. 4 Proposed OTA

4 Simulation Results

All the described OTAs and the proposed OTA are designed in Cadence Virtuoso environment with UMC 0.18 µm CMOS process technology. The power supply is set to be 1.8 V for all the OTAs. The dimensions of all the transistors of proposed OTA are listed in Table 1. The transconductance variations with input differential voltage for all the OTAs are shown in Fig. 5.

All the simulation results have been simulated based on single ended output which adds significant amount of distortion with simple current mirror. The plots in Fig. 5 are obtained for different values of parameters on which the respective transconductance is dependent. The constant G_m input swing for proposed OTA is maximum for $m_a = 3$, $R_s = 6$ kΩ and $V_{bias} = 1.5$ V. The total harmonic distortion plot for 10 MHz input differential signal is shown in Fig. 6. For 0.4 V_{pp} input signal, proposed OTA shows -79 dB HD$_3$. The same analysis is done for all the OTAs and results are compared in Fig. 7. From the results, it is clear that the proposed OTA has minimum HD$_3$ than the other OTA configurations up to 0.6 V_{pp} input voltage. The time domain response of the proposed OTA is shown in Fig. 8. The power consumed by the proposed OTA is 579 µW. The comparison of all the OTAs is listed in Table 2 and the performance comparison of proposed OTA with the existing OTAs available in the literature is tabulated in Table 3.

Table 1 Transistor sizes of proposed OTA

Transistor	W (µm)	L (µm)
M_1, M_2, M_5, M_6	5.22	1
M_3, M_4	19	1
M_7, M_8, M_9, M_{10}	22	1
M_{11}, M_{14}	1.03	0.36
M_{12}, M_{13}	2.2	0.36
M_{15}, M_{16}	10	1
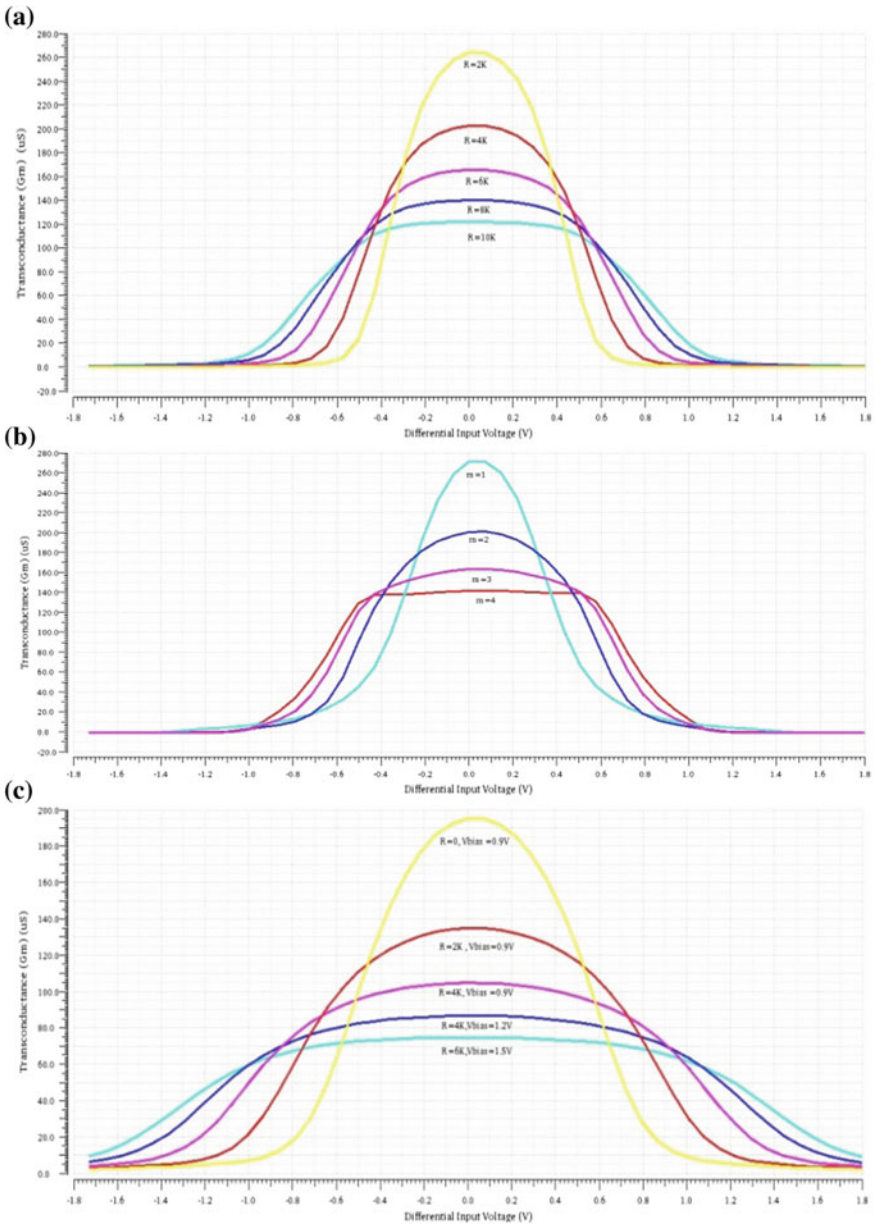
Fig. 5 a Transconductance versus signal input of source degenerated OTA. **b** Transconductance versus signal input of attenuated source degenerated OTA. **c** Transconductance versus signal input of source degenerated floating gate OTA. **d** Transconductance versus signal input of proposed OTA

(d)
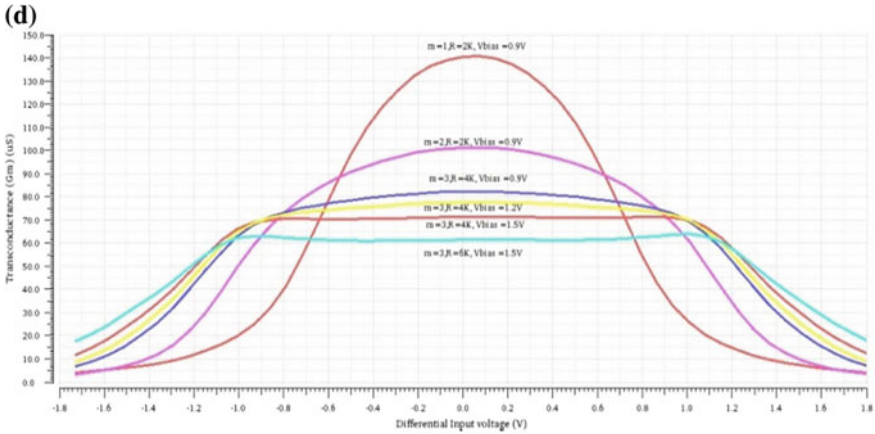
Fig. 5 (continued)
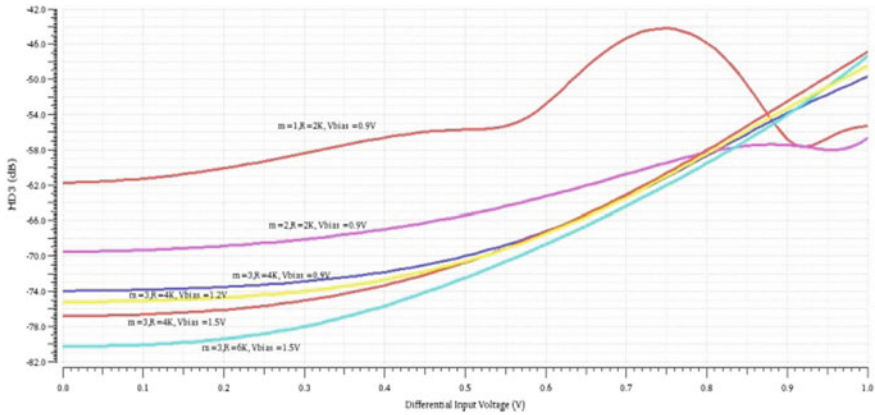
Fig. 6 HD$_3$ versus signal input of the proposed OTA for different values of m_a, R_s, and V_{bais}

5 Conclusions

This paper proposed a linear OTA comprises of combination of nonlinear attenuator with floating gate and source degeneration to obtain reduced third harmonic distortion hence increasing the overall linearity. The proposed circuit is designed with 0.18 µm CMOS technology and compared with different OTAs structure. Simulation results show that for input signal of 0.4 V$_{pp}$ input voltage at 10 MHz, the HD3 is below −79 dB which is better than other OTA configurations. The power consumed by proposed OTA is 579 µW which is partially higher than the other structures and can be considered as tradeoff for the linearity.
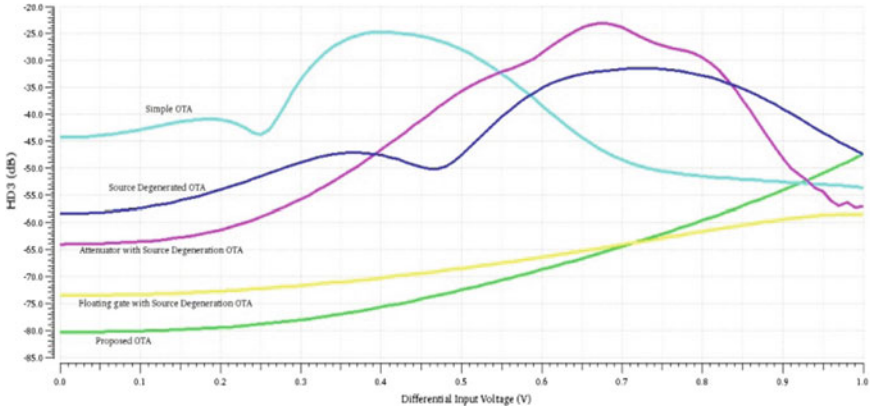
Fig. 7 Comparison of HD$_3$ versus signal input of the proposed OTA ($m_a = 3$, $R_s = 6\,\text{k}\Omega$ and V_{bias} = 1.5 V) with other basic OTAs
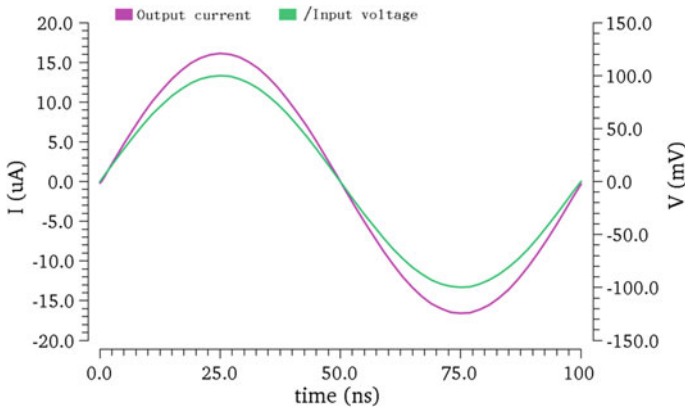
Fig. 8 Time domain response of the proposed OTA

Table 2 Performance summary of OTAs

Parameter	Simple OTA	SD OTA	Attenuated SD OTA	SD floating gate OTA	Proposed OTA
G$_m$ (μA/V)	390	170	140	100	70
HD$_3$ for 0.4 V$_{pp}$ @ 10 MHz (dB)	−41	−54	−62	−74	−79
Power (μW)	314	380	505	404	579

Table 3 Performance comparison of proposed OTA

Parameter	[3]	[8]	[12]	[15]	Proposed OTA
Technology (μm)	0.18	0.18	0.18	0.18	0.18
Power supply (V)	1	±0.6	1	1.8	1.8
G_m (μA/V)	90	75.5	110	6–32	70
HD_3(dB)@ V_{pp} @freq (MHz)	−70@0.6@5	−42@1@5	−61@0.5@5	−58@0.9@10	−79@0.4@10
Power (μW)	450	560	450	600	579

References

1. Voorman, H., Veenstra, H.: Tunable high-frequency Gm-C filters. IEEE J. Solid-State Circuits **35**, 1097–1108 (2000). https://doi.org/10.1109/4.859498
2. Kar, S.K., Sen, S.: Tunable square-wave generator for integrated sensor applications. IEEE Trans. Instrum. Meas. **60**, 3369–3375 (2011)
3. Kar, S.K., Sen, S.: A highly linear CMOS transconductance amplifier in 180 nm process technology. Analog Integr. Circuits Signal Process. **72**, 163–171 (2012). https://doi.org/10.1007/s10470-011-9796-1
4. Bhadauria, V., Kant, K., Banerjee, S.: Linearity enhancement of 0.18 μm transconductor using active attenuation technique. Can. J. Electr. Electron. Eng. **2**, 598–601 (2011)
5. Sharan, T., Bhadauria, V.: Fully differential, bulk-driven, class AB, sub-threshold OTA with enhanced slew rates and gain. J. Circuits Syst. Comput. **26**, 1750001 (2017). https://doi.org/10.1142/S0218126617500013
6. Sharan, T., Bhadauria, V.: Sub-threshold, cascode compensated, bulk-driven OTAs with enhanced gain and phase-margin. Microelectron. J. **54**, 150–165 (2016). https://doi.org/10.1016/j.mejo.2016.05.009
7. Khateb, F., Kulej, T., Vlassis, S.: Extremely low-voltage bulk-driven tunable transconductor. Circuits Syst. Signal Process. **36**, 511–524 (2017). https://doi.org/10.1007/s00034-016-0329-0
8. Dubey, T., Pandey, R.: Low-voltage highly linear floating gate MOSFET based source degenerated OTA and its applications. Inf. MIDEM **48**, 19–28 (2018)
9. Sánchez-Rodríguez, T., Muñoz, F., Galán, J., et al.: Low voltage linear tunable transconductor for high speed filters. Analog Integr. Circuits Signal Process. **82**, 329–333 (2015). https://doi.org/10.1007/s10470-014-0435-5
10. Shen, D.-L., Chu, Y.-J., Chen, H.-W.: A linearized technique in an all-MOS transconductance amplifier. Microelectron. J. **43**, 1023–1028 (2012). https://doi.org/10.1016/j.mejo.2012.07.017
11. Lewinski, A., Silva-Martinez, J.: A high-frequency transconductor using a robust nonlinearity cancellation. IEEE Trans. Circuits Syst. II Express Briefs **53**, 896–900 (2006). https://doi.org/10.1109/TCSII.2006.880025

12. Kar, S.K., Sen, S.: Linearity improvement of source degenerated transconductance amplifiers. Analog Integr. Circuits Signal Process. **74**, 399–407 (2013). https://doi.org/10.1007/s10470-012-9948-y

13. Ngamkham, W., Kiatwarin, N., Narksap, W., et al.: A linearized source-couple pair transconductor using a low-voltage square root circuit. In: 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2008, pp. 701–704 (2008)

14. Elamien, M.B., Mahmoud, S.A.: On the design of highly linear CMOS digitally programmable operational transconductance amplifiers for low and high-frequency applications. Analog Integr. Circuits Signal Process. p. 2 (2018). https://doi.org/10.1007/s10470-018-1128-2

15. Rezaei, F.: Linearity enhancement in the entire tuning range of CMOS OTA using a new tune compensated source degeneration technique. Microelectron. J. **66**, 128–135 (2017). https://doi.org/10.1016/j.mejo.2017.06.008

16. Nedungadi, A., Viswanathan, T.R.: Design of Linear CMOS transconductance elements. IEEE Trans. Circuits Syst. **31**, 891–894 (1984). https://doi.org/10.1109/TCS.1984.1085428

17. Khateb, F.: Bulk-driven floating-gate and bulk-driven quasi-floating-gate techniques for low-voltage low-power analog circuits design. AEU Int. J. Electron. Commun. **68**, 64–72 (2014). https://doi.org/10.1016/j.aeue.2013.08.019

Application of Petri Net Model in High-Level Scheduling Algorithm

Atul Kumar Srivastava, Siddharth Tiwari and Shubhanker Banerjee

Abstract Various perspectives of hardware design in VLSI can be easily understood with the help of Gajski Y-Chart. In High-Level Synthesis, Control and Data Flow Graphs are used to demonstrate the behavior of hardware on the basis of HDL and its implementation, it is observed that there are very few tools on which such design can be made and even then it is not possible to visualize the dynamic behavior of the system in its pre-processing stage. Unlike Control and Data Flow Graphs, in this research a foundation to link and merging of High-level Synthesis Data Flow Graph to Petri-Nets has been made. Various affirmative conclusions have been reached. Here an effort is made to map Control and Data Flow Graph to Petri-Nets.

Keywords High-level synthesis (HLS) · Control and data flow graphs (CDFG) · Petri-nets

1 Introduction

One of the initial steps in HLS is to collect the information about transformation from behavioral of any hardware into their internal representation. Hence most of the HLS uses CDFG as the representations. A composition of both control flow and data flow of design entry. The next step is to choose the functional units and memory units from the design library as per the scheduled and optimized graphs known as Allocation. The final process is to bind the assigned operation to their functional units, data transfer to their buses and wires, and variables to storage elements, etc.

Control and data flow graphs are directed graphs which have inputs and outputs representing and mapping with RTL to form hardware. There are certain limitations on CDFGs that can be taken care during the transformation of these Graphs to Petri-nets models. Petri-nets are mathematical modeling tool. Understanding the concepts of Scheduling and developing a scheduler with the help of Petri-Nets is the goal of this paper.

A. K. Srivastava (✉) · S. Tiwari · S. Banerjee
Jaypee Institute of Technology, Noida, India
e-mail: atul.srivastava@jiit.ac.in

2 Literature Survey

Three scholars M. C. McFarland, R. Camposno and A. C. Parker did a major study in the field of high-level synthesis digital systems in 1990 [1]. The article submitted by them is a worthy study of the introduction to the topic of the same with all important aspects. There are also some recent articles by Gajski and Ramachandran [2], Lin [3] that gives a brief idea on recent research on High-level Synthesis.

A fine intro on the topics of Data Flow and Control Flow graphs and its history is given in the articles of A. L. Davis and R. M. Keller in 1982 IEEE Computer [4]. Their studies also provide us an idea of parallel computing. Similar works are done by Lee [5–7]. Cases of combining the two graphs of data-control flow graphs [8] or control-data flow graphs [9] are also given. The definitions of high-level synthesis "scheduling", "allocation" and "binding" are given by M. Potkonjak and J. M. Rabaey in their international journals of circuit theory and application 1992 [8].

The major study on the algorithms like As Soon As Possible (ASAP), As Late as Possible (ALAP), Force Directed Scheduling (FDS), and List-based Scheduling (LS) in VLSI High-level synthesis [10]. It illustrates the prime examples of task resource and constraints' DFGs with the help of different example that are efficiently used in this research [11].

Detailed study on petri-nets and its application is done with the help of a book name "modeling with generalized stochastic Petri nets" written by M. Ajmone Marsan, Gianfranco Balbo, Gianni Conte, Susanna Donatelli and Giuliana Franceschinis [11]. This is one complete book on the fundamentals of various aspects and application of petri-nets. By following of this book we will be able to get a detailed idea on the vastness of Stochastic Petri nets and Generalized Stochastic Petri nets (GSPN) [12].

A paper published by Aalst W. M. P. from Eindhoven University of Technology, Department of Mathematics and Computer Science naming "Petri net Based scheduling" is the only paper which scaled our understanding of mapping of Scheduling problems to Petri-nets and helped in a great way for developing an idea to do the same [12].

For designing and modeling of our nets and testing its authenticity we used special tools like CPN Tool and HiPS that were peptized with the help of their websites [13, 14]. HiPS tool is invented by Department of Computer Science and Engineering, Shinshu University. Similarly CPN tool [15] was developed by CPN group, Aarhus University in ten years of hard efforts. Both of tools and their specifications, uses, pros and cones and implementations are mentioned in the subsequent chapters of this thesis.

3   High-Level Synthesis

High-level Synthesis basically transforms Specifications along with the algorithms to RTL. These specifications are represented by Control and Data Flow Graphs. So CDFGs in High-Level Synthesis is one of the widely accepted tools for this. In a pre-processing step we take a CDFG and go for the optimization followed by scheduling allocation and binding. High-level Synthesis takes your preprocessed design and drives to block level design or RTL design.

A behavioral description (i.e. Functional Specifications) is used as starting point of Synthesis. It specifies behavior in terms of its operations, and control constructs in HDL [16]. The basic sequenced steps of the High-level synthesis are summed up in Fig. 1.

The HLS is compilation of Hardware Description Language and transformation into the internal representation with control and data flow graphs techniques since it contains CDFG. These processes also involve a set of compilation ways examples optimization, dead code removal, redundant expression skipper or eliminator, etc.

Fig. 1 HLS design flow
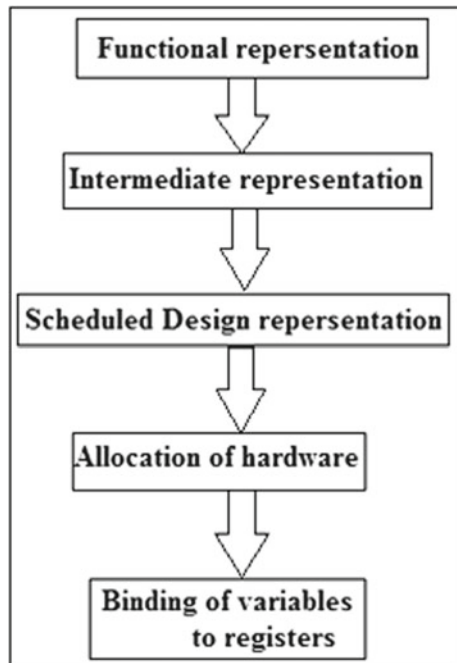
4   Control and Data Flow Graphs

Mathematically a graph is an ordered pair G = (V, E) where V is the set of nodes or vertices or points and E is the set of edges or lines or arcs The CDFGs used in high-level synthesis are having various components namely input/output nodes, conditional nodes where condition are checked and computational nodes where some operations are performed. These nodes are joined by one more component called directed arcs or directed edges.

5   Scheduling

The prime step of the high-level synthesis which generally plays a key role in transforming a Control and Data Flow Graphs to Register Transfer representation is Operational Scheduling or simply Scheduling. It involves in assigning the operations of Control and Data Flow Graphs into control steps that generally corresponds to the duration of system clock. These system clocks are basically the part of synchronous digital systems.

5.1   Scheduling Problem

Let us consider OP (o_1, o_2, o_3...) be the set of the entire operator to be scheduled that are obtained from the HDL code after synthesis. If we consider again that there is an operator OPj ϵ OP which be influenced by the result of alternative operation OPi ϵ OP, then OP_i requisite terminate its execution beforehand operation OP_j can initiate. Such instances we say that there is dependency of data between the two operations. This is the vibrant constraints before scheduling a set of operation from Hardware Description Language code i.e. to check the data dependency and make sure that the solution should not violate this. Data dependency is nothing but prioritizing the hardware to work one before another. We can also consider this as a way to sequence the operations.

For high-level synthesis platform, there happens a module library containing circuits of diverse functionalities such as adder, multiplier, register, etc. Furthermore these libraries also have information about various parameters of module that is to say occurrence, region, power and others. Let "U" be set of various modules which are obtainable. A prearranged operator "o", its function Uy: OP → U; Uy (o) = t. Infer operator "o" can function on unit of type t. Scheduling problem can be classified into four types of problems based on the some issues of constraints based on above basic formulation.

– Un-Constrained Scheduling problem (UCS): a partial order set OP is determined by precedence constraints. Therefore set of operations OP, and set U of different

types of functional modules, are unconstrained. The only constraint in such type of problem is the data dependency. As schedule is unconstrained, we need to see each of the elements of OP which are scheduled, appropriate module from U are taken and partial order is balanced i.e. data dependency.

– Time Constrained Scheduling problem (TCS): Likewise Unconstrained Scheduling problem the only thing that is additional in this is a time constraint (also known as deadline) D (i.e., maximum control steps) on the partial order on OP determined by the precedence constraints based partial order OP. Here time is the limiting factor of the problem.

– Resource Constrained Scheduling problem (RCS): Unlike the deadline constraints D as in TCS, RCS has resource constraints max ($1 \leq k \leq |U|$) for each functional module and a precedence constraints based partial order on O. Here we come across the limits of limited no of Resources in HLS design.

– Time-Resource Constrained Scheduling problem (TECS): In this problem both the deadline constraints D and the resource constraints max ($11 \leq k \leq |U|$) for every functional module together with the general partial order on OP of precedence constraints is present. Here both times as well as resource are the limiting factors of a scheduling problem.

5.2 Scheduling Algorithms

Depending up on the types of scheduling problem scheduling algorithms involuntarily assign control steps to operators with in design constraints criterion. For the past few years numerous algorithms were projected in the previous research that solves the above-mentioned types of scheduling problems. These algorithms are categorized into *heuristic* and *deterministic*. In practice the deterministic algorithm such as integer linear programming provides optimal scheduling solution but eat high processing time. If the execution time is not an issue then several algorithms based on greedy strategies have been proposed that make a series of local decisions, selecting at every point the single best operational-control steps pairing without back tracking or look ahead. This method may not reach the global optima solution but they can provide the solution quickly. Generally those solutions are very close to global optimal solution therefore we use these approaches widely and are in practice. Such algorithms are called heuristic algorithms. Examples of such kind of heuristic algorithms are—As Soon As Possible, As Late As Possible, Force Directed Scheduling, List Scheduling.

The best way to differentiate these algorithms is to provide a scheduling problem and see the approach through which they optimize the graphs. Let us consider a HDL code given to be as

$$Begin$$
$$x = [(a + b)/(c - d)] * a - [(e + f)/b);$$
$$y = (g * b) + f;$$
$$End$$

ASAP is one of the simplest scheduling approaches. Here the maximum numbers of control steps that are allowed are first determined. This algorithm schedules each operator at a time into the earliest possible steps keeping the partial order i.e. data dependency into consideration. Hence, for making this algorithm work for the given HDL codes it requires two multipliers, two divisors, three adders, and two subtractors.

6 Petri-Nets

The current prompt progresses in various technological fields have resulted to extraordinary increases in system intricacy in several uses. A representative illustration of this progression is provided by the field of computer engineering, where the advances in the fields of VLSI and software scheme have led to multifaceted scattered computing systems whose competences majorly surpass those of outdated computers. The operative utilization of the possibilities offered by technological novelties necessitates satisfactory tools to pronounce, to model, and to analyze the various design substitutes. In precise, since system (performance is regularly the crucial goal behind augmented system complexity, tools that permit the comparative virtues and the intrinsic class of the possible explanations to be quantified are seriously needed. Many different standards for the portrayal of complex systems have recently been established, but only a few of these allow the incorporation of routine analysis concepts within the description itself. Petri nets are one of the few archetypes allowing this incorporation. Petri Nets (PNs) are a graphical tool for the correct portrayal of systems whose dynamics are categorized by concurrency, synchronization, mutual exclusion, and conflict, which are distinctive topographies of distributed settings. PNs incorporate a concept of (distributed) state and a rule for state change that permit them to capture both the static and the dynamic properties of any present system [17]. PNs can be fruitfully functional in quite a varied variety of fields, from distributed computing in its architectural, algorithmic, hardware, and software components to malleable manufacturing, communiqué protocols, control systems, transportation, banking, and association of work.

In this paper we are mainly concerned in the use of PNs for concert estimation of distributed systems by means of simulation methods. Also we will be explaining mapping techniques to map the CDFGs to Petri-nets. As mentioned, Petri-Nets can

play a significant role in the life cycle of a distributed system, starting from the early phases of the design, inside the system development and to the conservation during its operational stage.

Historically speaking PNs were originally introduced by Carl Adam Petri within his doctoral thesis to define concurrent systems in terms of cause and effect relations deprived of explicit time respects. The overview of progressive concepts into PN models was suggested some years in future by C. Ramchandani, P. M. Merlin and D. J. Farber, and J. Sifakis, with diverse tactics. A variety of different proposals followed, typically based on the custom of deterministic timing. The first definitions of PNs with stochastic timing are by reasons of F. Symons, G. Florin, S. Natkin and M. Molloy. These proposals unlocked the possibility of connecting PNs to the field of performance evaluation, conventionally based on a stochastic modeling method. Such models, and their offspring, are nowadays cooperatively named Stochastic PNs (SPNs). An allowance of the approach proposed by Molloy was provided, where stochastic timing is motley with deterministic null delays, so that both the temporal as well as the logic progression of a system can be described within one model. This then resulting in modeling pattern, named Generalized SPNs (GSPNs).

6.1 Components of Petri-Nets

A PN includes *places, transitions, and arcs* what can be expressing as its essential components. Places are in the habit of describing possible system states (named conditions or situations). Arcs lay down the similarity between restricted states as well as actions in two types: they are indicative of the local state in which the event can happen, and the local state renovations affected by the event. Transitions designate events which adjust system states.

Places can be represented by circles or ovals, Transition is represented by bars or rectangle and arcs is of two types directed and inhibitor arcs and they are represented by directed lines or lines having end with dots respectively. Figure 2 shows all the three components.

A straightforward sample of a PN model is given in Fig. 3a, b, where we have places *p on* and *p off* and transitions *tswitch_on* and *tswitch_off* are associated with four arcs. Places describe conditions (that we may call the "on condition" and "off condition"). Status shown in given diagram is shows that place *p off* contains one
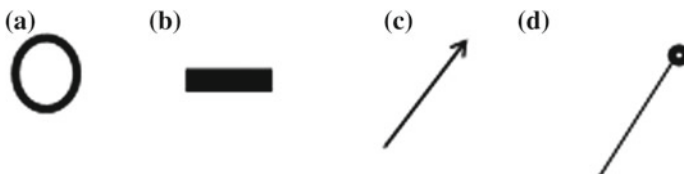
Fig. 2 a Place, **b** transition, **c** directed arcs and **d** inhibitor arcs
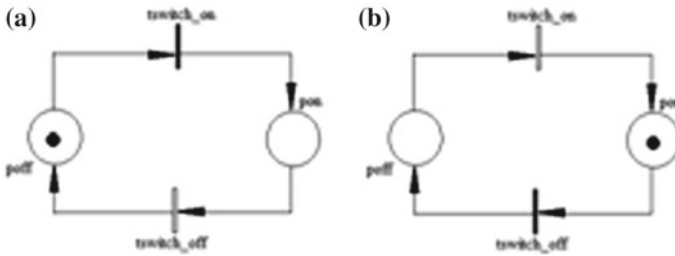
Fig. 3 **a**, **b** Simple switch PN model with different stages

token; and the "off condition" is true and place p on is empty, thus the "on condition" is not true. In the given example PN model given in Fig. 3a, b, transition tswitch_on is not disabled, it can fire, eliminating one token from p off, placing one token in p on, so that the new state results in the "on condition" being true and the "off condition" is not true. In the resultant state, transition tswitch_off is not disabled, it can fire, restoring the state shown in Fig. 3a, b. This simple PN model canister clearly is understood as the PN explanation of the conduct of a simple switch. Usually, the firing of a transition defines the result of whichever a reasonable condition flattering true in the system, or the finishing point of an activity. The second understanding is the reason for connecting timing with transitions, as several authors did in their proposals for the definition of temporal concepts in PNs.

6.2 Dynamic Behavior of PN

The dynamic behavior of PN is administrated according to firing rule. Transition can fire, if all transition input places, comprise at least unity token. Such cases have transition enabled. The firing of such a transition takes away one token at each of its input places, and creates one token in each of its output places [18]. When arc weights more than one are used, the number of tokens essential in each input place for the transition being enabled and the number of tokens produced in each output place by the transition firing are resolute by the weight of the arc connecting the place and the transition. The firing of a transition is an infinitesimal operation. An "enabling rule" and a "firing rule" are related with transitions. The enabling rule states the conditions in which transitions are allowed to fire. The firing rules undergo change in the marking which is induced by the transition firing. Informally, we can say that the enabling rule defines the conditions that permit a transition t to fire, and the firing rule specifies the change of state produced by the transition. Both the enabling and the firing rule are quantified through arcs. In particular, the enabling rule encompasses input as well as inhibitor arcs, whereas the firing rule hinge on input and output arcs. Note that input arcs play a dual role, since they are tangled both in enabling as well as in firing. A transition t is enabled if and only if

– Each input place comprises a number of tokens more or equal than a certain threshold, and
– Each inhibitor place contains a number of tokens strictly smaller than a given threshold.

6.3 Petri-Net Model Definition

A way in which Petri net model can be defined is as following:

A PN model is an 8-tuple
M = {P, T, I, O, H, PAR, PRED, MP};
where P is the set of places. T is the set of transitions, $T \cap P = \phi$. I, O, H: T $\rightarrow$ Bag(P), are the input, output and inhibition functions, respectively, where Bag(P) is the multiset on P.
PAR is a set of parameters.
PRED is a set of predicates restricting parameter ranges.
MP: P $\rightarrow$ IN U PAR is the function that associates with each place either a natural number or a parameter extending on the set of natural numbers. Functions I, O, and H describe the input, output, and inhibitor arcs of transitions, correspondingly.
Assumed a transition $t \in T$, we denote with
•t = {p $\in$ P: I (t, p) > 0} and with
t• = {p $\in$ P: O (t, p) > 0} the input and output sets of transition t and with
∘t = {p $\in$ P: H (t, p) > 0} the inhibition set of transition t. Since Bag (P) can be seen as a function from P to IN, then both the following are syntactically correct:

– I(t), in lieu of the multiset of input places of transition t
– I(t, p) denoting the multiplicity of element p in the multiset I(t) For convenience, I, O, and H can also be represented by |T| × |P| matrices of natural numbers.

Formally, a marking is M: P $\rightarrow$ IN, and we refer to M(p), for p $\in$ P, as the number of tokens in place p in marking M. For suitability, the marking can also be viewed whichever as a |P|-component vector, or as an element of Bag(P).

6.4 Analysis Techniques of Petri-Nets

State space analysis techniques are fundamentally nonparametric in comparison to the initial marking, since they necessitate its comprehensive instantiation. The reachability analysis is based on the construction of the Reachability Graph (RG) of the PN, and it is practicable only when the Reachability Set (RS) and the RG are predetermined. Even if the RS and RG are finite, their magnitude can vary usually with

the PN structure and the number of tokens in initial marking M0. Innumerous cases, the progress in the number of markings can be combinatorial equally in the number of places and in the number of tokens in M0. Reachability investigation methods are very powerful, since they permit the proof of most properties of attention by inspection, as the RS and RG contain all conceivable evolutions of the PN system. However, it often happens that the space and time intricacy of the RG construction algorithm surpasses any acceptable boundary. Once the RG has been built, properties may be tested using classical graph analysis algorithms. Let us consider disjointedly each property discoursed in this section.

- Reachability: To test whether a marking M0 is reachable from a marking M in a specified PN system, it is necessary to check if RG contains a directed path joining M to M′. The concatenation of the tags of the arcs making the path gives the transition order.
- Reversibility: To check whether a given marking M is a home state, it is sufficient to build the set of all the markings that can be reached by following backwards in the RG any path entering M, and to verify that this set is RS (M). A PN system is reversible if and only if its initial marking is a home state. This is true if and only if the RG has a unique strongly connected component that contains M0.
- Absence of deadlock: The occurrence of a deadlock can be checked on the RG by observing for a dead marking, i.e. a node with no output arcs. Then on appearance of such a node is sufficient assurance of the absence of deadlock.
- Liveness: Transition t of a PN system is live if and only if from every single marking M ϵ RS it is likely to reach a new marking M′ such that t ϵ E(M′). t is live if and only if

  The RG contains no dead marking
  – t labels some arc of every strongly connected component of RG.

- Boundedness: A given place p ϵ P of a PN system is n-bounded if and only if n = maxM(p) M ϵ RS

Of course this can be computed only for finite reachability sets. These are some of the basic properties of Petri-Nets in very brief. If anyone wishes to dig deep into their properties kindly refer to the reference links.

7 Simulation Result

Simulation of Task Resource and Precedence example of adder is shown in Fig. 4a–c.
Conditional data flow Simulation is shown in Fig. 4b. Iterative construct Simulation result as shown in Fig. 4c.

(a)

(b)

(c)
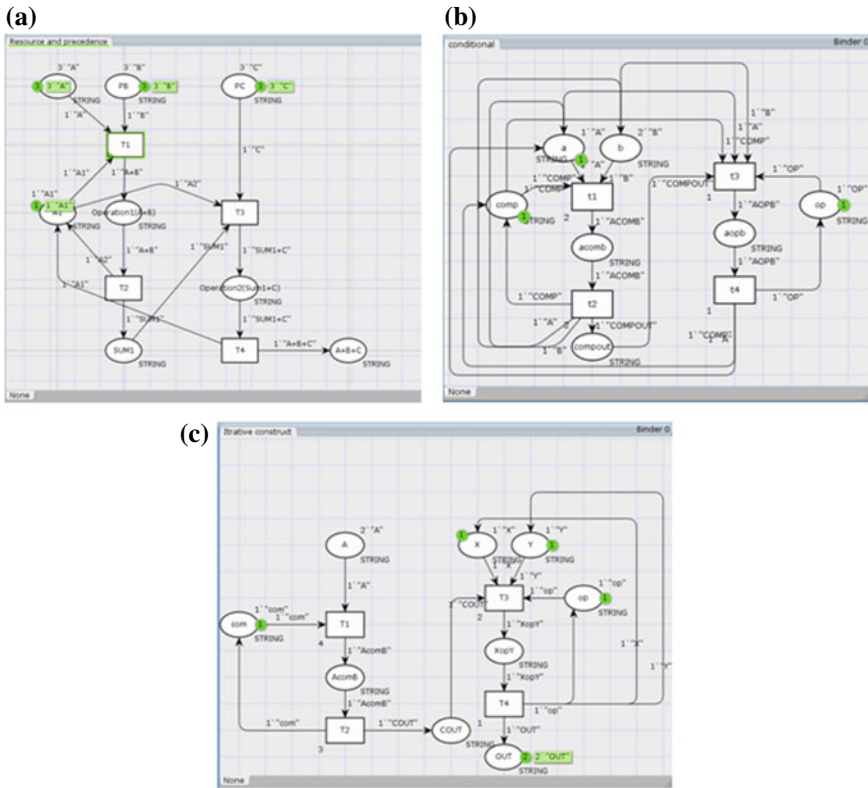
Fig. 4 **a** Simulation of task resource and precedence example of adder. **b** Different stage of simulation of condition net. **c** Different stages of simulation iterative construct

8   Conclusion and Future Scope

Here are some of the major conclusions we come across when we did the mapping, modeling and analyzing the High-level Synthesis data flow graph using petri-net on their two respective tools. By the use of modeling steps we are able to model the CDFG to petri-nets. Petri-nets demonstrate the dynamic behavior of a pre-processing stage of HLS which was not possible in CDFG. We need various numbers of assumptions when we go for CDFG rather than PN. Petri-net, with its property of bipartite connectivity in its constituent components, illustrates more precisely the events and changing of the state of a design from one to another in a proper sequential manner. We were able to peptize tool which are able to model, simulate and analysis of Petri-nets. There are certain Pros and Cons in HiPS versus CPN tool. CPN tool concentrates over the state space analyses of a designed model only. Hips can have various types of analyses including Structural and behavioral analysis. The major Drawback of using HiPS Tool is that priority of the transition cannot be set. Since

priority of transition can't be set it cannot be used for complex and multiple transitions and it will not have a precise design. With the help of examples like Conditional, Iteration and task resource we are able to map and model basic High-level synthesis codes which are used to formulate RTL for further use. The state space analysis can be made with the help of CPN tool and various attributes of the analysis like Liveness, Boundness, and Deadlocks can be analyzed efficiently. For the drawing of CDFG and its analysis there are a few tools available and they are not so promising to use. Unlike this Petri-net tools like CPN Tools can be used and learned very easily. Petri-Net tools can illustrate the movement of token in the design very efficiently but there are hardly any such tools for developing graphs of CDFG and executing this with each step of stage. Since the dynamic behavior of a petri-net can be easily visualized, it is recommended that one should drift and convert/map their design of CDFG to Petri-Net. Since PN is a bipartite graph, we have seen that there will be very less problem in putting scheduling algorithms on to the design. A whole new world of possibility for the designers can develop if they can use Petri-net as their tool for pre-processing of HLS.

References

1. McFarland, M.C., Parker, A.C., Camposano, R.: High level synthesis of digital system. Proc. IEEE **78**(2), 301–318 (1990)
2. Gajski, D.D., Ramachandran, L.: Introduction to high level synthesis. IEEE Design Test Comput. 44–54 (1994) (Winter)
3. Lin, Y.L.: Recent development in high level synthesis. ACM Trans. Design Autom. Electron. Syst. **2**(1), 2–21 (1997)
4. Davis, A.L., Keller, R.M.: Data flow program graph. IEEE Comput. 26–41 (1982)
5. Lee, E.A.: Recurrence, iteration and conditionals in statically scheduled block diagram languages. In: Brodersen, R.W., Moscowitz, H.S. (eds.) VLSI Signal Processing III, pp. 330–340. IEEE Press, New York (1988)
6. Lee, E.A.: Consistency in data flow graphs. IEEE Trans. Parallel Distrib. Syst. **2**(2), 223–235 (1991)
7. Lee, E.A., Parks, T.M.: Data flow process networks. Proc. IEEE **83**(5), 773–799 (1995)
8. Potkonjak, M., Rabaey, J.M.: Scheduling algorithms for hierarchical data control flow graphs. Int. J. Circuit Theory Appl. **20**, 217–233 (1992)
9. Paulinand, P.G., Knight, J.P.: Force directed scheduling for the behavioral synthesis of ASIC's. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **8**(6), 661–679 (1989)
10. Gerez, S.H.: Algorithms for VLSI Design Automation. High Level Synthesis, pp. 253–260 (1998)
11. Ajmone Marsan, M., Balbo, G., Conte, G., Donatelli, S., Franceschinis, G.: Modelling with Generalised Stochastic Petri Nets (1995)
12. Lectures on Formal Methods and Performance Analysis. Springer Nature America, Inc (2001)

13. Aalst, W.M.P.: Eindhoven University of Technology, Department of Mathematics and Computer Science, Petri-Net based Scheduling, pp. 1–10
14. HiPS Tool: User manual for (en) for 1.x, 2 Mar 2016, at 03:16; http://hips-tools.sourceforge.net/wiki/index.php/UsersManual(en)for1.x
15. CPN Tool: CPN tool. Eindhoven University of Technology. http://cpntools.org/start (2015)
16. Ferrandi, F., Lanzi, P.L., Palermo, G., Pilato, C., Sciuto, D., Tumeo, A.: An evolutionary approach to area-time optimization of FPGA designs. In: 2007 International Conference on Embedded Computer Systems: Architectures, Modeling and Simulation (2007)
17. Design of it Infrastructures of Data Centers: An Approach Based on Business and Technical Metrics: Quantitative Assessments of Distributed Systems (2015)
18. Rezvanian, A., Saghiri, A.M., Vahidipour, S.M., Esnaashari, M., Meybodi, M.R.: Recent Advances in Learning Automata. Springer Nature (2018)

Design of Full Adder with Self-checking Capability Using Quantum Dot Cellular Automata

Shahneela Jamal Kidwai, Divya Tripathi and Subodh Wairya

Abstract Conventional CMOS Technology is losing its efficiency to fulfil the needs of this technically super advanced era in which we need devices with lesser power, lesser area and higher speed than ever. CMOS will reach its limitation due to concepts of fundamental physics. If we further decrease the size of transistor, effects of quantum physics will come in picture and limit further miniaturization of device size. Now designers are shifting towards an emerging technology based on Polarization of Electrons that is Quantum Dot Cellular Automata. QCA gives efficient results in area reduction and power reduction to ultra-low level. How Design of Parity-preserving Reversible Gate is optimized than this Gate is used to make Full Adder is explained in the paper. In transmission of digital data the process of error detection becomes decisive. In this gate input parity is always equal to output parity so no extra testing circuitry is needed. The Proposed PPRG requires 8.77% lesser cell count and 4.54% lesser area and proposed Full Adder has 6.29% lesser cell count and 4.90% lesser area when compared to existing design

1 Introduction

Conventional CMOS has dominated our fabrication industry for last few decades and it has proved to be a better alternative than earlier technologies but very soon a day will come when our conventional CMOS technology of device designing will reach its limitation and we will have to migrate to a new technology. Quantum Dot Cellular Automata shows all signs of becoming powerful and better alternative to Conventional CMOS Technology. In the year 1993, C S Lent introduced Quantum

S. J. Kidwai (✉) · D. Tripathi · S. Wairya
Institute of Engineering and Technology, Lucknow, India
e-mail: shahneelakidwai@gmail.com

D. Tripathi
e-mail: divyatripathi.cest@gmail.com

Dot Cellular Automata and in the year 1997 QCA was physically verified [1]. CMOS Technology is hard to upgrade further because of limitations posed by short channel effect and ever decreasing size of gate oxides at nano scale. Also there are many more scaling restrictions in CMOS technology. Numerous researches are working on designing nano scale devices using alternative techniques such as Quantum Dot Cellular Automata (QCA). It is interesting because of its attractive characteristics of high speed performance, low power consumption and small dimension [2]. Use of Fault-Tolerant Full Adders is inevitable in digital transmission where accuracy plays an important role. In this paper first a Parity-preserving Reversible Gate is modified using an efficient XOR gate then using this Parity Gate design of Full Adder is optimized. Existing Parity-preserving Reversible Gate and Full Adder has been taken from [3]. XOR Gate has been taken from a paper which uses very less number of cells [4]. The functionality of PPRG, Full Adder and XOR Gate has been verified on QCA designer. Proposed Parity-preserving Reversible Gate and Full Adder are designed and simulated on QCA Designer.

2 QCA Overview

Basic functioning of QCA is performed by QCA Cell [5]. QCA cell transfers the data from input side to output side and also performs the computation. QCA cell is a conjectural square-shaped space in which there are four places where electrons can reside. Possible potential space for electron is represented by a dot inside the QCA cell. The dots are coupled through quantum mechanical tunnelling barriers and electrons tunnel through them depending on the state of the system. Two electrons reside in each cell. In that hypothetical square space electrons occupy diagonal position which is farthest possible distance between them due to inter columbic repulsion. In QCA cell, polarization means relative position of electrons inside the cell. Two types of polarizations are there in a QCA cell 90° and 45° shown in Fig. 1. Binary '1' and Binary '0' are represented by cell polarization. QCA cell gets the power to process the input from clock.

(a) Polarisation '1'   (b) Polarisation '0'   (c) Null State

Fig. 1 QCA cell at 90°

Transmitting binary '0' in QCA wire
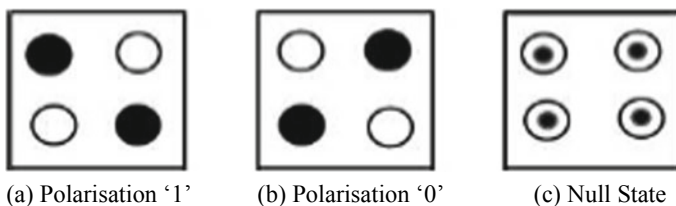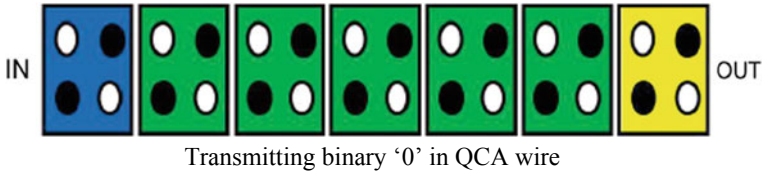
Fig. 2 Information transfer in QCA wire

2.1 QCA Wire

By placing two cells adjacent to each other and forcing the first cell into a certain state, the second cell will assume the same state in order to lower its energy. Many QCA cells are arranged in a cascaded manner to make a QCA Wire. Attraction and repulsion of electrons occur due to columbic force between the adjacent cells which causes the polarization of a cell so that it can align according to its neighbouring cells so the transmission of information along the array of cells occur. QCA wires transmitting binary "1" using 90° cells in shown in Fig. 2 here the input cell is driven by an external source and is strongly polarized in one direction. By stringing cells together in this way, a "pseudo-wire" can be made to transport a signal, in contrast to a real wire, however no current flows. The input cell drives other cells in the NULL state which tend to align themselves to the input cell polarization to reach the system's ground state [1].

2.2 Majority Voter

Majority Voter can be considered as basic building block for most of the QCA. 3 input Majority voters consist of five QCA cells in which output comes according to the input which is in Majority. The output F of a majority voter is defined as (1). Truth table of majority voting scheme is shown is listed in literature [6, 7].

$$F = AB + BC + AC \qquad (1)$$

2.3 Majority Gate

Another primary gate in this tool is QCA majority gate. There are two types of majority gates in general.

(1) Original Majority Gate (OMG)
(2) Rotated Majority Gate (RMG)
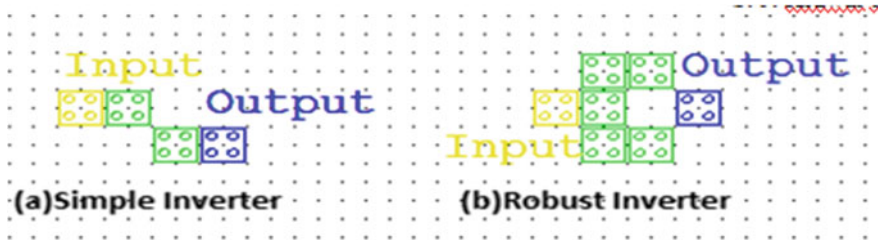
Fig. 3 Basic inverter layout
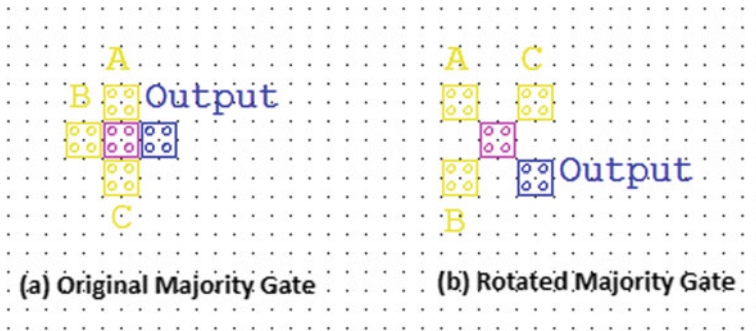
Fig. 4 QCA majority gate

QCA majority gate is made of five QCA cells with a cross-shaped structure. The polarity of the central cell, too recognized as device cell is enforced, via the coulomb repulsion to be equal to the output cell in OMG and opposite to the output cell in RMG. Thus, a combination of the QCA majority gate and an inverter is sufficient to get a complete logic set for designing any circuit [8] (Figs. 3 and 4).

2.4 QCA Clocking

The data stream in QCA circuits is overseen and controlled by the clock. Timing gives the power pick up and maintains a strategic distance from the meta-stable states [9, 10]. Regarding QCA cell the meta-stable state compares to polarization of a cell that cannot be particularly recognized as rationale 1 or rationale 0. The check in QCA innovation is not same as they are in conventional CMOS circuits, QCA timing plan comprises of four stages: switch (unpolarized cells are driven by some information and get energized relying upon their neighbour's polarization), hold (cells are held in same twofold state with the goal that it can be utilized as a contribution to different cells), discharge (boundaries are brought down and cells

Fig. 5 Four phases of QCA clock
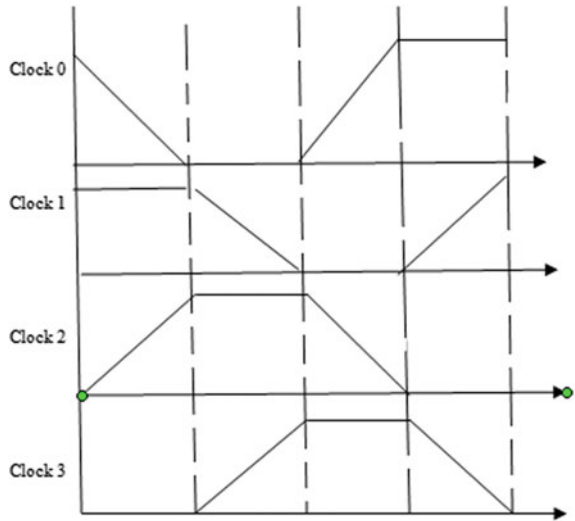
move toward becoming unpolarized) and Relax (cells remain unpolarized) [10, 11]. There is a stage distinction of quarter cycle in all these timing stages which can be actualized by creating four timekeepers each with π/2 stage contrast from past one appeared in Fig. 5.

2.5 QCA Designer

QCA Designer is a tool to simulate complicated QCA circuit layouts. ATIPS Laboratory, University of Calgary, developed the QCA Designer for the first time. There are 3 simulation engines present in the current version of QCA Designer. All three simulation engines have their own specifications but each of them is able to perform an exhaustive verification of the system or a set of user selected vectors [12, 13].

2.6 Reversible Gates and Reversibility

Logic gates which we use in conventional digital designs find themselves incapable of restoring inputs from outputs. Power will be lost in a circuit if the information is lost in processing from input to output. This fact was proven by Landauer and Bennett in their research. If we are unable to reproduce input with the help of output information will be lost in such case. Reversible gates are those which have same number of inputs and outputs and do not loose information. Landauer proved that computations which are not done by reversible circuits loses KT * ln2 Joules of heat

energy on each bit loss of information. Functioning and design of many Reversible gates such as Fredkin Gate [14], Toffoli Gate [15] and Peres Gate [16] is explained in literature. A Reversible gate of N input and N outputs can be represented as Eqs. (2) and (3). Where input vectors are represented by I_v and output vectors are represented by O_v.

$$I_v = I_1, I_2, I_3, I_4, \ldots, I_N \tag{2}$$

$$O_v = O_1, O_2, O_3, O_4, \ldots, O_N \tag{3}$$

3 Proposed Full Adder and Presentation

In this section design of a Parity-preserving Reversible Gate is modified using a more efficient XOR [4] gate and by altering clocking and then this modified gate is used to make a Full Adder. The proposed PPRG and Full Adder are more efficient than previous design.

$$P = A \oplus B$$
$$Q = AC + B'C'$$
$$R = A'C + B'C'$$

3.1 Parity-Preserving Reversible Gate

Existing PPRG [3] had more number of cells so the complexity was more. XOR, Multiplexer and Inverters are utilized in implementing the Boolean expressions. Input Parity is equal to output parity so that any error in transmission or computation can easily be detected by observing the inputs and outputs. Number of inputs is also equal to number of outputs which gives this gate feature of Reversibility. Two dimensional wave clocking is used in designing the layout [17]. Conventional 1-D clocking mechanism [18] could not provide efficient results in lesser area. So a new type of clocking is proposed in literature [19] which reduces computational time and simplifies the layout considerably. The given clocking scheme partitions the circuitry along two dimensions and utilizes parallel execution and computation in clocking zones within a different timing framework. Truth Table [3] and Boolean expressions of PPRG is given below. Layout of proposed PPRG and its simulation waveform is shown in Figs. 6 and 7 respectively. A different XOR is used in this PPRG and

Fig. 6 Existing PPRG [3]
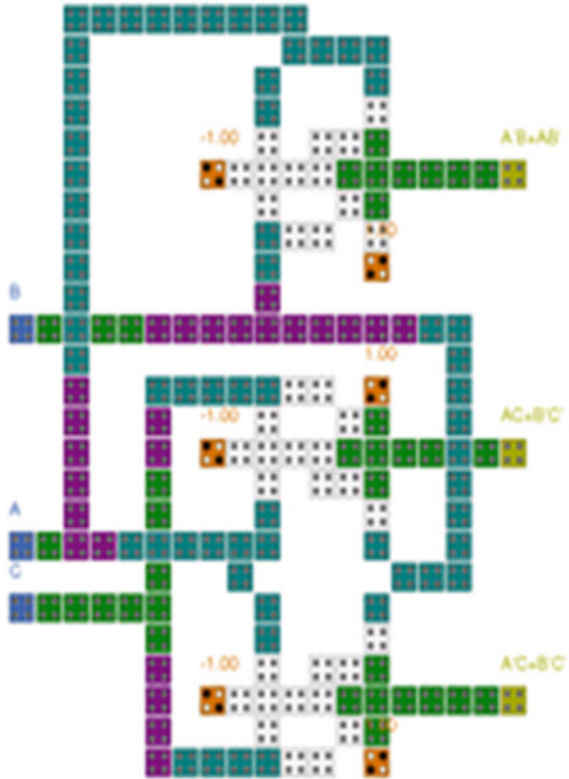
clocking technique is changed slightly in pre existing layout shown in Figure [5, 6] (Figs. 8 and 9) (Table 1).

3.2 Full Adder Using PPRG

Use of Full Adder is inevitable in computational systems and processors where it is used to calculate Memory Address, table indices, etc. Modified PPRG is used in such a way so that it provides functionality of Full Adder with reduced complexity and lesser number of cells. Connections of subsequent PPRG's are taken from literature [3] and are shown below. PPRGs used to design Full Adder are connected in cascaded fashion and parity of input set will always be equal to parity of output set so fault can be detected by observing input and output. Fault detection procedure is explained in [3]. Layout and Simulation waveform of proposed Full Adder is shown in Figs. 10 and 11 respectively. In output waveform of Full Adder, there is a latency of 3.
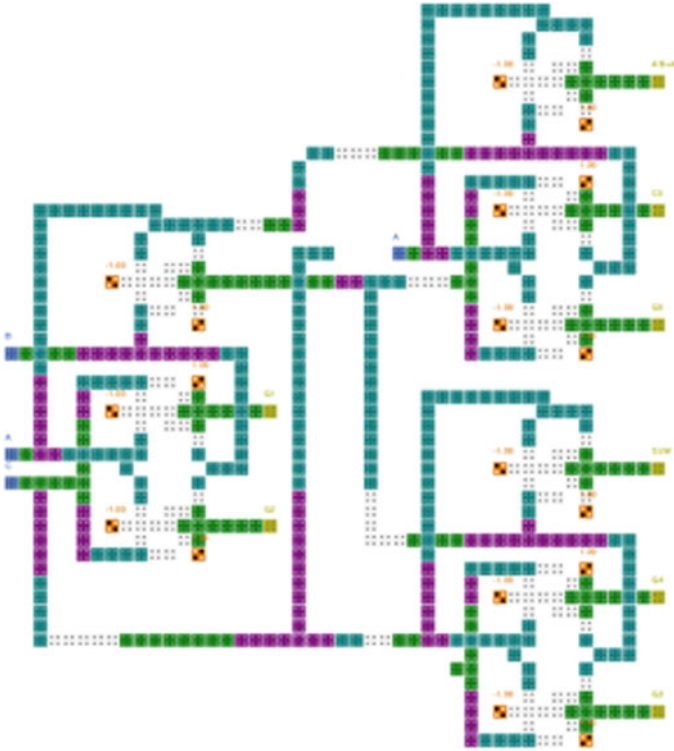
Fig. 7 Exisitng full adder [3]

PPRG1: *First Input = A, Second Input = B, Third   Input = C*
         Output P = P1 = A $\oplus$ B
PPRG2: *First Input = C, Second Input = A, Third Input = P1*
         *Output Q = **Q2** = C.P1 + A.P1*
            = C.(A $\oplus$ B) + A.(A $\oplus$ B)
            *= A.B.C +A.B.C + A.B = A.B + B.C + A.C = **Cout***
PPRG3: *First Input = P1, Second Input = C, Third Input = not used*
         P3** = P1 $\oplus$ C = A $\oplus$ B $\oplus$ C = **SUM

4 Result

For the verification and simulation 2.0.3 version of QCA Designer [19] has been used and bistable approximation engine is used. Dimension of the cell is 18 × 18 nm and the diameter of the quantum dot is 5 nm. To test the functionality of proposed PPRG which is the building block of Full Adder various input vectors are taken and
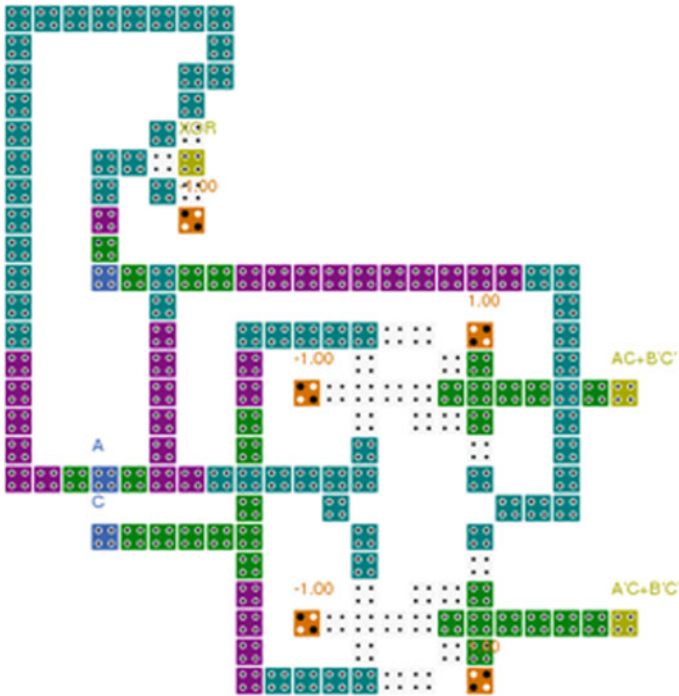
Fig. 8 QCA layout of optimized PPRG

output is verified. Proposed Full Adder circuit is also verified on QCA Designer v2.0.3. Proposed PPRG requires 8.77% lesser cell count and 4.54% lesser area and Full Adder has 6.29% lesser cell count and 4.90% lesser area when compared to existing design (Fig. 12) (Table 2).

5 Conclusion

As logic gates are regimental for most of digital circuits, having high speed, less complex and reduced area designs are significantly imperative. Gate designs which save energy by using alternative emerging technologies allow extensions in beneficial directions by utilizing concept of reversibility and by including inherent fault observability. The availability of appropriate clocking mechanism in QCA enables the utilization of parity-preserving gates. Proposed PPRG and Full Adder design saves significant amount of energy as power in QCA circuits is provided by Clocks and number of cell count is reduced here. A reversible full adder which is fault-tolerant is made using 581 QCA cells which occupies only 1.02 μm^2 area which proves its utility in the area of nanotechnology. Also the PPRG which has only 156 QCA cells and has 0.21 μm^2 area is made which can be used to realize almost all
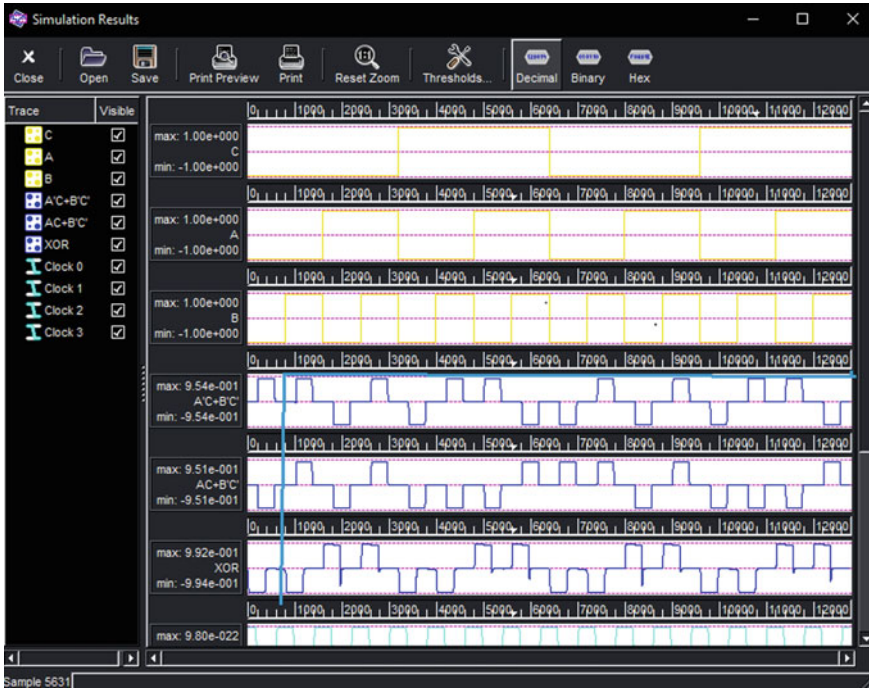
Fig. 9 Simulation result of optimized PPRG

Table 1 Truth Table of PPRG

A	B	C	P	Q	R
0	0	0	0	1	1
0	0	1	0	0	1
0	1	0	1	0	0
0	1	1	1	0	1
1	0	0	1	1	1
1	0	1	1	1	0
1	1	0	0	0	0
1	1	1	0	1	0

Boolean functions. Apart from lesser energy, concept of reversibility is also utilized. More efficient clocking will allow further improvement and applications of this type of parity-preserving gates. In gain, obtained results, check with the corresponding truth table. Still a lot of research and attention is needed in this direction to enter in the era of quantum computing.
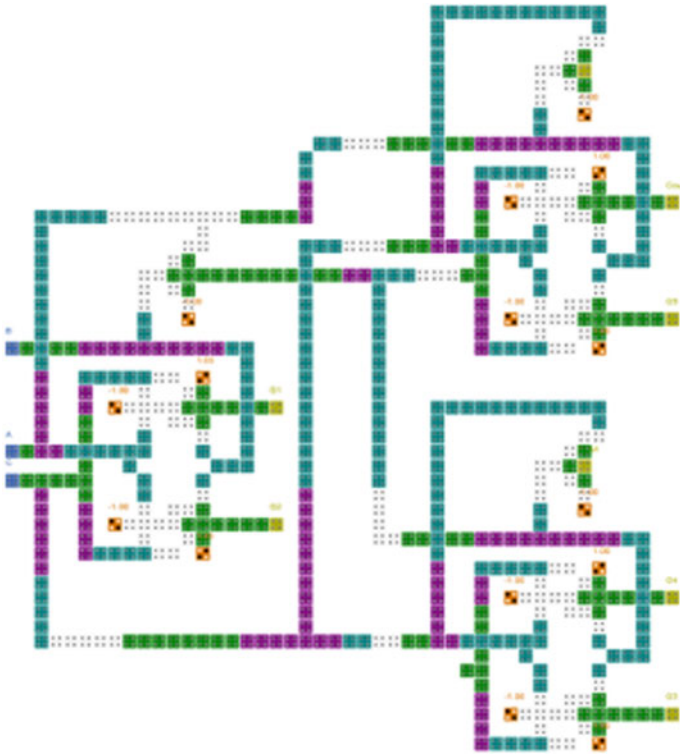
Fig. 10 QCA layout of proposed full adder
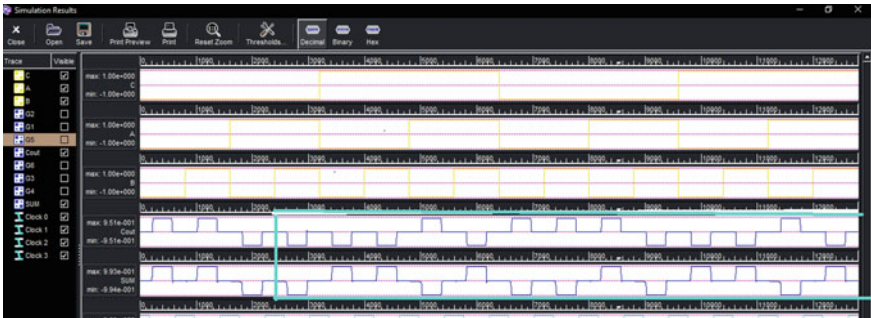
Fig. 11 Simulation result of proposed full adder

Fig. 12 Comparative representation of previous design with proposed design
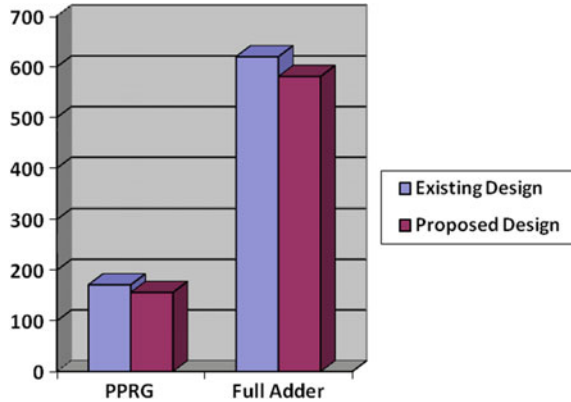
Table 2 Comparison of proposed design with previous one

Layout	Cell count	Cell count reduced	Area μm^2	Area reduced
PPRG [2]	171		0.22	
Optimized PPRG	156	8.77%	0.21	4.54%
Full adder [2]	620		1.02	
Proposed full adder	581	6.29%	0.97	4.90%

References

1. Tougaw, P.D., Lent, C.S.: Logical devices implemented using quantum cellular automata. J. Appl. Phys. **75**(3), 1818–1825 (1994)
2. Kassa, Sankit R., Nagaria, R.K.: A novel design of quantum dot cellular automata 5-input majority gate with some physical proofs. J. Comput. Electron. **15**(1), 324–334 (2015)
3. Roohi, A., Zand, R., Angizi, S., Demara, R.F.: A parity-preserving reversible QCA gate with self-checking cascadable resiliency. IEEE Trans. Emerg. Top. Comput. (2016)
4. Bahar, A.N., Waheed, S., Hossain, N., Asaduzzaman, M.: A novel 3-input XOR function implementation in quantum dot-cellular automata with energy dissipation analysis. Alex. Eng. J. (2017)
5. Niemier, M., Kogge, P.: Problems in designing with QCAs: Layout = Timing. Int. J. Circuit Theory Appl. **29**(1), 49–62 (2001)
6. Walus, K., Schulaf, G., Jullien, G.A., et al.: High level exploration of quantum dot automata. In: Proceedings of the IEEE Nanotechnology Conference, vol. 2, pp. 30–33 (2004)
7. Walus, K., Schulaf, G., Julliaen G.A., et al.: Circuit design based on majority gates for application with quantum dot cellular automata. In: Proceeding of the IEEE Nanotechnology conference, vol 4, pp. 1350–1351
8. Kassa, Sankit R., Nagaria, R.K.: An innovative low power full adder design in nano technology based quantum dot cellular automata. J. Low Power Electron. **12**(2), 107–111 (2016)
9. Lent, C.: Aquinas: a quantum interconnected network array simulator. In: Proceedings of Fifth International Workshop on Computational Electronics
10. Niemier, M., Kogge, P.: Problems in designing with QCAs: Layout = Timing. Int. J. Circuit Theory Appl. **29**(1), 49–62 (2001)

11. Niemier, M., Rodrigues, A., Kogge, P.: A potentially implementable FPGA for quantum dot cellular automata. In: 1st Workshop on Non-Silicon Computation, NSC-1, vol. 69, pp. 38–45 (2002)
12. Walus, K., Dysart, T., Jullien, G., Budiman, R.: QCADesigner: a rapid design and simulation tool for quantum-dot cellular automata. IEEE Trans. Nanotechnol. **3**(1), 2629 (2004)
13. Walus, K., Schulhof, G., Jullien, G.A.: High level exploration of quantum-dot cellular automata (QCA). In: Conference Rec. 38th Asilomar Conference on Signals, Systems and Computers, vol. 1, pp. 3033 (2004)
14. Fredkin, E., Toffoli, T.: Conservative logic. Int. J. Theor. Phys. **21**, 219–253 (1982)
15. Toffoli, T.: Reversible computing. Technical Report Tech Memo MIT/LCS/TM-151, MITLab for Computer Science (1980)
16. Peres, A.: Reversible logic and quantum computers. Phys. Rev. A Gen. Phys. **32**(6), 3266–3276 (1985)
17. Vankamamidi, V., Ottavi, M., Lombardi, F.: Two-dimensional schemes for clocking/timing of QCA circuits. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **27**(1), 34–44 (2008)
18. Lent, C.S., Tougaw, P.D.: A device architecture for computing with quantum dots. Proc. IEEE **85**(4), 541–557 (1997)
19. Thapliyal, H., Ranganathan, N., Kotiyal, S.: Reversible Logic Based Design and Test of Field Coupled Nano computing Circuits, pp. 133–172. Springer, Berlin, Heidelberg (2014)

A Novel Approach for Reversible Realization of 4 × 4 Bit Vedic Multiplier Circuit

Vandana Shukla , **O. P. Singh** , **G. R. Mishra** and **R. K. Tiwari**

Abstract The availability of fast and efficient processing systems is the basic requirement of current era. In digital systems, multiplications is one of the major operations, which limit the speed and efficiency of the system. This paper describes a novel approach for the Reversible realization of 4-Bit Vedic multiplier circuit with optimized performance parameters. Vedic multipliers are based on the concept of Vedic mathematics. It is a very fast multiplier, as it generates all the partial products and their sum in single step only. Moreover, designing of this multiplier using reversible approach will lead to the low loss fast multiplier circuits for digital systems. Some parameters indicating performance of the circuit, such as number of gates (TG), constant inputs (CI), garbage outputs (GO) and quantum cost (QC) of proposed multiplier design is also compared and analyzed with the earlier designs.

Keywords Vedic multiplication · Reversible logic gates · Reversible logic approach · Urdhva Tiryakbhyam · Ripple carry adder

1 Introduction

Various microprocessors, signal processing applications and communication techniques, etc., requires widely use of multiplier circuits. In bigger bit sizes, multiplier requires more numbers of adders to perform addition of partial products generated. Multiplier circuits basically restrict the speed and performance of the electronic

V. Shukla (✉) · O. P. Singh · G. R. Mishra
Amity University Uttar Pradesh, Lucknow Campus, Lucknow, India
e-mail: vandanashuklaec05@gmail.com

O. P. Singh
e-mail: opsingh@amity.edu

G. R. Mishra
e-mail: gr_mishra@rediffmail.com

R. K. Tiwari
Dr. Ram Manohar Lohia Avadh University, Faizabad, India
e-mail: rktiwari2323@yahoo.co.in

systems. Further, Vedic mathematics provide the solution to this issue [1–3]. Vedic mathematics is given by Sri Jagadguru Bharathi Krishna Tirthaji Maharaja (1884–1960) depending upon his rigorous study on Vedas [4, 5]. Vedic mathematics consists of a total of sixteen sutras which can be very effectively utilized in different parts of engineering and science.

Moreover, Reversible design approach is widely applicable in digital signal processing, optical computing calculations, low power CMOS design, etc., due to its ability to generate low loss efficient digital design [6–8]. In this paper, we have combined these two approaches to design efficient, high speed and low loss design of 4×4 size multiplier circuit with reduced parameters. This paper is illustrated in six sections. Initially, introduction of the work is presented, which is followed by basic concepts of Vedic mathematics and structure of 4×4 size Vedic multiplier circuit using conventional approach in Sects. 1 and 2 respectively. Further, fundamental issues and characteristics of Reversible logic approach are discussed in Sect. 3, whereas Sect. 4 presents the proposed approach for 4×4 size Vedic multiplier circuit using Reversible logic gates. After that, obtained results are discussed and analyzed as compared to the earlier designs in Sect. 5 followed by conclusion of the work in Sect. 6 at the end.

2 Basic Concepts of Vedic Multiplier

Vedic mathematics provides very fast and error-free calculations for any number of digits as compared to other methods. Vedic mathematics is derived from forgotten formulae of Atharva Veda by his holiness Jagadguru Shankaracharya Bharati Krishna Teerthaji Maharaja after his rigorous study and research of Vedas [9, 10].

Vedic mathematics contains a total of 16 sutras for various calculations in the field of science, engineering and other high-order calculating applications. These sutras provide very fast as well as flexible calculations for mathematical problems. All sixteen sutras are briefly described in Fig. 1.

Urdhva Tiryakbhyam is most widely applicable sutra for providing fast multiplication result of two numbers of any bit size. This method is also described as vertically and crosswise method. In this method bits are multiplied starting with vertical direction and then moving towards crosswise multiplication and calculations. Figure 2 clearly shows all seven steps involved in 4×4 size Vedic multiplication process using this method.

As shown in Fig. 2, two four bit binary numbers A ($A_3A_2A_1A_0$) and B ($B_3B_2B_1B_0$) are multiplied to provide the required multiplication result. Let multiplication result is represented as M ($M_7M_6M_5M_4M_3M_2M_1M_0$), these bits are calculated as follows:

1. Anurupye/Shunyamanyat (If one is in ratio, the other is zero)	2. Chalana Kalanabyham (Differences and similarities)
3. Ekadhikina Purvena (By one more than the previous One)	4. Ekanyunena Purvena (By one less than the previous one)
5. Gunakasamuchyah (Factors of the sum is equal to the sum of factors)	6. Gunitasamuchyah (The product of sum is equal to sum of the product)
7. Nikhilam Navatashcaramam Dashatah (All from 9 and last from 10)	8. Paraavartya Yojayet (Transpose and adjust)
9. Puranapuranabyham (By the completion noncompletion)	10. Sankalana- vyavakalanabhyam (By addition and by subtraction)
11. Shesanyankena Charamena (The remainders by the last digit)	12. Shunyam Saamyasamuccaye (When the sum is same then sum is zero)
13. Sopaantyadvayamantyam (The ultimate and twice the penultimate)	14. Urdhva-tiryakbhyam (Vertically and crosswise)
15. Vyashtisamanstih (Part and Whole)	16. Yaavadunam (Whatever the extent of its deficiency)

Fig. 1 Sutras of vedic mathematics

(i) $M_0 = A_0.B_0$

(ii) $M_1 = A_1.B_0 + A_0.B_1$

(iii) $M_2 = A_2.B_0 + A_1.B_1 + A_0.B_2 +$ carry forward M_1

(iv) $M_3 = A_3.B_0 + A_2.B_1 + A_1.B_2 + A_0.B_3 +$ carry forward M_2

(v) $M_4 = A_3.B_1 + A_2.B_2 + A_1.B_3 +$ carry forward M_3

(vi) $M_5 = A_3.B_2 + A_2.B_3 +$ carry forward M_4

(vii) $M_6 = A_3.B_3 +$ carry forward M_5

(viii) $C_7 =$ carry forward from M_6

Here, only final multiplication bits are calculated directly instead of first generating four partial products and then adding them. This is the main reason to make Vedic multiplier a fast calculating circuit.

Further, utilizing 2 × 2 size Vedic multiplier and adder circuits, we may design this 4 × 4 size multiplier circuit as shown in Fig. 3. Here, we require a total of four 2 × 2 size Vedic multiplier circuits in combination with three 4-bit Adder circuits.

Utilizing this concept for designing of 4 × 4 bit Vedic multiplier circuit, we further proceed for the concept of low power digital circuit design approach, named as Reversible logic in next section. Further, we will combine fast multiplication technique of Vedic multiplier with low power designing to provide our proposed 4 × 4 Vedic multiplier design with optimized performance parameter.
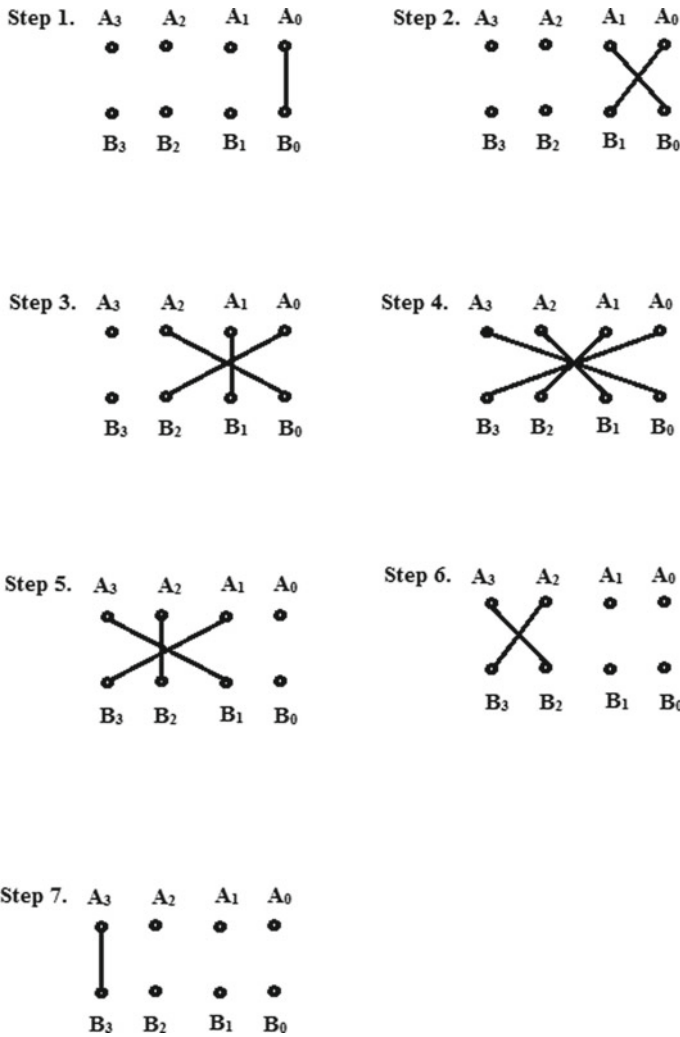
Fig. 2 Steps of 4-bit vedic multiplication

3 Reversible Logic Approach

The conventional use of digital circuit designing contains the source of heat loss from the systems because of the bit loss after the digital operation. Researcher R. Landauer has calculated the amount of heat loss/power generation per bit loss in the year 1961 [11]. This concept was that a total of k.T. ln 2 joules of energy is emitted for every lost bit. Here k is Boltzmann's constant ($= 1.38 \times 10^{-23}$ m^2 kg s^{-2} k^{-1}) and T is temperature of the digital device measured in kelvin.
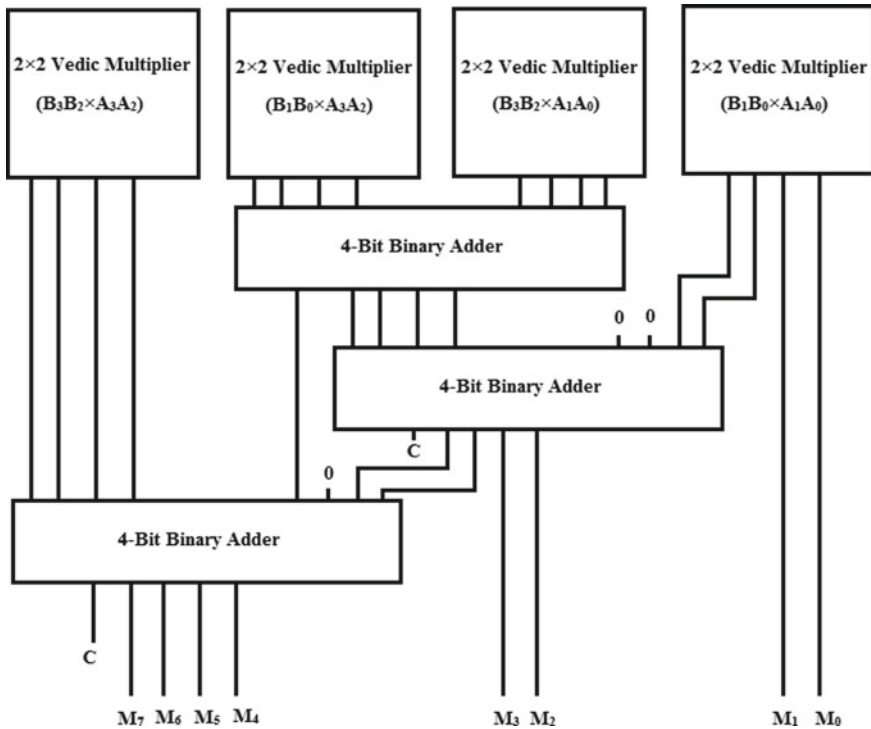
Fig. 3 4 × 4 size vedic multiplier circuit

Moreover, G. E. Moore has predicted the nature of increase in component density for selected chip area in the year 1965, famously known as "Moore's law" [12]. As shown in Fig. 4, this law states that component density basically doubles in the time duration of 2 years approximately.

Landauer's findings and Moore's law in combination indicates that the measure of heat generated from specific device area increases at exponential rate which in turn, deteriorates the efficiency and overall performance of the digital circuit. Further, in the year 1973, C. H. Bennett has given the concept of Reversible designing [13]. It was stated by him that, Reversible design entities will lead to the creation of digital systems with zero loss in ideal conditions. Utilizing this concept, various designers have proposed and optimized different digital circuits.

Some basic concepts of reversible logic approach are described as follows.

3.1 Reversible Logic Gates

These are digital logic gates of n input signals and n output signals, i.e., number of input and output are equal [14–18]. These gates have basically following properties:
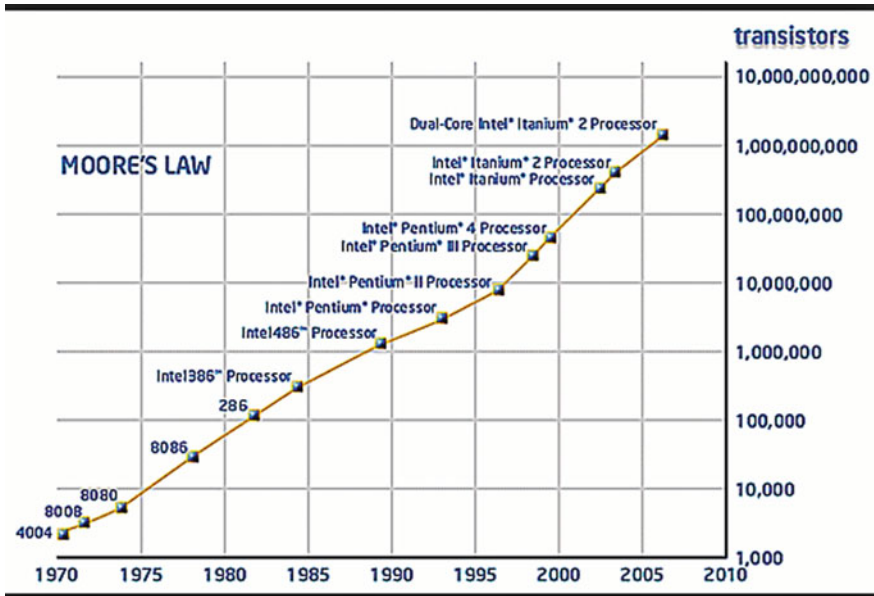
Fig. 4 Moore's law [12]
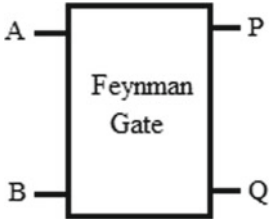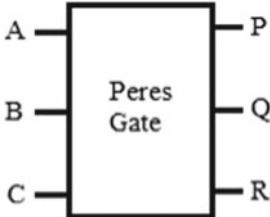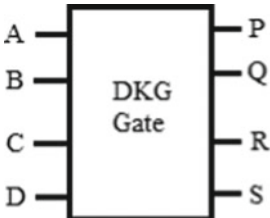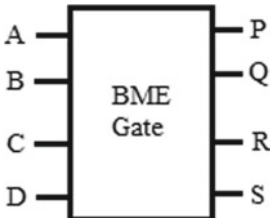
(i)   Same number of input and output lines.
(ii)   Input and output line have one to one mapping.
(iii) Each output line is at logic high for half the number of total input combinations possible.
(iv)   Applied input combination can be correctly retrieved by the knowledge of generated outputs at any instance of time.

Based on these characteristics, till now various Reversible logic gates of different sizes has been already proposed and utilized to design different digital circuits [19–21]. Table 1 represents some of these reversible logic gates with corresponding output equations and quantum cost.

3.2   Performance Parameters for Reversible Designs

During the designing of digital circuits using Reversible approach, following parameters play vital roles.

Table 1 Some examples of reversible logic gates

S. No.	Block diagram	Equations	Quantum cost
1	Feynman Gate (A → P, B → Q)	$P = A$; $Q = A \oplus B$;	1
2	Peres Gate (A → P, B → Q, C → R)	$P = A$; $Q = A \oplus B$; $R = A.B \oplus C$;	4
3	DKG Gate (A → P, B → Q, C → R, D → S)	$P = B$; $Q = A'.C + A.D'$; $R = (A \oplus B).(C \oplus D) \oplus C.D$; $S = B \oplus C \oplus D$;	6
4	BME Gate (A → P, B → Q, C → R, D → S)	$P = A$; $Q = A.B \oplus C$; $R = A.D \oplus C$; $S = A'.B \oplus C \oplus D$;	6

3.2.1 Total Reversible Gates

This parameter basically indicates the number of total Reversible logic gates utilized in the aimed digital circuit. As the number of Reversible gates increases, it will make the performance of digital system degrade. So, this number should be minimized as much as possible.

3.2.2 Garbage Outputs

These signals are generated from the realized circuit using Reversible approach in addition to the required ones. Garbage signals in turn leads to the bit loss from the circuit, which is in contrast to the concept of Reversible approach. So these outputs should also be minimized to follow the concept of Reversible approach.

3.2.3 Constant Inputs

Sometimes, we require "Logic Low" or "Logic High" signals to be applied to the Reversible logic gates connected in the aimed circuit for acquiring the correct working of designed Reversible circuit. These input signals are additional overheads on the circuit which tend to deteriorate the efficiency of the Reversible designs. Thus, it is required to minimize constant input signals applied to Reversible designs.

3.2.4 Quantum Cost

This parameter is described as basic Quantum/Reversible gates required to design the Reversible realization of any digital circuit. Here, cost of 1-input and 2-output Quantum/Reversible logic gate are zero and one respectively. Higher quantum cost of designed circuits deteriorate the overall performance of the system. Thus, this parameter is tried to be minimized as low as possible.

These performance parameters basically lay down the base to plan the strategy for the designing of any reversible circuit with better performance and low power loss.

3.3 Reversible Circuit Design Approach

In this design approach Reversible logic gates are utilized as design identities for designing of low power digital circuits instead of irreversible logic gates, as in the case of conventional digital design approach [22–25]. Here following ideal characteristics are aimed for designed Reversible circuit:

1. Ideally single Reversible logic gate should be used (minimum number of reversible gates in practical situations).
2. No garbage output should be generated from the designed Reversible design (otherwise garbage outputs should be minimized in practical scenario).
3. No constant input signal should be applied ideally (Number of constant input signals should be minimized practically).
4. Low quantum cost of the designed Reversible circuit should be aimed.
5. There should be no feedback connection in the designed Reversible circuit.

These aimed conditions for performance parameters related to Reversible designs provide a right direction for researchers to proceed. Till now, numerous designs have been proposed and optimized utilizing different combinations of available and new Reversible design entities [26–28]. Here, we utilize available Reversible logic gates to design optimized design for 4 × 4 size Vedic multiplier circuit having focus on above-mentioned characteristics. This design approach is described in detail in the subsequent section.

4 Proposed Design for 4 × 4 Vedic Multiplier Using Reversible Logic Approach

Here, we aim to design 4 × 4 size Vedic multiplier circuit using the concept of design shown in Fig. 3 earlier. In this figure, four 2 × 2 size Vedic multiplier blocks are connected to three 4-bit Adder block to provide the multiplication result M $(M_7M_6M_5M_4M_3M_2M_1M_0)$ of two 4-bit binary numbers A $(A_3A_2A_1A_0)$ and B $(B_3B_2B_1B_0)$. In the following subsections, we discuss 2 × 2 size Vedic multiplier circuit and 4-bit Adder circuit using Reversible approach, which are further utilized to design the aimed 4 × 4 size Vedic multiplier using Reversible approach.

4.1 Reversible Realization of 2 × 2 Size Vedic Multiplier Circuit

The optimized Reversible design of 2 × 2 size Vedic multiplier is already presented by authors in 162(2) issue of Sylwan Journal [10]. As shown in Fig. 5, this design utilizes a combination of BME, Peres and Feynman gates.

Here, multiplicand and multiplier bits are applied to individual BME gates. Multiplication result bits are taken out from second output line of Feynman gate. This design requires three constant input signals and generates four garbage outputs with a total of 17 quantum cost.

4.2 Reversible Realization of 4-Bit Adder Circuit

Here, 4-bit Adder circuit is designed using four DKG gates in cascade connection as shown in Fig. 6. This approach is already presented in International Conference on Industrial and Information Systems in the year 2016 by authors [29].

As shown in Fig. 6 constant input signals are applied to this adder block which generates a total of 8 garbage output signals with a total of 24 quantum cost. These
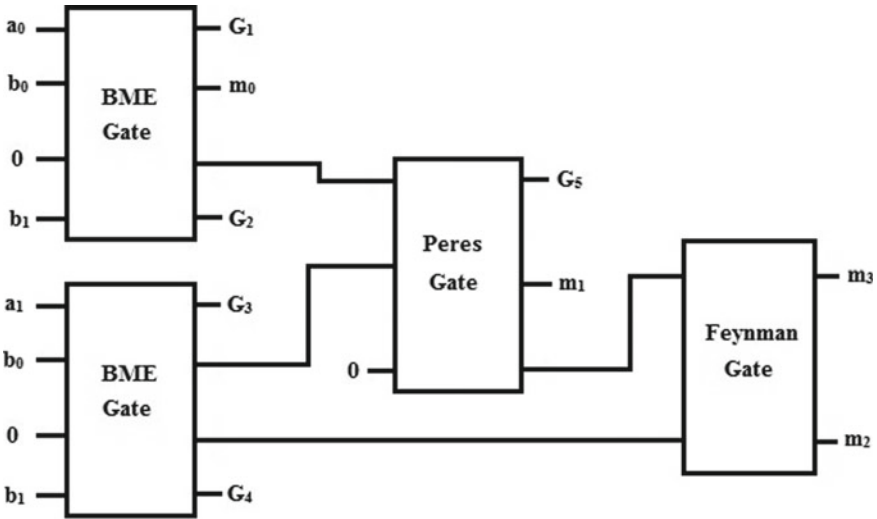
Fig. 5 Reversible realization of 2×2 size vedic multiplier circuit
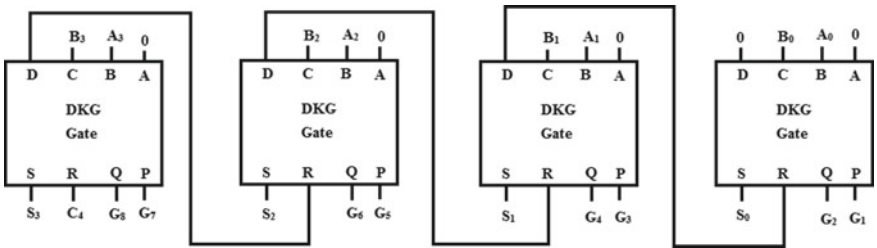
Fig. 6 Reversible realization of 4-bit adder circuit

designs are connected according to circuit presented in Fig. 3 to design our proposed 4×4 size Vedic multiplier circuit.

5   Result and Analysis

Our proposed design for 4×4 size Vedic multiplier circuit requires 30 constant input signals and generate 44 garbage output signals. Moreover, this design utilizes a combination of 28 Reversible gates with 140 quantum cost. Proposed design is compared with existing design as show in Table 2.

Table 2 Comparison table

Performance parameters	Total gates	Constant inputs	Garbage outputs	Quantum cost
Existing design 1 [2]	37	29	62	162
Existing design 2 [3]	32	32	60	166
Proposed design	28	30	44	140

Total number of Reversible gates utilized in the design (TG), number of constant inputs applied to the circuit (CI), total number of garbage output generated from the circuit (GO) and quantum cost of overall circuit (QC) are considered as the basis of comparison for proposed design over the existing designs. From the comparison table, it is clearly visible that proposed approach to design for 4 × 4 size Vedic multiplier circuit is providing most optimized performance parameters which in turn lead to low power high performance circuit.

Furthermore, this design in simulated and synthesized for ModelSim tool and Xilinx software using Device XC3S500E and behavioral approach on VHDL programming. Figures 7 and 8 are showing simulated waveforms and synthesized circuit of proposed Multiplier design.

Fig. 7 Simulated waveforms of proposed multiplier
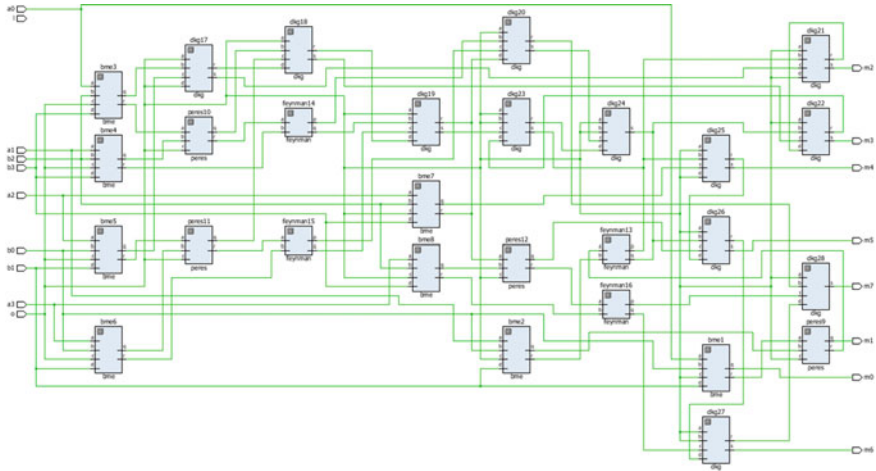
Fig. 8 Synthesized circuit of proposed multiplier

6 Conclusion

High-speed calculations and low power systems are two different dimensions of today's processing world. Vedic Mathematics overcomes the first requirement of high-speed processing up to much extent, whereas Reversible design approach targets on low power digital circuit designing. This paper combines both technologies to design a 4 × 4 size Vedic multiplier circuit using Reversible approach. Multiplier design presented here is analyzed and compared with the existing designs on the basis of TG, CI, GO and QC. The comparison shows that proposed approach is most optimized in terms of selected parameters. This design may be considered for further optimization of other high-speed low-power applications.

References

1. Anju, V.K., Agrawal, V.K.: FPGA implementation of low power and high speed vedic multiplier using vedic mathematics. IOSR J. VLSI Signal Process. **2**(5), 51–57 (2013)
2. Rakshith, T.R., Saligram, R.: Design of high speed low power multiplier using reversible logic: a vedic mathematical approach. In: International Conference on Circuits, Power and Computing Technologies (ICCPCT), 2013, IEEExplore, pp. 775–781 (2013)
3. Shivarathnamma, G.J., Kurian, M.Z.: A novel approach of 4*4 vedic multiplier using reversible logic gates. In: 1st International Conference on Innovations in Computing & Networking (ICICN16), CSE, RRCE, International Journal of Advanced Networking & Applications (IJANA), pp. 538–542 (2016)
4. Paramasivam, M.E., Sabeenian, R.S.: An efficient bit reduction binary multiplication algorithm using vedic methods. In: IEEE 2nd International Advance Computing Conference, 2010, ISBN: 978-1-4244-4791-6/10, pp. 25–28 (2010)

5. Kanhe, A., Das, S.K., Singh, A.K.: Design and implementation of low power multiplier using vedic multiplication technique. Int. J. Comput. Sci. Commun. **3**(1), 131–132 (2012)
6. Thapliyal, H., Srinivas, M.B.: Novel design and reversible logic synthesis of multiplexer based full adder and multipliers. IEEE **2**, 1593–1596 (2005)
7. Rice, J.E.: A new look at reversible memory elements. In: Proceedings International Symposium on Circuits and Systems (ISCAS) 2006, Kos, Greece, 21–24 May 2006, pp. 243–246 (2006)
8. Noor Mahammad, S.K., Sastry Hari, S.K., Shroff, S., Kamakoti, V.: Constructing online testable circuits using reversible logic. In: Proceedings of the 10th IEEE VLSI Design and Test Symposium (VDAT), Goa, India, August 2006, pp. 373–383 (2006)
9. Murugesan, G., Lavanya, S.: Design and implementation of high speed multiplier using vedic mathematics. ARPN J. Eng. Appl. Sci. **10**(16), 6758–6764 (2015)
10. Shukla, Vandana, Singh, O.P., Mishra, G.R., Tiwari, R.K.: Design and analysis of reversible realization of vedic multiplier circuit. Sylwan **162**(6), 32–43 (2018)
11. Landauer, R.: Irreversibility and heat generation in the computational process. IBM J. Res. Dev. **5**, 183–191 (1961)
12. Moore, G.E.: Cramming more compounds onto integrated circuits. Electronics **38**(8) (1965)
13. Bennett, C.H.: Logical reversibililty of computation. IBM J. Res. Dev., pp. 525–532 (1973)
14. Toffoli, T.: Reversible Computing. Tech memo MIT/LCS/ TM-151, MIT Lab for Computer Science (1980)
15. Fredkin, E., Toffoli, T.: Conservative logic. Int. J. Theor. Phys. **21**, 219–253 (1982)
16. Peres, A.: Reversible logic and quantum computers. Phys. Rev. A **32**, 3266–3276 (1985)
17. Feynman, R.: Quantum mechanical computers. Opt. News **11**, 11–20 (1985)
18. Sengupta, D., Sultana, M., Chaudhuri, A.: Realization of a novel reversible SCG gate and its application for designing parallel adder/subtractor and match logic. Int. J. Comput. Appl. **31**(9), 30–35 (2011). (0975-8887)
19. Thapliyal, H., Vinod, A.P.: Design of reversible sequential elements with feasibility of transistor implementation. In: Proceedings of the 2007 IEEE International Symposium on Circuits and Systems, pp. 625–628, New Orleans, USA (2007)
20. Singla, P., Malik, N.K.: A cost-effective design of reversible programmable logic array. Int. J. Comput. Appl. **41**(15) (2012)
21. Shukla, Vandana, Singh, O.P., Mishra, G.R., Tiwari, R.K.: Novel design of a multiplexer circuit using reversible logic. IJCER **3**(10), 30–35 (2013)
22. Shukla, H.P., Rao, A.G., Mall, P.: Design of low power comparator circuit based on reversible logic technology. In: 1st International Conference on Emerging Trends and Applications in Computer Science (ICETACS), IEEE, pp. 6–11 (2013)
23. Shukla, V., Singh, O.P., Mishra, G.R.: An efficient approach for the reversible realization of 2:4 decoder circuit. J. Int. Acad. Phys. Sci. **21**(4) (2017)
24. Frank, M.P.: Introduction to reversible computing: motivation, progress and challenges. In: Proceedings of the 2nd Conference on Computing Frontiers, pp. 385–390 (2005)
25. Wang, J.C., Pang, Y., Xia, Y.: A BCD priority encoder designed by reversible logic. In: IEEE, pp. 318–321 (2012)
26. Aradhya, R., Chinmaye, R., Muralidhara, K.: Design, optimization and synthesis of efficient reversible logic binary decoder. Int. J. Comput. Appl. **46**, 45–51 (2012)
27. Picton, P.: Multi-valued sequential logic design using Fredkin gates. Multiple-Valued Logic Journal **1**, 241–251 (1996)
28. Nagamani, A.N., Jayashree, H.V., Bhagyalakshmi, H.R.: Novel low power comparator design using reversible logic gates. Indian J. Comput. Sci. Eng. (IJCSE) **2**(4), 566–574 (2011)
29. Shukla, V., Singh, O.P., Mishra, G.R., Tiwari, R.K.: Performance parameters optimization and implementation of adder/subtractor circuit using reversible logic approach. In: 2016 11th International Conference on Industrial and Information Systems (ICIIS), IEEExplore, pp. 323–328 (2017)

Design and Implementation of 32-bit MIPS-Based RISC Processor

Sumit Patra, Sunil Kumar, Swati Verma and Arvind Kumar

Abstract With the continuously technology scaling, there exists a huge scope of undesirable hazards in processors. To solve the hazards, additional circuits are required in addition to conventional design and due to these additional circuits, the parameters area, power, and timing have been affected. Therefore, to make a processor having more number of operations without much affecting these parameters is a quite challenging task. In this paper, design and verification of 32-bit RISC CPU using 90 nm SCL CMOS technology is presented in detail. MIPS-based RISC architecture having operations like addition, subtraction, etc. Also having pipeline stages of five named as IF (Instruction Fetch), ID (Instruction Decode), EXE (Execute), MEM (Memory Access), WB (Write Back) to increase the throughput of the processor without degrading its latency. In this paper, all existing instructions as well as the new instructions, multiplication, and division are functionally verified. The analysis of performance parameters like area and power dissipation is done using synopsys design compiler with typical libraries of TSMC 90 nm technology.

Keywords RISC · MIPS · Low power · Pipeline architecture

S. Patra · S. Kumar (✉) · S. Verma · A. Kumar
Motilal Nehru National Institute of Technology Allahabad, Allahabad, India
e-mail: sunmnnit@gmail.com

S. Patra
e-mail: sumit.patra0010@gmail.com

S. Verma
e-mail: vermaswati.93@gmail.com

A. Kumar
e-mail: arvindk@mnnit.ac.in

1 Introduction

There are numbers of processors available in the market but designing of the processors using Hardware descriptive language like Verilog-HDL or VHDL are few in market. Using of HDL languages in the design help designer to use them in any embedded application. Design of the processor based on two ISA (Instruction Set Architecture): RISC and CISC [1, 2].

In a simple and short, Instruction Set Architecture defines working of the processor on an instruction. A large number of instructions is found in CISC processors and they perform various complex tasks. These instructions are typically of different length and different format, which requires very complex control unit and it also, requires larger area on chip. In many cases, they require different clock cycles to execute. Due to these reasons, Pipelining is not easy in CISC processors [3]. On the other hand, RISC ISA is having less number of instructions and has generally large number of general-purpose registers. These instructions are of fixed length and are designed in such a way that they can perform a single execution in a single cycle, which make it useful in pipelining. Arithmetic as well as Logical operations always involve registers and do not involve memory directly; to access memory LOAD and STORE instructions are used, so it is also called LOAD-STORE architecture. The above features of RISC design make it powerful in "System-on-Chip" market [4].

In a current time it is mandatory to use a processor or a machine that is efficient in terms of the three most important parameters, i.e., area, speed and power. Although there are trade-off between these three performance parameters, research work is going on to satisfy the above performance parameters [5, 6]. We have chosen MIPS as it requires low power and lesser area; as accessing time of register is faster than accessing time of memory and most of the instructions access register rather than memory, the execution time of instructions is less. The architecture of all MIPS based processors remains same while the implementation may vary in single cycle, multi-cycle and pipelined processors [7, 8].

1.1 MIPS Framework

1.1.1 MIPS Instruction Set

The instruction types of MIPS can be divided into three types:

a. Register or R-type instruction format
b. Immediate or I-type instruction format
c. Jump or J-type instruction format.

a. **Register type**

The format for R-type instruction is as follows (Fig. 1).
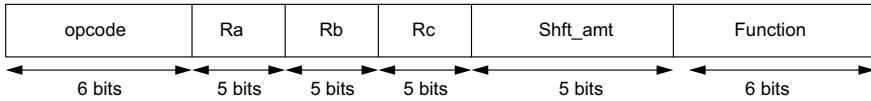
Fig. 1 Format of R-type instruction
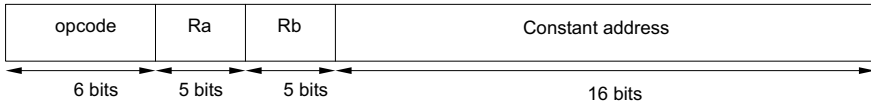
Fig. 2 Format of I-type instruction
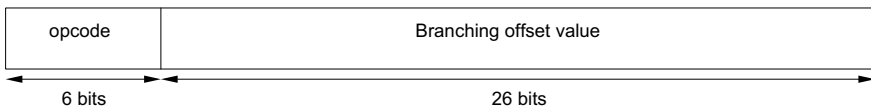
Fig. 3 Format of J-type instruction

Opcode field is machine code representation of basic operation of the instruction. Ra and Rb are two source registers (operands) respectively. Rc is destination register. It stores result of the operation. Shft_amt represents Shift amount, function field, often called the *function code*, and selects the specific variant of the operation in the op field. For an example: add $s1, $s2, $s3. Here, $s2 and $s3 represent source registers; $s1 represents destination register. The content of $s2 and $s3 are added together and result is stored in $s1 register.

b. **Immediate type**

I-type format is used for immediate and data transfer instructions. The format for R-type instruction is as follows (Fig. 2).

c. **Jump type**

The format for J-type instruction is as follows (Fig. 3).

The first 6 bits are for opcode and remaining 26 bits are for branch offset in form of 2's complement, which is shifted left by two and then added to PC to get branch target address.

1.1.2 MIPS Architecture

Instruction Fetch (IF)

It consists mainly of two blocks one is program counter (PC) and other is instruction memory. The instruction memory stores the data in their respective memory locations
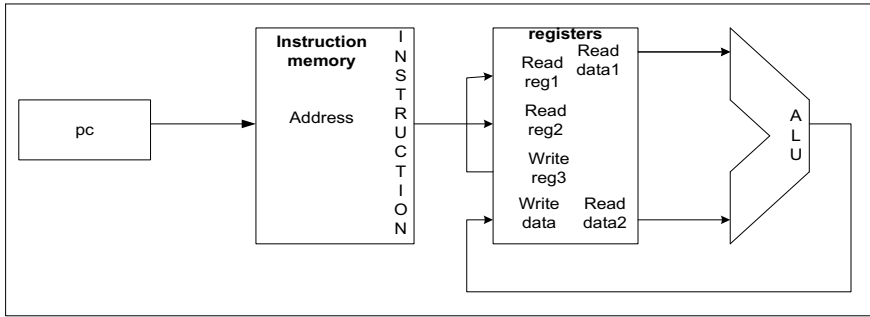
Fig. 4 Data path of Register type instructions [1]

and the program counter stores the address of the current instructions to be performed, which will direct the instruction memory. In the end, the value of next memory location (PC+4) and output of the instruction memory both are stored in the IF/ID register.

Data Path of R-Type Instruction

Here pc value gives the address to the instruction memory and based on that data two register's values are read; ALU does the operation on those values, depending on the instruction and the result is stored into another register as shown in Fig. 4 [9].

Data Path of I-Type Instruction

The data path for I-type instruction is shown in Fig. 5. Here the immediate data is last 16-bit data starting from least significant bit, which is sign, extended to 32 bit and is given to ALU as input.

Data Path for J-Type Instruction

Jump type instruction's data path is shown in Fig. 6. To get jump address we need to shift lower 26 bits by 2 to the left and then placing 00 as lower order bits and then concatenating upper 4 bits from high-order bits of (PC+4) which makes a 32 bit jump address.
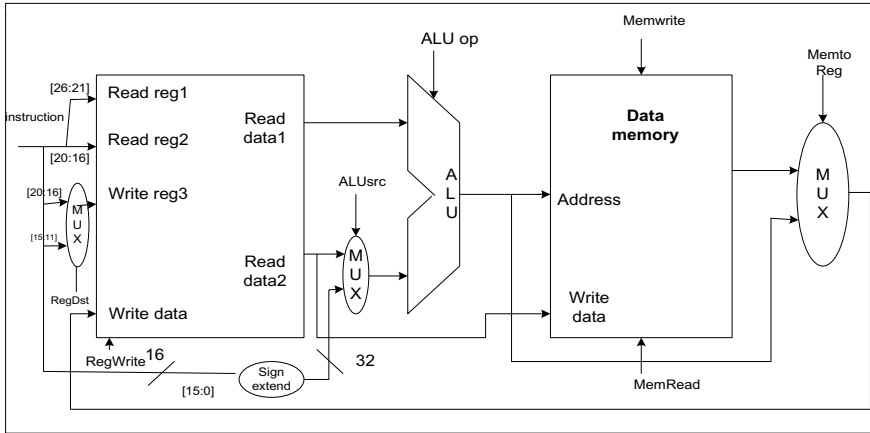
Fig. 5 Data path of Register type instructions [1]
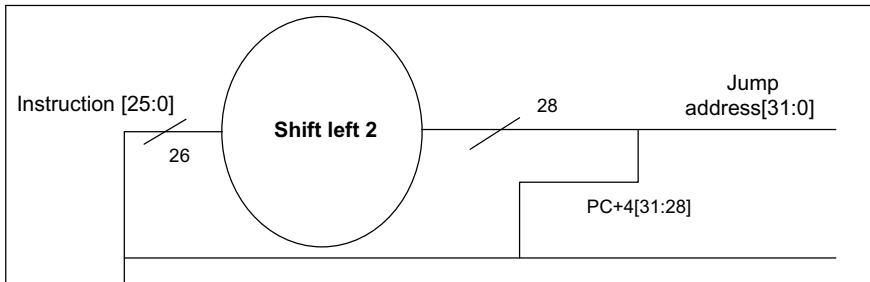
Fig. 6 Data path for J-type instruction [1]

MIPS Pipelined Architecture

MIPS is a Harvard type architecture in which the instruction memory and the data memory are different. In addition, a RISC based processor is having the reduced number of instructions, more number of general-purpose registers and mainly load-store architecture. The pipelining is added in the processor to increase the overall performance by executing the different instructions at the same time. It is possible for a multi-cycle CPU to process many tasks simultaneously. Suppose one task is processing without affecting it, we can start another new task. To make this possible, the overall processes are divided into a number of pipelined stages where each stage performs a specific task. The result of each stage is stored into a register before the next stage after every clock. This makes MIPS design to use every stages simultaneously, which in turn increases throughput of the processor.

Pipelining of MIPS is done using five stages [1]. The first one is Instruction Fetch or IF; second one is Instruction Decode or ID; third, one is Execution or EX; fourth, one is Memory access or MEM; and the last one is Write Back or WB. Every
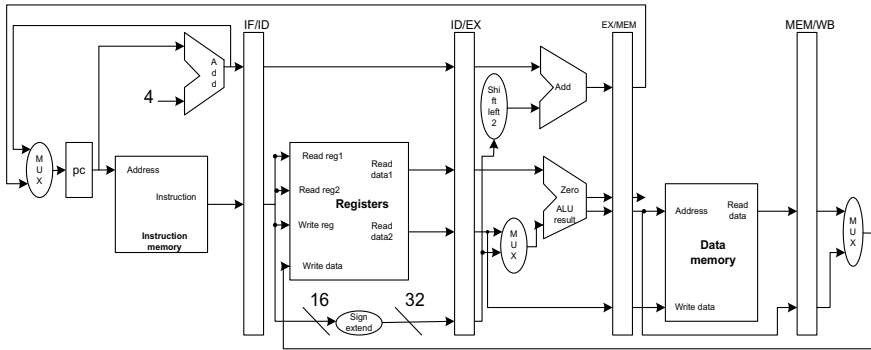
Fig. 7 MIPS pipelined data path [1]

two stages are separated by pipelined registers, which prevent data confliction due to execution of multiple instructions at a time. These registers are IF/ID, ID/EX, EX/MEM and MEM/WB (Fig. 7).

Improved Data Path of Pipelined Architecture

To improve the data path of pipelined architecture, two extra modules are added which are Data forwarding unit and Hazard detection unit.

a. **Hazard detection unit**

Hazards in pipeline can be categorized in three: Data hazard, structural hazard, and control hazard. Here we will concentrate only on data hazard. Data hazard in pipeline occurs when result of one instruction is needed in the next instruction, where both the instructions are still in the pipeline. In this case, pipeline has to be stalled.

For an example, if we have an AND instruction immediately followed by an OR instruction that uses the result $s1.

AND $s1, $t3, $t2
OR $t4, $s1, $t1

Due to data hazard, pipeline can be stalled for severe clock cycle. In the above example, result of the AND instruction cannot be written before fifth stage, which makes pipeline to waste three clock cycles.

b. **Data forwarding unit** [1]

From the above example, as soon as the ALU creates the AND result $s1 for AND instruction, we can supply it as an input for the next instruction. Adding extra hardware can supply the missing item early from the internal resources; this is called forwarding or bypassing. In Fig. 10, it shows that value of register $s1 is forwarded earlier to the next instruction by using data forwarding unit.

2 Proposed ALU Design

The ALU unit of 32-bit MIPS RISC processor is modified with two extra features, multiplication and division along with arithmetic instructions addition, subtraction and logical instructions AND, OR.

a. **Multiplication**

Multiplication is done by using Vedic multiplier, which is based on Vedic Mathematics [10]. Vedic mathematics is mainly based on sixteen word-formulae that are termed as Sutras. This is very fast multiplier. The block diagram for 32-bit Vedic multiplier is shown below in Fig. 8.

b. **Division**

The division is done by using Long division method. Process, which is repeated further are following steps:

Division: we start with first digit of dividend and then it is divided by divisor. The division result is quotient that may be zero.

Multiplication: Now to obtain product we multiply divisor with dividend.

c. **Subtraction**

To get remainder we subtract the product from the working dividend.

Bringing down the next digit: The last step is to bring the next digit down from current dividend to the remainder and combine it to get next working dividend.
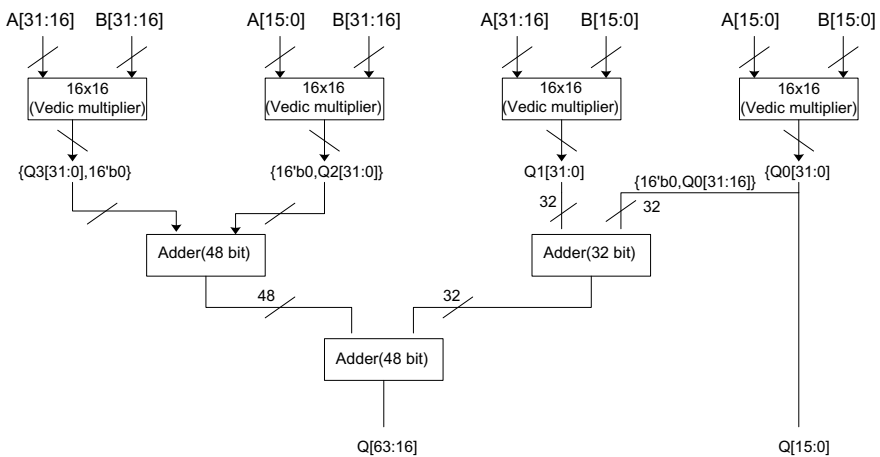
Fig. 8 Block diagram of 32-bit Vedic multiplier [10]

3   Result and Analysis

The functionality of these instructions is verified using cadence NCLaunch tool. The RTL schematic and test bench waveform are shown in Fig. 9 and Fig. 10 respectively. From the test bench we can see that ALU is having two inputs A and B, with ALUctl as control unit. Depending on the control signal, it performs different tasks and shows the result as ALU out, zero, and carryout.

The MIPS based RISC architecture is designed using Verilog-HDL and simulated using cadence NCLaunch tool. The designed architecture is synthesized using synopsys design compiler at 90 nm technology. The area and power report are shown
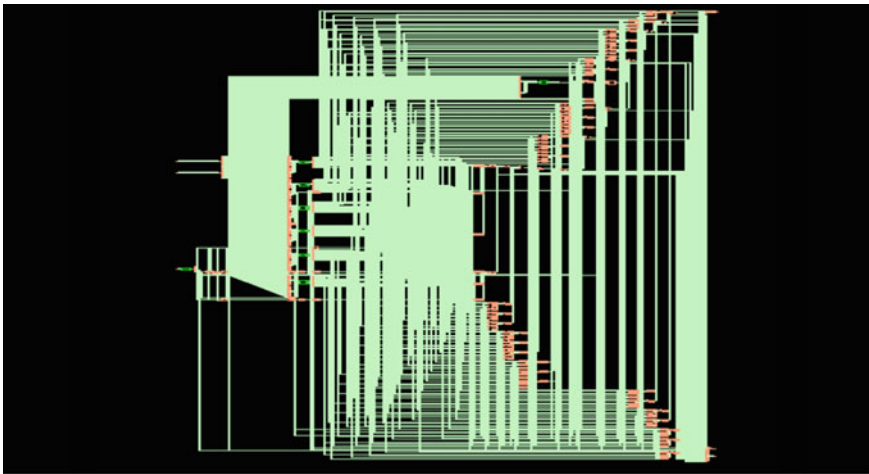
Fig. 9 RTL schematic of ALU
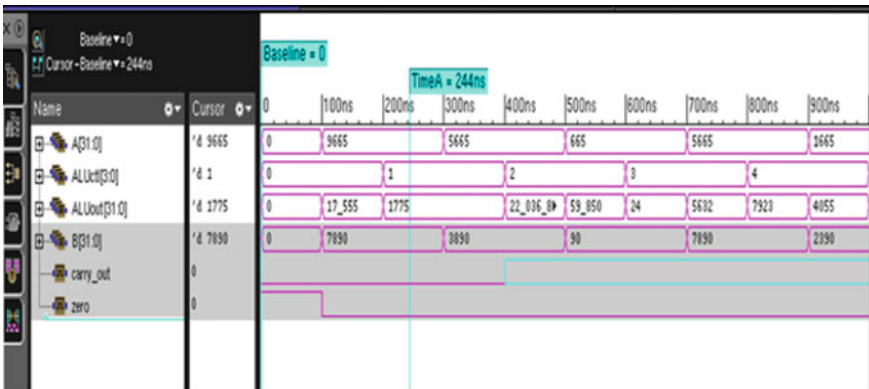
Fig. 10 ALU test bench waveform

```
report_area

*******************************************
Report : area
Design : Kogge_stone_1
Version: L-2016.03
Date   : Wed Mar  7 14:37:28 2018
*******************************************

Library(s) Used:

    tcbn90gbwp14ttc (File: /home/2015REL05/dc/ref/models/tcbn90gbwp14ttc.db)

Number of ports:                      1691
Number of nets:                      33107
Number of cells:                     31516
Number of combinational cells:       10416
Number of sequential cells:          21082
Number of macros/black boxes:            0
Number of buf/inv:                     901
Number of references:                   19

Combinational area:            68767.931748
Buf/Inv area:                   3803.183942
Noncombinational area:        314444.826790
Macro/Black Box area:              0.000000
Net Interconnect area:     undefined  (Wire load has zero net area)

Total cell area:              383212.758538
Total area:                undefined
```

Fig. 11 Area report of MIPS

below in Fig. 11 and Fig. 12 respectively. The total dynamic power consumption is 62.8953mW and cell leakage power is 907.9777 µW.

4 Conclusion

In this paper, 32-bit MIPS processor with five stages pipeline functionality is designed using Verilog-HDL and synthesized using synopsys design compiler tool using libraries of TSMC 90 nm technology. The logical instructions AND, OR, and the arithmetic instructions like multiplication (which is done by 32-bit Vedic multiplier), division (which is done by long division method) are tested through test bench. The problem of data dependencies in pipeline is solved by using hazard detection unit and data forwarding unit. Implementation of these two units was critical however it done successfully.

```
Global Operating Voltage = 1
Power-specific unit information :
    Voltage Units = 1V
    Capacitance Units = 1.000000pf
    Time Units = 1ns
    Dynamic Power Units = 1mW    (derived from V,C,T units)
    Leakage Power Units = 1nW


  Cell Internal Power  =  27.6519 mW   (44%)
  Net Switching Power  =  35.2433 mW   (56%)
                          ---------
Total Dynamic Power     =  62.8953 mW   (100%)

Cell Leakage Power      = 907.9777 uW

Information: report_power power group summary does not include estimated clock t
ree power. (PWR-789)
```

Power Group (%) Attrs	Internal Power	Switching Power	Leakage Power	Total Power
io_pad (0.00%)	0.0000	0.0000	0.0000	0.0000
memory (0.00%)	0.0000	0.0000	0.0000	0.0000
black_box (0.00%)	0.0000	0.0000	0.0000	0.0000
clock_network (0.00%)	0.0000	0.0000	0.0000	0.0000
register (0.00%)	0.0000	0.0000	0.0000	0.0000
sequential (43.27%)	26.8302	1.0584e-02	7.6580e+05	27.6066
combinational (56.73%)	0.8193	35.2327	1.4218e+05	36.1937
Total	27.6495 mW	35.2433 mW	9.0798e+05 nW	63.8003

Fig. 12 Power report of MIPS

References

1. David, A.P., John, L.H.: Computer Organization and Design: The Hardware/Software Interface, vol. 1, p. 998. San mateo, CA: Morgan Kaufmann Publishers (2005)
2. Palnitkar, S.: Verilog HDL: A Guide to Digital Design and Synthesis, vol. 1. Prentice Hall Professional (2003)
3. Topiwala, M.N., Saraswathi, N.: Implementation of a 32-bit MIPS based RISC processor using Cadence. In: 2014 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT). IEEE (2014)
4. Kadam, S.U., Mali, S.D.: Design of risc processor using VHDL. 2016 Int. J. Res. Granthaalaya **4**(6) (2016)

5. Oliver, J.P., et al.: Clock gating and clock enable for FPGA power reduction. In: 2012 VIII Southern Conference on Programmable Logic (SPL). IEEE (2012)
6. Bhosle, P., Moorthy, H.K.: FPGA implementation of low power pipelined 32-bit RISC Processor. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **1**(3) (2012)
7. Tina, G., Daruwala, R.D.: Performance improvement of MIPS architecture by adding new features. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(2) (2013)
8. Kadam, S.U., Mali, S.D.: Design of instruction fetch unit and ALU for pipelined RISC processor. Int. J. Eng. Comput. Sci. **5**(2) (2016)
9. Mansoor, C.M.M.: An approach for automatic design of application specific instruction set processors (ASIP) (2017)
10. Kumar, G.G., Charishma, V.: Design of high speed vedic multiplier using vedic mathematics techniques. Int. J. Sci. Res. Publ. **2**(3) (2012)

Signal Processing

Image Compression Using Hybrid Approach and Adaptive Scanning for Color Images

Ayush Kumar Nigam, Priyank Khare and Vinay Kumar Srivastava

Abstract This paper aims at the design of an algorithm for image compression. For color images RGB components are extracted and converted to YCbCr as a pre-processing step and then block based Hybrid Transform is performed. After this preprocessing step image is quantized and projected lossless scanning is done to ensure the excellent quality of the recovered image. In this paper adaptive scanning approach is anticipated to compress the length of codeword to be transmitted and thus achieving higher compression ratios. Principle behind image compression is to reduce the number of bits required per pixel.

Keywords Discrete cosine transform (DCT) · Adaptive scanning · Discrete wavelet transform (DWT)

1 Introduction

Compression is mainly a technique of reducing the total space required to store any data and then recover it back with good quality. In today's era with the growing use of Internet and other technologies there is a huge demand for data storage. So to meet this demand efficiently we have to compress this data to maximum extent possible, ensuring that it can be recovered back with good quality. Since in an image pixels are highly correlated so this characteristic of an image can be used for compression that is we reduce correlation among the pixels or try to transmit only difference between the pixels. There are basically two types of compression techniques:

A. K. Nigam (✉) · P. Khare · V. K. Srivastava
Department of Electronics & Communication Engineering, MNNIT, Allahabad, India
e-mail: ayushnigam2715@gmail.com

P. Khare
e-mail: rel1606@mnnit.ac.in

V. K. Srivastava
e-mail: vinay@mnnit.ac.in

- Lossless Image Compression: In this case quality of an image is absolutely conserved and compression is achieved by reducing the redundancy among the pixels.
- Lossy Image Compression: In this case quality of an image is not totally conserved in order to increase the amount of compression but distortion is kept within the desired limits to ensure good quality of the recovered image.

In an image compression, transforms such as Discrete Cosine Transform [1–3] or Discrete Wavelet Transform [4, 5] are basically used to concentrate the amount of energy associated with the pixels into a few coefficients and thus reduce the number of bits required to represent an image. In this paper adaptive scanning approach is used to achieve high compression ratios. To implement this adaptive scanning approach we are transmitting a matrix of binary codes for all the blocks. This matrix represents the type of scan used in each block. During the decompression phase this matrix is used to perform inverse scan of the blocks at the receiver side.

2 Proposed Method

The method elaborated below is devoted to image compression using Hybrid Transform and Adaptive Scanning. It is devoted to provide the excellent quality of recovered image by using lossless Adaptive Scanning. This method aims at preserving the maximum possible excellence of an image and thereby achieving high compression ratios. The block diagram shown in Fig. 1 has two phases, first is the compression phase and second is the reverse that is the decompression phase.
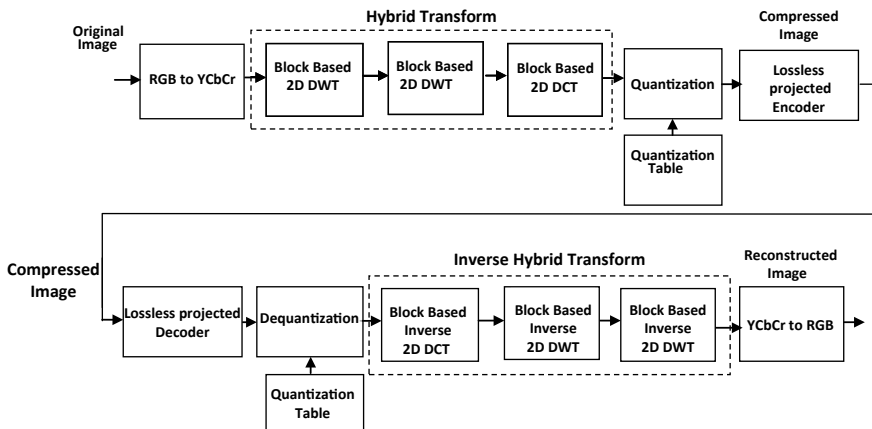
Fig. 1 Block diagram of image compression: **a** Compression stage. **b** Decompression stage

2.1 RGB to YCbCr Conversion

RGB to YCbCr conversion is done to achieve high compression ratios. As in YCbCr format most of the signal energy is present in the Y plane so maximum compression can be achieved by setting high compression ratios for Cb as well as Cr planes and thus conserving the quality of the image.

The RGB to YCbCr conversion is performed as follows:

$$
\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168736 & -0.331264 & 0.5 \\ 0.5 & -0.418688 & -0.081312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \tag{1}
$$

However, the reverse transformation is given by:

$$
\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.344136 & -0.714136 \\ 1 & 1.772 & 0 \end{bmatrix} = \begin{bmatrix} Y \\ Cb - 128 \\ Cr - 128 \end{bmatrix} \tag{2}
$$

2.2 Hybrid Transform

For every color image after RGB to YCbCr conversion is done, block-based hybrid transform is applied. The main purpose of this union transform is to utilize the benefits of both the transforms [6–12]. The advantage of DCT is to store most of the important information into very less number of coefficients and that of DWT is its multi resolution ability. One-dimensional DCT is defined as in [13]:

$$
C(k) = a(k) \sum_{d=0}^{N-1} f(d) \cos \left[\frac{(2d+1)k\pi}{2N} \right] \tag{3}
$$

where k = 0, 1, 2,……, N − 1
1 D Inverse DCT is defined as:

$$
f(d) = \sum_{d=0}^{N-1} a(k) c(k) \cos \left[\frac{(2d+1)k\pi}{2N} \right] \tag{4}
$$

where d = 0, 1, 2, …, N − 1

$$
a(k) = \sqrt{\left(\frac{1}{N} \right)} \text{ for } k = 0
$$

$$a(k) = \sqrt{\left(\frac{2}{N}\right)} \text{ for } k = 1, 2, 3, \ldots, N - 1.$$

In DWT, a picture corresponds to computation of wavelet functions, comprising dissimilar position and scale. Wavelet can be defined as a mathematical function by which an image is transformed as a computation of wavelet functions [14, 15]. DWT applies multi-resolution property for image breakdown, i.e., concurrent demonstration of picture on dissimilar resolution levels.

Two dimension wavelet transform requires a couple of waveforms for the decomposition of an image. This decomposition results in high frequency which is equivalent to the comprehensive component of the picture representing wavelet function and one more low frequency or smooth component of the picture representing scaling function. Figure 2 shows wavelet filter decomposition.

- LLn shows the approximation image at the nth stage of breakdown, ensuing from low-pass filtering in both vertical and horizontal directions.
- LHn shows the parallel details at the nth stage of disintegration as an outcome of raster low-pass filtering and vertical high-pass filtering.
- HLn shows the taken out vertical details or edges at the nth stage of disintegration resulting from vertical low-pass filtering and raster high-pass filtering.
- HHn corresponds to the diagonal details at the nth stage of disintegration resulting from high-pass filtering in both directions.

Coding method [16]:

(a) Compression Process

YCbCr image is first broken into block size of 16×16 pixels. After that 2D-DWT is applied on every block of size 16×16 which produces four details. From these four resultant sub-band details, approximation detail or sub-band of all the blocks are collected to form a matrix of all approximation details/sub-bands. This matrix is further broken into block size of 16×16 pixels and after that 2D-DWT is applied

Fig. 2 Wavelet filter decomposition

again on every block of size 16 × 16, which further produces four details. From these four resultant sub-band details, approximation detail or sub-band of all blocks are again collected to form a matrix of all approximation details/sub-bands. Then this approximation details/sub-bands matrix is further broken into block size of 8 × 8 and block based DCT is implemented. After that quantization and encoding gives the compressed image.

(b)   Decompression Process

At the recipient side, we decode the quantized DCT coefficients. Then block based dequantization is done. After that we compute two dimensional block based inverse DCT (IDCT) and further take inverse wavelet transform of the IDCT blocks. Since IDCT gives only LL components, i.e., approximation detail/sub-band of each block so we consider other three sub-bands (LH, HL, HH) as a matrix of zeros of size equal to approximation sub-band size while computing IDWT. As the level of breakdown at compression stage is two, two times inverse wavelet transform is performed following this procedure to obtain the original block size, i.e., 16 × 16. This process is repeated block wise for the whole matrix obtained. After the entire dequantized matrix received is decompressed by the procedure explained above, we get the blocks of size 16 × 16 which are rearranged to obtain the reconstructed image.

2.3   Quantization

After Block-Based DCT quantization is done. It is basically a process of converting real values of DCT matrix to discrete.

Quantization is done using the following equation:

$$C(k, d)_{quantized} = \text{round}\left(\frac{C(k, d)}{Q(k, d)}\right) \tag{5}$$

And reverse quantization process is represented as:

$$C(k, d)_{dequantized} = C(k, d)_{quantized} \times Q(k, d) \tag{6}$$

The quantization matrix for luminance (Q_l) and chrominance (Q_c) components is defined as follows:

$$Q_l = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \tag{7}$$

$$Q_c = \begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 18 & 21 & 26 & 66 & 99 & 99 & 99 & 99 \\ 24 & 26 & 56 & 99 & 99 & 99 & 99 & 99 \\ 47 & 66 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \end{bmatrix} \tag{8}$$

2.4 Lossless Projected Encoder

In encoding first adaptive scanning of blocks is done to ensure the best scan is adopted for each block and then run-length followed by Huffman coding is implemented to get the encoded sequence [17]. The block diagram of Lossless Projected Encoder is shown in Fig. 3.

Adaptive Scanning
The quantized coefficient blocks received are now scanned using adaptive scanning technique. In adaptive scanning technique we perform four scans (zigzag, horizontal, vertical and Hilbert) on each block and calculate the run-length of each block. This produces two matrixes first containing elements and second containing the number of runs of these elements. Then the length of matrix containing number of runs is calculated for each type of scan. Scan giving minimum length for matrix containing number of runs is chosen as the best scan. Likewise each quantized block is finally
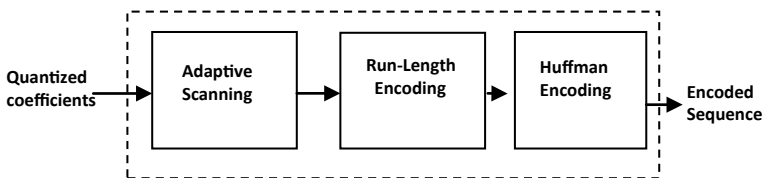
Fig. 3 Block diagram of lossless projected encoder

Fig. 4 The different ways for DCT blocks scanning

| Zigzag | Horizontal |
| Vertical | Hilbert |
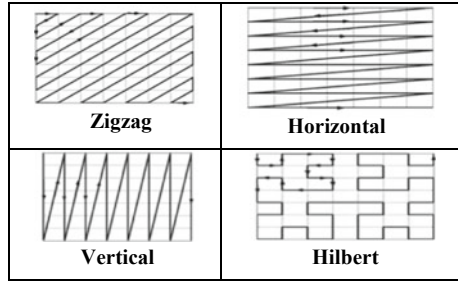
Table 1 Table of codes for various scans

Type of scan	Code
Zigzag	00
Horizontal	01
Vertical	10
Hilbert	11

scanned using the best scan for that block. Adaptive scanning proves to improve the compression performance as single scan can give overall good results for the entire image but might not give good results for each block of that image so adaptive scanning helps in using best scan for each block thus giving better performance than single scan for whole image (Fig. 4).

A matrix is maintained to store the information about the final scanning technique used in different quantized blocks to recover the blocks at the decompression level. This matrix is maintained using the codes given in Table 1. After adaptive scanning, Run-Length and Huffman coding techniques are implemented to compress the sequence more effectively. The entire compression and decompression process is represented in Figs. 5 and 6) correspondingly.

3 Experimental Results and Comparison

Some well known standard images like Airplane, Peppers, Lena, Baboon, etc. of size 256×256 are being used for testing the effectiveness of the proposed method (Fig. 7 and Table 2). To measure the effectiveness of the compressed image we are using some parameters like Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR).

$$\text{MSE} = \frac{1}{N \times T} \sum_{i=0}^{N-1} \sum_{j=0}^{T-1} \left(x_{ij} - \hat{x}_{ij} \right)^2 \tag{9}$$
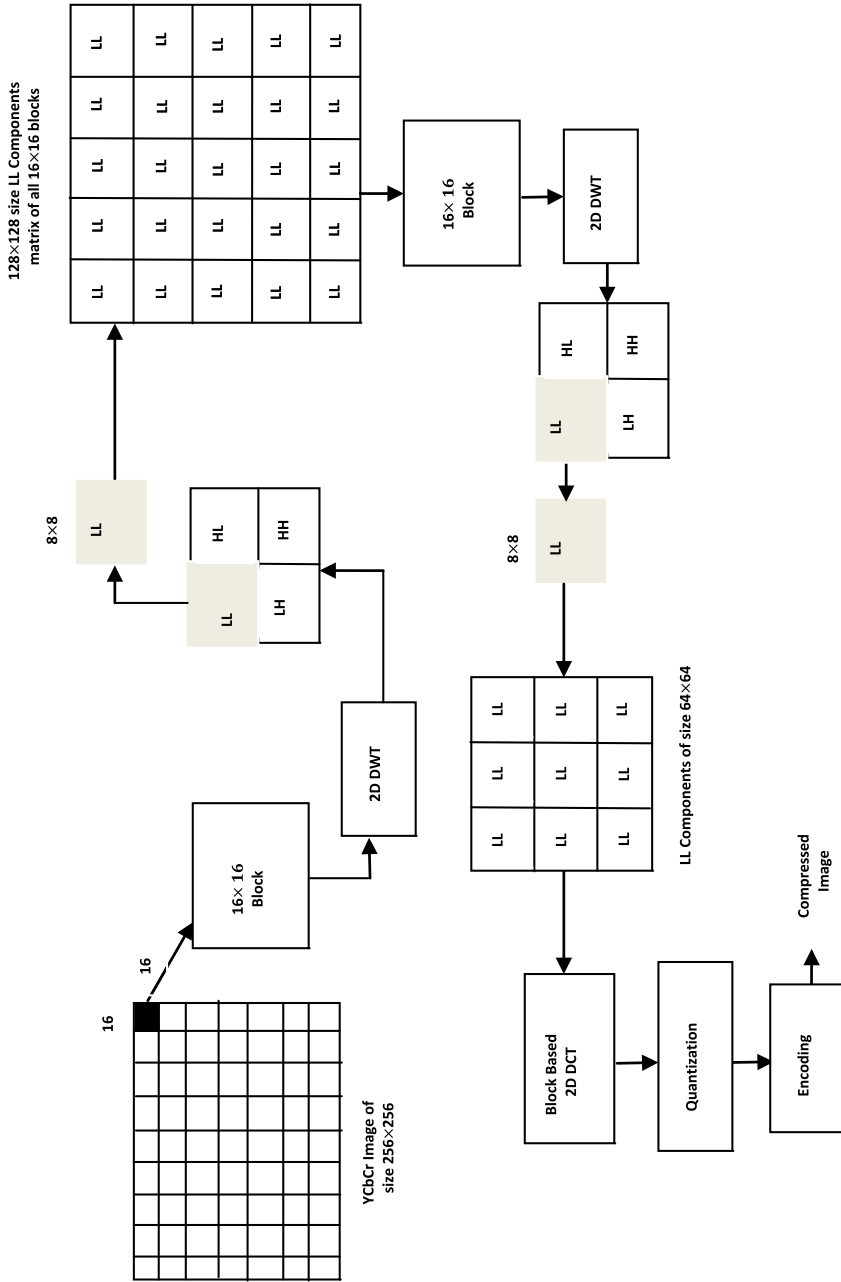
Fig. 5 Compression technique using hybrid transform
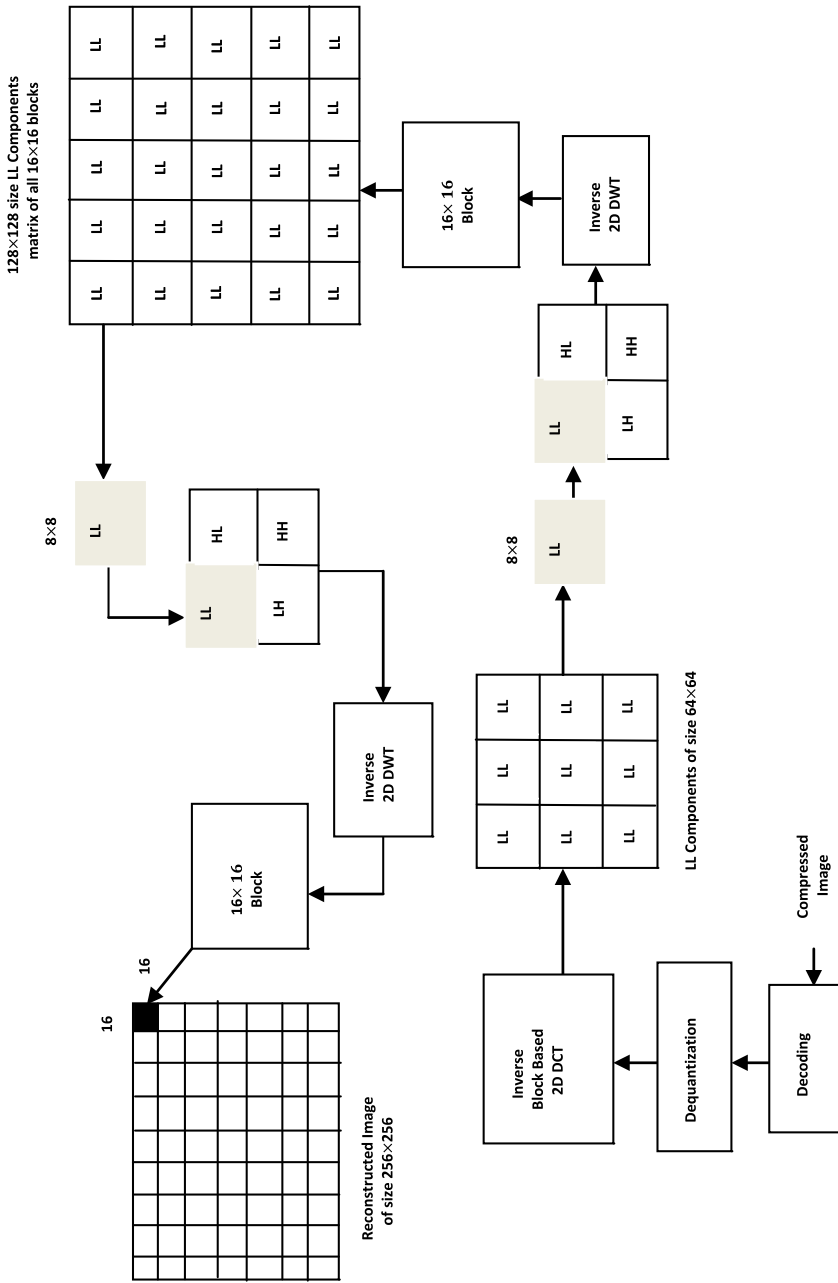
Fig. 6 Decompression technique using hybrid transform

(a)

(b)
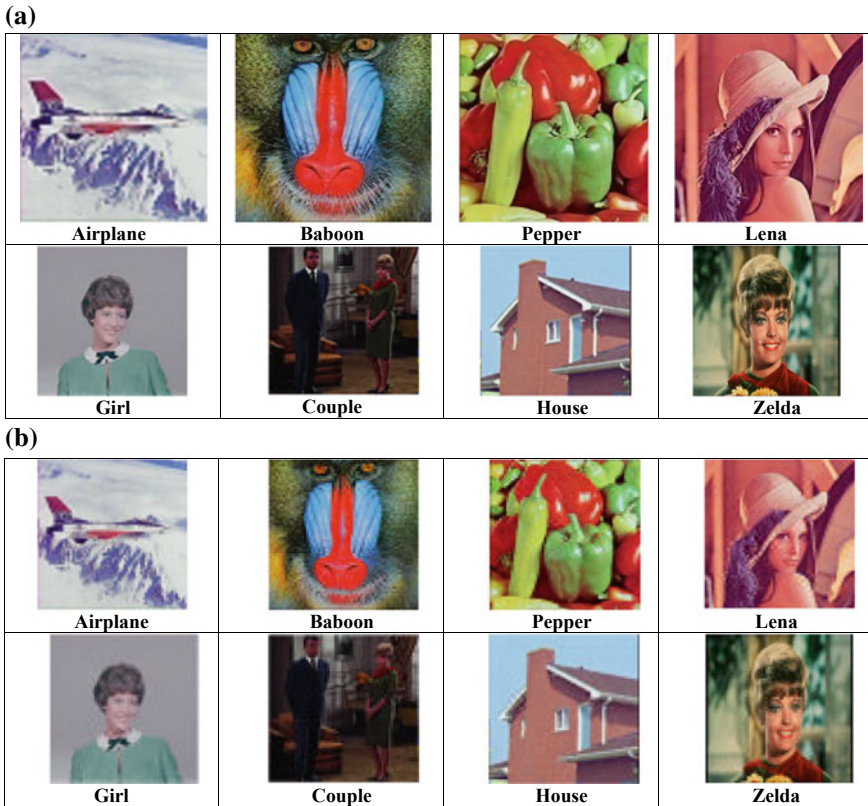
Fig. 7 Images: **a** Original test images, **b** Reconstructed images

Table 2 Table showing performance of the Proposed Method versus the standard DCT JPEG algorithm for color images in terms of PSNR and Bpp

Image	Proposed Method		JPEG	
	PSNR	Bpp	PSNR	Bpp
Airplane	32.1797223	0.64	35.636544	3.50
Baboon	29.7926640	0.80	31.9315114	5.25
Peppers	31.6047583	0.90	35.6831634	4.06
Lena	31.5561465	0.73	34.3919058	3.76
Girl	36.5146016	0.36	44.4478466	1.42
Couple	33.6564953	0.57	40.7658724	2.54
House	33.6571095	0.58	40.3144454	2.60
Zelda	32.2426209	0.72	37.9456113	3.28
Average	32.6505	0.6625	37.6396	3.30125

where x_{ij} and $\hat{x}_{ij}$ are correspondingly the intensities of original and reconstructed R, G and B planes. PSNR for gray level images is calculated by:

$$PSNR = 10 \times \log_{10} \frac{255^2}{MSE} \tag{10}$$

For color RGB images, PSNR is calculated by:

$$PSNR = 10 \times \log_{10} \frac{255^2 \times 3}{MSE(R) + MSE(G) + MSE(B)} \tag{11}$$

Size of the compressed image is calculated by obtaining the compression ratio (CR) or with bit per pixel (bpp) which are given by:

$$CR = \frac{\text{Size of original image in bits}}{\text{Size of compressed image in bits}} \tag{12}$$

$$bpp = \frac{24\text{bits}}{CR} \tag{13}$$

Since at the time of compression while performing two level 2D DWT we are only taking approximation detail or sub-band (LL) and discarding other three sub-bands so size of original image gets reduced which results in reduction of the computation time for the proposed scheme though we are using Hybrid Transform along with RGB to YCbCr conversion and adaptive scanning (Fig. 8 and Table 3).

(a)
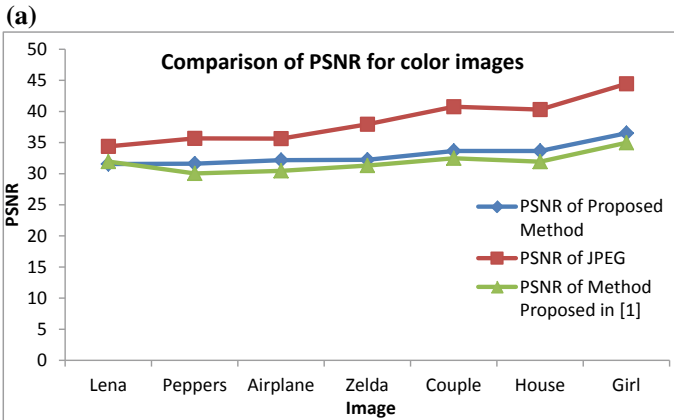
Fig. 8 Comparison graphs: **a** PSNR, **b** Bpp, **c** Encoding and Decoding time

(b)

(c)
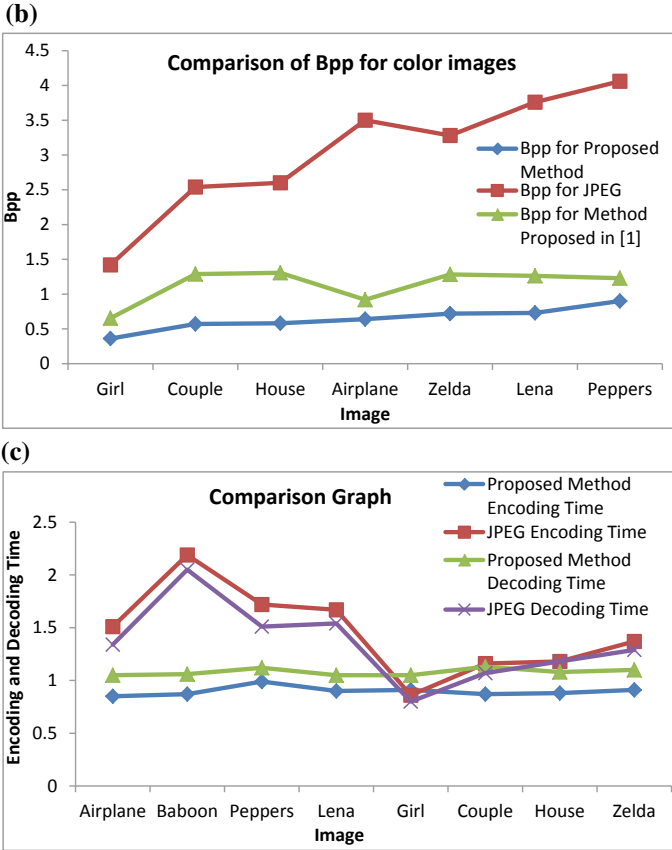
Fig. 8 (continued)

Table 3 Table showing performance of the Proposed Method versus the standard DCT JPEG algorithm for color images in terms of Encoding and Decoding time

Image	Proposed Method		JPEG	
	Encoding time	Decoding time	Encoding time	Decoding time
Airplane	0.85	1.05	1.51	1.34
Baboon	0.87	1.06	2.19	2.05
Peppers	0.99	1.12	1.72	1.51
Lena	0.90	1.05	1.67	1.54
Girl	0.91	1.05	0.86	0.80
Couple	0.87	1.13	1.16	1.07
House	0.88	1.08	1.18	1.18
Zelda	0.91	1.10	1.37	1.29
Average	0.8975	1.08	1.4575	1.3475

4 Conclusion and Future Scope

An efficient image compression technique to maximize the compression ratio along with achieving good values of PSNR is used in this paper. The graphs show that with slight reduction in PSNR value we can get a significant improvement in bit per pixel (Bpp) thus achieving high compression ratios. This work also proposed solution to blocking artifacts, false contouring and ringing effect by proposing a Hybrid (DWT and DCT based approach) Transform. Hybrid DWT-DCT algorithm is implemented to utilize the benefits of the two transforms (DWT as well as DCT). Simulation results elucidate that the proposed algorithm can significantly maximize compression ratio with good values of PSNR. In future, the projected technique can also be used for video coding. Also the algorithm can be realized in hardware accomplishment. As luminance and chrominance quantization matrix of JPEG algorithm are used in this scheme which would not be optimum for the proposed method so optimum quantization matrix for the proposed scheme can be found as a part of future work.

References

1. Douak, F., Benzid, R., Benoudjit, N.: Department of Electronics, Faculty of Engineering, University of Batna, Batna 05000, Algeria. Color image compression algorithm based on the DCT transform combined to an adaptive block scanning. Int. J. Electron. Commun. (AEU) **65**, 16–26 (2011)
2. Raid, A.M., Khedr, W.M., El-dosuky, M.A., Ahmed, W.: Mansoura University, Faculty of Computer Science and Information System Zagazig University, Faculty of Science. JPEG Image Compression Using Discrete Cosine Transform A Survey. Int. J. Comput. Sci. Eng. Surv. (IJCSES) **5**(2) (2014)
3. Messaoudi, A., Srairi, K.: Colour image compression algorithm based on the DCT transform using difference lookup table. Electron. Lett. **52**(20), 1685–1686 (2016)
4. Rabbani, M., Joshi, R.: Eastman Kodak Company, Rochester, NY 14650, USA. An overview of the JPEG2000 still image compression standard. Signal Process.: Image Commun. **17**, 3–48 (2002)
5. Skodras, A., Christopoulos, C., Ebrahimi, T.: The JPEG 2000 still image compression standard. IEEE Signal Process. Mag. **18**(5), 36–58 (2001)
6. Ebrahimi, F., Chamik, M., Winkler, S.: JPEG versus JPEG 2000: an objective comparison of image encoding quality. In: Applications of Digital Image Processing XXVII, vol. 5558. International Society for Optics and Photonics (2004)
7. Verma R.L., Ojha, D., Gupta, P., Gupta, M.: Excellence of DCT-DWT a hybrid approach for image compression. Int. J. Electr., Electron. Data Commun. **1**, ISSN (P): 2320-2084 (2013)
8. Deshlahra, A., Shirnewar, G.S., Sahoo, A.K.: A comparative study of DCT, DWT & hybrid (DCT-DWT) transform (2013)
9. Gautam, B.: Image compression using discrete cosine transform & discrete wavelet transform. Department of Computer Science and Engineering, National Institute of Technology, Rourkela (2010)
10. Saroya, N., Kaur, P.: Analysis of image compression algorithm using DCT and DWT transforms. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **4**(2) (2014)
11. Kurniawan, A., Purboyo, T.W., Prasasti, A.L.: Implementation of Image Compression Using Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). Int. J. Appl. Eng. Res. **12**(23), 13951–13958 (2017)

12. Rafael, C., Gonzalez, R.E., Woods, S.L.: Digital Image Processing, Pearson Edition (2005)
13. Cabeen, K., Gent, P.: Image compression and the discrete cosine transform. College of the Redwoods (1998)
14. DeVore, Ronald A., Jawerth, B., Lucier, Bradley J.: Image compression through wavelet transform coding. IEEE Trans. Inf. Theory **38**(2), 719–746 (1992)
15. Barbhuiya, A.J.I., Laskar, T.A., Hemachandran, K.: An approach for color image compression of JPEG and PNG images using DCT and DWT. In: 2014 International Conference on Computational Intelligence and Communication Networks (CICN), IEEE (2014)
16. Verma, R.L., et al.: Excellence of DCT-DWT: a hybrid approach for image compression. Int. J. Electr. Electron. Data Commun. **1**, 2084–2320 (2013)
17. Huffman, D.A.: A method for the construction of minimum-redundancy codes. Proc. IRE **40**(9), 1098–1101 (1952)

MR Image Denoising Using Adaptive Wavelet Soft Thresholding

Sima Sahu, Harsh Vikram Singh, Amit Kumar Singh and Basant Kumar

Abstract Medical imaging has played an important role in medical disease detection, diagnosis, and research related findings and therapy. Removal of noise in a medical image is mandatory step for quality assessment. This paper presents a feature preserved Magnetic Resonance (MR) denoising algorithm which is based on well-accepted multiresolution and statistical modeling. An adaptive wavelet soft thresholding method is designed to remove Gaussian noise from MR image. Orthogonality property of the wavelet transform is used to find the noise variance, using a Median Absolute Deviation (MAD) estimator. Distribution parameters of the wavelet coefficients are estimated by modeling the coefficients using a Normal Inverse Gaussian (NIG) probability density function (PDF). The threshold value is calculated considering the noise and signal information and the wavelet coefficients are updated accordingly. Estimation and analysis of the performance of the proposed method is performed using peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) performance evaluation parameters; these parameters demonstrate the noise suppression ability of the proposed method. Superiority of the proposed denoising method is further established through visual inspection of the denoised images obtained from proposed as well as other contemporary reported methods.

Keywords Multiresolution · Denoising · Soft thresholding · Statistical modeling

S. Sahu (✉)
Dr. A. P. J. Abdul, Kalam Technical University, Lucknow, Uttar Pradesh, India
e-mail: simahal@rediffmail.com

H. V. Singh
Department of Electronics, Kamla Nehru Institute of Technology (KNIT), Sultanpur,
UttarPradesh, India
e-mail: harshvikram@gmail.com

A. K. Singh
Department of Computer Science and Engineering, NIT, Patna, India
e-mail: amit.singh@nitp.ac.in

B. Kumar
Department of ECE, Motilal Nehru NIT, Allahabad, India
e-mail: singhbasant@mnnit.ac.in

1   Introduction

To study the functional characteristics and structural features, Magnetic Resonance imaging (MRI) is one of the most important techniques [1]. Induced thermal noise in the radio frequency coils due to the movement of charged particles produce noise in MRI. Preamplifiers generally have small abnormalities also cause noise in the MRI. Noise may obstruct the computer aided diagnosis and visual inspection. Thus removal of noise through different techniques is highly required.

Wavelet domain approaches in denoising applications have been used and most popular procedures. These approaches use wavelet decomposition procedure which divides the image into different scales and resolutions called subbands. Approximation and detailed subbands result from the wavelet decomposition. Further, hard and soft thresholding the coefficients are the next procedures generally applied to the detailed wavelet coefficients to obtain the true signal components. Neighshrink, Sureshrink, Visushrink and Bayesshrink are the wavelet thresholding approaches proposed in literatures [1]. The signal energy is concentrated to a small number of large wavelet coefficients, thus application of discrete wavelet transform (DWT) produce satisfactory noise reduction [2]. The soft thresholding function is defined as follows by Eq. (1) [3].

$$\widehat{Y} = \begin{cases} 0; & |Y| \leq T_h \\ sgn(Y)[|Y| - T_h]; & |Y| > T_h \end{cases} \tag{1}$$

where $\widehat{Y}$ denotes estimated value of wavelet coefficients, T_h is the threshold parameter and sgn represents the *sign* function. Threshold value is defined as [3]

$$T_h = \frac{noise\ variance}{standard\ deviation\ of\ noisefree\ signal} \tag{2}$$

The denoising efficiency of the wavelet threshold method is dependent on the threshold value T_h [4]. Accurate estimation of the threshold value results increased denoising efficiency. This paper implements the statistical modeling approach for finding the noise variance and standard deviation of the noise-free signal. Statistical modeling of wavelet coefficients have been proposed in literature for denoising different medical images [5, 6]. The distribution of wavelet coefficients has a sharp peak at the center and is heavy tailed in structure [7]. Different probability distribution functions (PDFs) such as Cauchy, Levy, NIG, Alpha-stable and Gaussian mixture model have been used by the researchers for modeling the wavelet coefficients [8]. A NIG PDF is used in this paper to model the wavelet coefficients because it has been verified as an excellent choice for modeling. The parameters of NIG PDF are estimated using maximum likelihood (ML) estimator [6].

The paper is organized as follows. Section 2 discusses about modeling of wavelet coefficients for estimating signal and noise parameters. Section 3 gives the proposed

methodology. The simulation result in terms of performance parameters and visual comparisons are discussed in Sect. 4. Finally Sect. 5 concludes the paper.

2 Background

This section discusses some preliminaries about the wavelet transform, its modeling with NIG distribution, estimation of NIG parameters, and estimation of signal and noise variances. Let LL_m, LH_i, HL_i, and HH_i for level i = 1, 2, 3...m (m is the maximum level of decomposition) are the subbands formed after decomposing the image in wavelet domain [9]. LL is the approximation subband and LH, HL and HH are detail subbands. The approximation subband coefficients are left unchanged as they carry signal information. The detail subbands coefficients are modeled to get the true signal information. The modeling of the wavelet coefficients is carried out by choosing a suitable distribution which should fit the empirical distribution of wavelet coefficients as shown in Fig. 1.
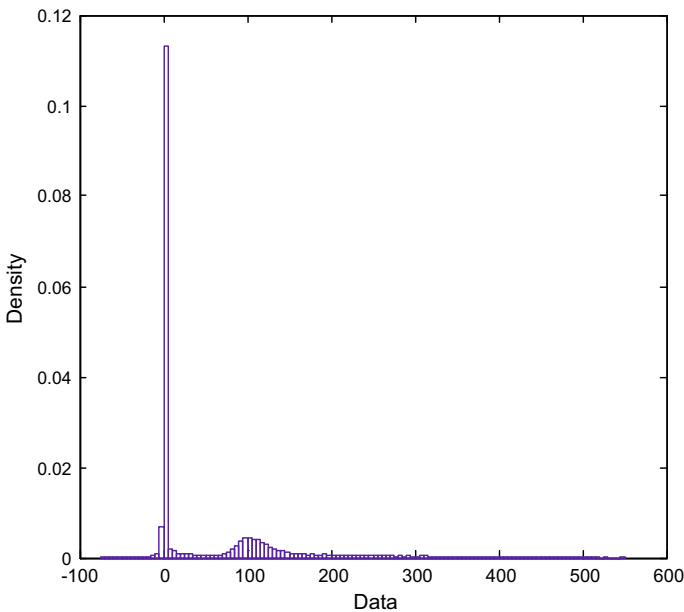
Fig. 1 Distribution of empirical wavelet coefficients

2.1 NIG Distribution and Goodness-of-Fit Test

NIG distribution was proposed by Barndorff-Nielsen in the year 1997 [10]. This distribution is the mixing distribution of inverse Gaussian and normal distribution. NIG distribution is expressed as

$$X(x) = \alpha \delta e^{\alpha \delta} k_l \left(\alpha \sqrt{\delta^2 + x^2} \right) / \pi \sqrt{\delta^2 + x^2} \tag{3}$$

k is the modified Bessel function with l(index) $= 1$, and δ, α are the distribution parameters. α defines tail-heaviness and it controls the steepness of the distribution. δ is the scale parameter [11]. This distribution effectively models the heavy-tailed data, and the goodness-of-fit curves are shown in Figs. 2, 3, 4 and 5 through PDF and probability plots.

Figure 2 shows the PDF plot for the empirical data for HL subband at level-1. The goodness-of-fit for NIG and Gaussian distribution are also shown in this figure. It is seen that the empirical data is modeled effectively by NIG PDF than Gaussian PDF. Figure 3 shows the probability plot for the empirical data for HL subband at level-1 and the goodness-of-fit for NIG and Gaussian distribution. Circles, in this figure represent the empirical data. This figure shows that NIG fits better than the Gaussian distribution. Figures 4 and 5 show the PDF and probability plot for LH subband at
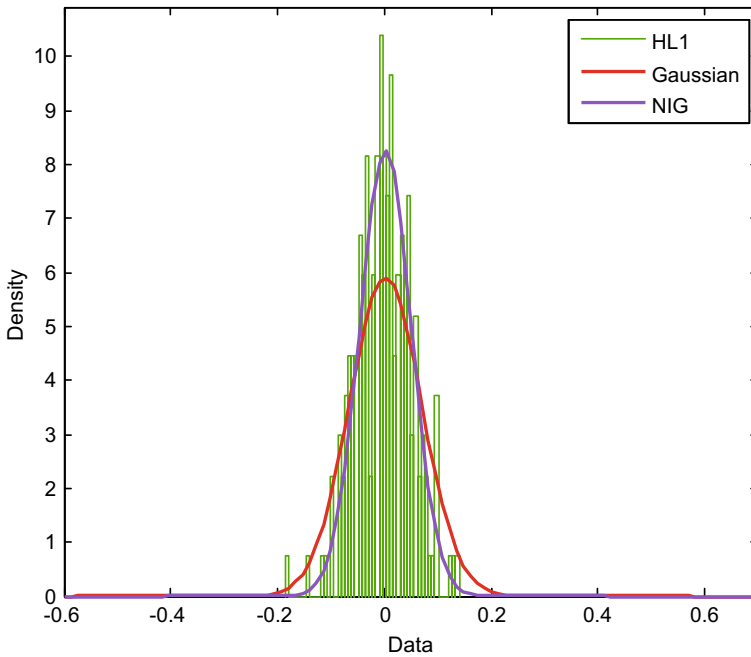
Fig. 2 Goodness-of-fit curve for HL subband at level-1(PDF plot)
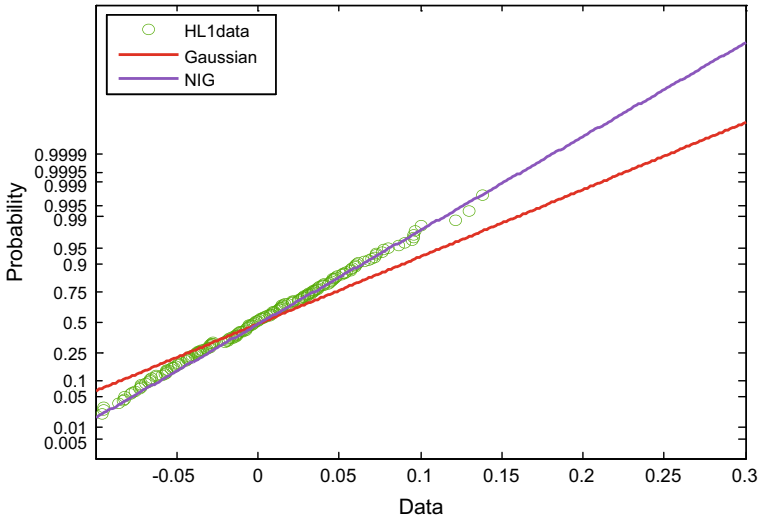
Fig. 3 Goodness-of-fit curve for HL subband at level-1(Probability plot)
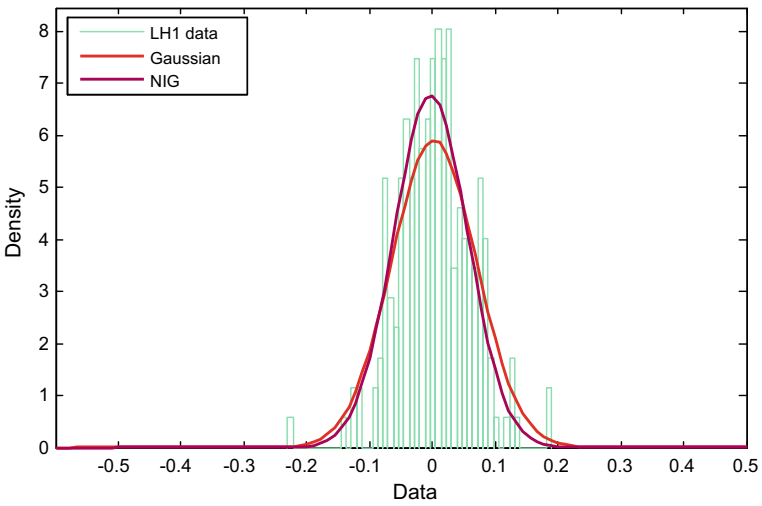
Fig. 4 Goodness-of-fit curve for LH subband at level-1(PDF plot)
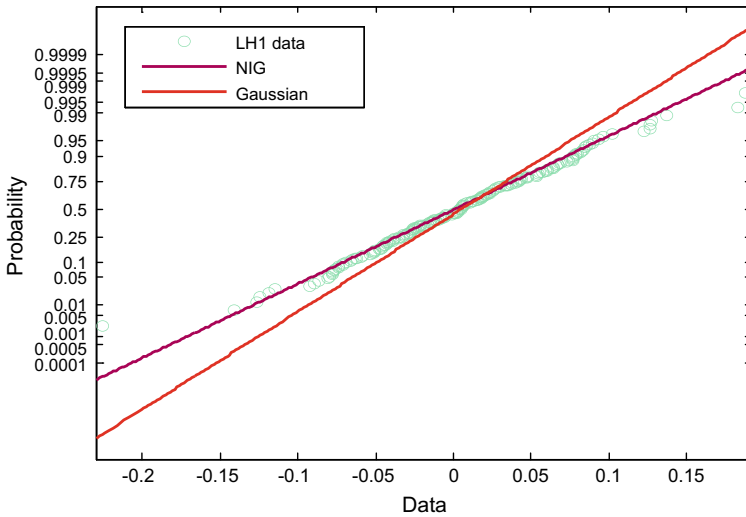
Fig. 5 Goodness-of-fit curve for LH subband at level-1(Probability plot)

level-1 respectively. Both the figures show the goodness-of-fit for NIG and Gaussian distribution. For both the figures NIG fits well for HL subband at level-1.

ML estimator has been proposed in this paper to estimate the NIG distribution parameters. By using the Hermite Gauss quadrature rule, the maximum likelihood estimation of distribution parameters is given by

$$\hat{\alpha} = \arg \max_{\alpha} \sum_{k=1}^{S} \ln \left(\left(\frac{1}{\sqrt{\pi}} \right) \sum_{j=1}^{R} w_j X \left(\left(Y(k) - \sqrt{2}\sigma_n x_j \right) \right) \right) \tag{4}$$

$$\hat{\delta} = \arg \max_{\delta} \sum_{K=1}^{S} \ln \left(\left(\frac{1}{\sqrt{\pi}} \right) \sum_{j=1}^{R} w_j X \left(\left(Y(k) - \sqrt{2}\sigma_n x_j \right) \right) \right) \tag{5}$$

where S = size of the subband, R = order of Hermite Polynomial, x_j = Root of Hermite Polynomial and w_j = weight of the root. $Y(k)$ is the kth wavelet coefficient. σ_n is the noise standard deviation, estimated using MAD estimator discussed in the next Sect. 2.2. The fundamental theorem of probability is used to find the signal standard deviation and is given by [12]:

$$signal\ standard\ deviation = \sqrt{\sum_x x^2 \widehat{X}(x) - \left(\sum_x x \widehat{X}(x) \right)^2} \tag{6}$$

$\widehat{X}(x)$ is the NIG distribution considering $\hat{\alpha}$ and $\hat{\delta}$ as the distribution parameters.
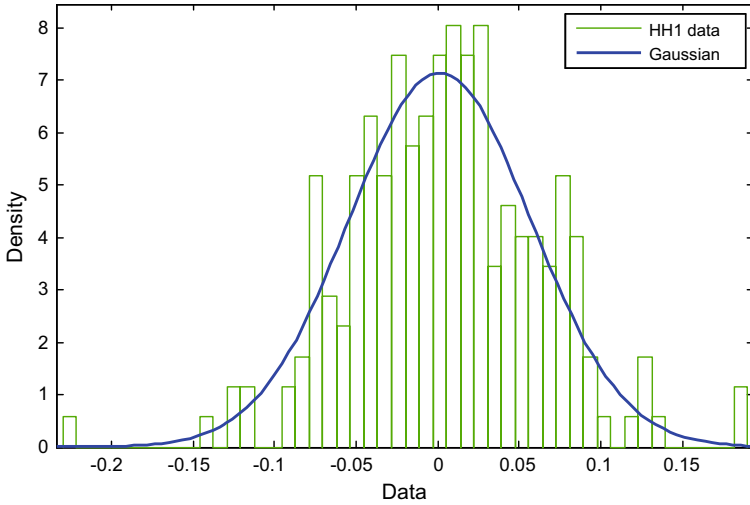
Fig. 6 Goodness-of-fit curve for HH subband at level-1(PDF plot)

2.2 MAD Estimator

Noise variance is calculated using the MAD estimator. This estimator is applied on HH subband at level-1. The Gaussian noise variance is estimated by modeling the HH subband by a Gaussian PDF at level-1 and it is expressed as [13]

$$noise\ variance(\sigma_\eta^2) = \left(\frac{median(HH1)}{0.6745}\right)^2 \tag{7}$$

Figures 6 and 7 show the goodness-of-fit curve for HH subband at level-1.

3 Proposed Methodology

The algorithm of the proposed method is discussed through the following steps. The block diagram is shown in Fig. 8.

Step 1 Input the MR image.
Step 2 Application of discrete wavelet transform (DWT) on the input image.
Step 3 Estimation of noise variance parameter by using MAD estimator using Eq. (7).
Step 4 Estimation of noise-free signal standard deviation by modeling wavelet coefficients with NIG PDF and using Eq. (6).
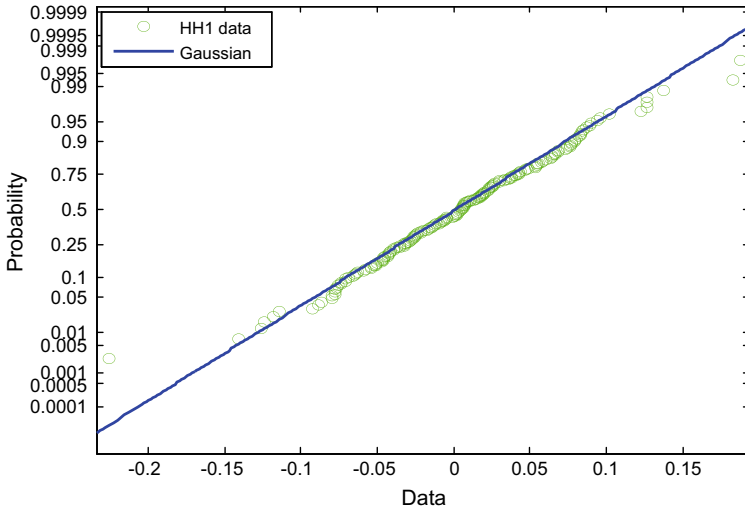
Fig. 7 Goodness-of-fit curve for HH subband at level-1(Probability plot)

Fig. 8 Block diagram of the proposed denoising procedure
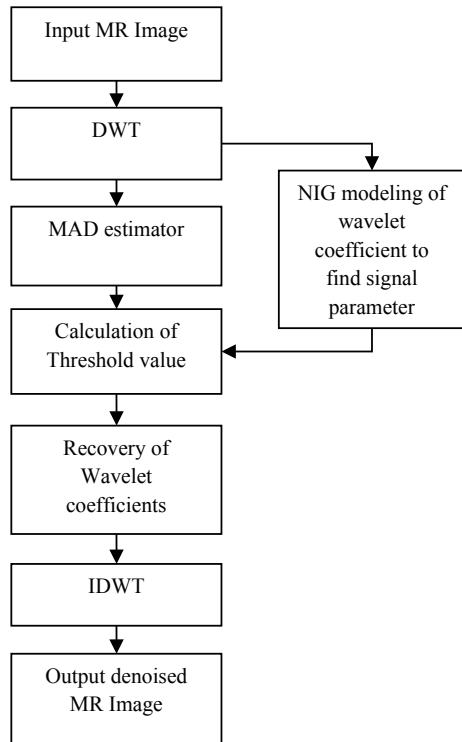
Step 5 Calculation of threshold value T_h using Eq. (2) and generation of noise-free wavelet coefficients using Eq. (1).

Step 6 Application of Inverse discrete wavelet waveform (IDWT) to the result obtained from step 5.

4 Simulated Results

Simulation of the proposed method is carried out on MR image of size 400×400 [14]. State-of-the-art methods such as Donoho's Soft threshold [15], Bayes–shrink [8] and Zaki et al. [3] are considered and compared with the proposed method.

Wavelet decomposition of level-3 is used and the simulation is carried out in MATLAB environment. PSNR and SSIM parameters are used as performance criterion [16]. Tables 1 and 2 give PSNR (dB) and SSIM values for state-of-the-art and proposed methods for different noise standard deviation values. From both the tables it can be seen that the proposed method performed better than the reported state-of-the-art methods. Figure 9 shows the original MR image, corresponding noisy image and denoised images results obtained from the proposed and other reported methods. Through visual inspection, it is clearly observed that the proposed denoising method gives the best visual quality among all considered methods.

Table 1 Comparison of PSNR (dB) performance for MRI for varying noise standard deviation

Method	Noise standard deviation				
	0.1	0.2	0.3	0.4	0.5
Donoho's soft threshold [15]	30.89	29.11	27.23	26.35	24.34
Bayes–shrink [8]	31.66	30.68	29.60	27.47	25.81
Zaki et al. [3]	33.15	31.93	30.12	28.39	26.96
Proposed method	35.79	32.47	31.66	29.66	27.88

Table 2 Comparison of SSIM performance for MRI for varying noise standard deviation

Method	Noise standard deviation				
	0.1	0.2	0.3	0.4	0.5
Donoho's soft threshold [15]	0.821	0.792	0.768	0.761	0.757
Bayes–shrink [8]	0.855	0.836	0.829	0.816	0.812
Zaki et al. [3]	0.949	0.925	0.913	0.899	0.894
Proposed method	0.967	0.951	0.943	0.920	0.925

Fig. 9 Denoising performance. **a** Original image. **b** Image corrupted with Gaussian noise of standard deviation 0.3. **c** Donoho's soft threshold [15]. **d** Bayes–shrink [8]. **e** Zaki et al. [3]. **f** Proposed method
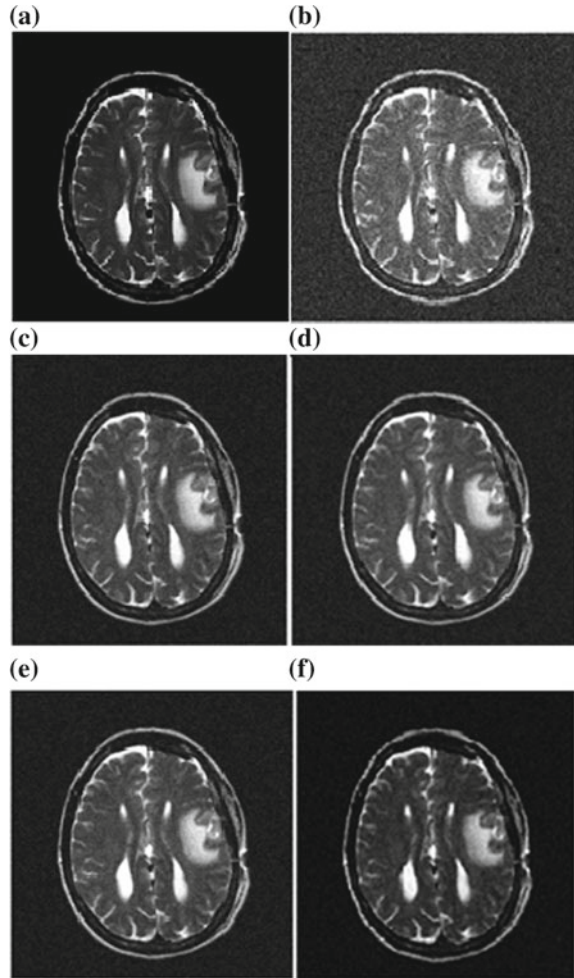
5 Conclusion

This paper presented a new statistical modeling based wavelet soft thresholding method to remove Gaussian noise from MR image. The detailed subband wavelet coefficients were modeled statistically using NIG distribution to estimate the signal parameters. A MAD estimator was used to find the noise variance. Utilizing the signal and noise parameters, a threshold value was obtained. Further improved wavelet soft thresholding was implemented to update the wavelet coefficients. From the obtained simulation result, it is concluded that the proposed denoising method performs better than the other considered contemporary denoising methods in terms of performance parameters such as PSNR and SSIM, as well as through visual inspection, for MR medical images.

References

1. Anand, C.S., Sahambi, J.S.: Wavelet domain non-linear filtering for MRI denoising. Magn. Reson. Imaging **28**(6), 842–861 (2010)
2. Bhuiyan, M.I.H., Ahmad, M.O., Swamy, M.N.S.: A new method for denoising of images in the dual tree complex wavelet domain. In: 2006 IEEE North-East Workshop on Circuits and Systems, pp. 33–36. IEEE (2006)
3. Zaki, F., Wang, Y., Yuan, X., Liu, X.: Adaptive wavelet thresholding for optical coherence tomography image denoising. In: Computational Optical Sensing and Imaging, pp. CTh4B–4. Optical Society of America (2017)
4. Sonali, Sahu, S., Singh, A.K., Ghrera, S.P., Elhoseny, M.: An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. Opt. Laser Technol. (an International Journal of Elsevier). https://doi.org/10.1016/j.optlastec.2018.06.061 (2018)
5. Sahu, S., Singh, H.V., Kumar, B., Singh, A.K.: Statistical modeling and gaussianization procedure based de-speckling algorithm for retinal OCT images. J. Ambient. Intell. Humaniz. Comput. (AIHC), (an International Journal of Springer). https://doi.org/10.1007/s12652-018-0823-2 (2018)
6. Sahu, S., Singh, H. V., Kumar, B., Singh, A. K.: A bayesian multiresolution approach for noise removal in medical magnetic resonance images. J. Intell. Syst. (2018)
7. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
8. Chang, Y.P., Hung, M.C., Liu, H.U.I.M.E.I., Jan, J.F.: Testing symmetry of a NIG distribution. Commun. Stat. Simul. Comput.® **34**(4), 851–862 (2005)
9. Sahu, S., Singh, H.V., Kumar, B.: A heavy-tailed levy distribution for despeckling ultrasound image. In: Fourth **IEEE** International Conference on Image Information Processing (ICIIP), pp. 1–5. Himachal Pradesh, India. December 21–23 (2017). https://doi.org/10.1109/ICIIP.2017.8313674
10. Barndorff-Nielsen, O.: Exponentially decreasing distributions for the logarithm of particle size. Proc. R. Soc. Lond. A **353**(1674), 401–419 (1977)
11. Bhuiyan, M.I.H., Ahmad, M.O., Swamy, M.N.S.: Wavelet based image denoising with the normal inverse Gaussian prior and linear MMSE estimator. IET Image Proc. **2**(4), 203–217 (2008)
12. Leon-Garcia, A.: Probability, statistics, and random processes for electrical engineering, 3rd edn. Prentice Hall (2007)
13. Rabbani, H., Vafadust, M., Gazor, S., Selesnick, I.: Image denoising employing a bivariate cauchy distribution with local variance in complex wavelet domain. In: 4th Digital Signal Processing Workshop, 12th Signal Processing Education Workshop, pp. 203–208. IEEE (2006)
14. Osirix.2014.OSIRIX DICOM image library. http://www.osirix-viewer.com/resources/diacom-image-library/
15. Donoho, D.L.: De-noising by soft-thresholding. IEEE Trans. Inf. Theory **41**(3), 613–627 (1995)
16. Sahu, S., Singh, H.V., Kumar, B., Singh, A.K.: De-noising of ultrasound image using Bayesian approached heavy-tailed Cauchy distribution. Multimed. Tools Appl. 1–18 (2017)

A Brief Survey on Hardware Realization of Two-Dimensional Adaptive Filters

Prabhat Chandra Shrivastava, Prashant Kumar, Manish Tiwari and Amit Dhawan

Abstract The efficient recognition of hardware of two-dimensional (2-D) adaptive filters is an immense problem of present state of art. The concept of the adaptive filter is given by Widrow in the decade of sixty and the mathematical expression of 2-D adaptive filters is introduced by Hadhoud in the decade of ninety. Further, several researchers give the different type of adaptive algorithms for the hardware realization of 2-D adaptive filters. The least mean square (LMS) algorithms are too renowned due to its accomplished convergence properties and simplicity to implement in hardware. In this paper, we present a concise compendium of the efficient hardware structure of 2-D adaptive filters.

Keywords 1-D and 2-D adaptive filter · LMS algorithms · Mean square error · Normalized LMS · 2-D LMS algorithm

1 Introduction

The adaptive technique is very important in digital signal processing, communication, an adaptive equalizer, image enhancement, image compression, and so on [1–3]. By the selection of convenient time constants and other dimensional (2-D) adaptive digital filters have a broad range of applications in image and video processing. The hardware realization of 2-D adaptive digital filters can easily be simulated on general purpose processor but these processors may not be appropriate for application

P. C. Shrivastava (✉) · P. Kumar · M. Tiwari · A. Dhawan
Department of Electronics & Communication Engineering, MNNIT, Allahabad 211004, India
e-mail: prabhatphd@gmail.com

P. Kumar
e-mail: pkumar.mnnit@gmail.com

M. Tiwari
e-mail: mtiwari@mnnit.ac.in

A. Dhawan
e-mail: dhawan@mnnit.ac.in

in real time. Hence, applications such as real-time image and video processing that involve high data rate which needs devoted hardware architecture to find the high throughput requirements. The introduction of VLSI technology has a significant impact on the realization and the development of dedicated hardware such as ASIC design. Due to the complexity of the development of hardware realization of 2-D adaptive algorithms, has been, however, much slower than the 1-D case.

For real-time applications, high-speed architectures are required. The high-speed system is very essential for high data-rate-based applications. To achieve such high-speed applications two design methodologies pipelining in addition to parallel processing are used. For dedicated VLSI processors, the systolic-array-based architecture, in exclusive, is receiving a large amount of consideration. Systolic arrays are a set of connections of several similar type cells which satisfy the limitation of modularity, position and pipelineability. Due to the regular and identical structures the systolic arrays providing an identical and high-speed throughput rate. For the designing of adaptive filters, the least mean squares (LMS) algorithm is broadly used for real-time adjustment of filter coefficients [4]. The LMS algorithm is simple but its performance is robust. For VLSI design of LMS algorithm without considering the delays is very difficult. The performance of adaptive filter has been improved due to its real-time updation of the weight of filter coefficient. The improved feedback signal approaches to the desired signal after several iterations. The number of iteration depends upon the quantity of the feedback error.

For parallel and/or pipelined execution of the LMS algorithm latches have to be proposed to execute the proper adaptation process and then LMS algorithm is called delayed least mean squares (DLMS) algorithm. The delays disgrace the execution of the structure when analogizing to the LMS algorithm. A VLSI design of Delay-LMS (DLMS) algorithm for approximating LMS adaptive filter has been suggested by Long et al. [5]. Further, some researches [6, 7] have been proposed DLMS adaptive filter which architecture is based on the systolic array. A new structure suggested in [8, 9] that come close to the convergence factor of the LMS algorithm. In this suggested technique, they required the larger area to calculate feedback error. The throughput of the structure has low if delay D decreases. Alternatively, a much larger amount of computations delays are required for 2-D adaptive digital filter than the 1-D, so high throughput and possible characteristics are desirable.

Firstly, Hadhoud and Thomas [10] have introduced the 2-D adaptive LMS algorithm (TDLMS), which is an extension of the LMS adaptive algorithm which is discussed in [11]. An IIR filter based adaptive filter architecture is introduced in [12]. A weakness of the methods described in [10–12] is the procedure to select constant step size. For adaption of desire filter coefficients of 2-D adaptive filter, we have used a convergence factor in each iteration. The adjustment of the proper convergence factor is very sensitive because a small variation can produce an unfortunately large amount of change in adaptation speed and accuracy. Hence, to get the suitable convergence factor value, hit and trial methods are expected, and the result is commonly a balanced achieved between low evaluated errors versus fast convergence. Furthermore, when the incoming signals for filtering are time-varying, which is generally happen then constant convergence factor cannot consentaneously

be modified for the filter parameters. To overcome the problem of constant convergence factor, a new technique has been proposed where the time-varying optimum convergence factors is used. The variable step-size approach was established effectively in the one-dimensional case [13, 14]. The convergence characteristic of the LMS is negotiated by the step-size factor μ [1]. Due to better-quality performance and simple calculation, the Least Mean Square (LMS) algorithms are commonly used an algorithm for adaptive digital filters.

In the concise review paper presents a survey on the hardware architecture of 2-D adaptive filter based on two-dimensional least mean square (TDLMS) algorithm. This algorithm is suitable for hardware implementation because they do not require any a former acquaintance concern signals. The arrangement of the paper is specified as in the upcoming section; we have deliberated about the basic technique of 2-D adaptive filter with suitable mathematical expression and related features. Section-III explains a concise survey on hardware architecture of 2-D adaptive filters. Finally, in Sect. 4, concluding remarks of the paper are discussed.

2 Basic Technique of Adaptive Filters

In this Section, we discuss the ultimate concepts and algorithms, error calculation, LMS and DLMS based algorithms of adaptive filters. The main focus of this section is mathematical expression and architecture of 2-D adaptive filter. The few decades before, in the field of DSP, and exclusively adaptive signal processing, has evaluated extremely due to the enhancement of the technology and algorithms. The adjustment mechanism of filters coefficients of adaptive filters are in real-time manner while the calculations of conventional filter coefficient are carried out in an off-line mode. Moreover, an adaptive filters, track the optimal behaviour of a slowly varying environment due to its real-time self-modifying characteristic of filter coefficients. Before the hardware designing of adaptive filters many important aspects are taken in account such as filter structures, rate of convergence, misadjustment and efficient tracking of filter coefficients. The filter structure means which type of filters are used in designing of adaptive filter. Here, we may be use both finite impulse response (FIR) and infinite impulse response (IIR). Actually, the adaptive filter structure has two units one is filtering and other is weight-updation unit as depicted in Fig. 1. In Fig. 1, the input signal is denoted by $x(n)$, $d(n)$ represents the expected output signal, $y(n)$ is the adaptive filter output then the error signal is

$$e(n) = d(n) - y(n) \tag{1}$$

The adaptive filter coefficient vector $w(n)$ of the filter updated according to the values of the error signal $e(n)$. The adaption process is going to continue till it reaches the desire signal in a statistical sense. In Fig. 1, we are using FIR filter for filtering.

Fig. 1 Block diagram of
FIR based adaptive filter
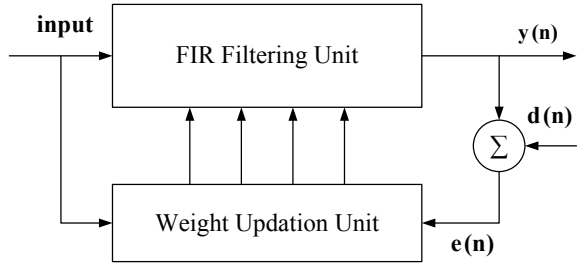
2.1 Error Measurements

The measurements of error is very important in hardware designing of adaptive filters
because the time taken by the adaption process is totally depends upon the amount
of the error. The error is also defined by error function $e(n)$ or cost function, defined
as difference between the expected or reference signal to the adaptive filter output.
Three types of norm of error function are defined which are weighted least squares
(WLS), mean square error (MSE), and the instantaneous square error (ISE) among
all of three, MSE is very easy and suitable for error calculation in adaptive filters.
Here, we have focused only on MSE.

The mean square error (MSE) is defined as

$$\xi(n) = E\big[e^2(n)\big] = E\big[|d(n) - y(n)|^2\big] = E\Big[\big|d(n) - w^T(n)x(n)\big|^2\Big] \qquad (2.1)$$

$$= E\big[d^2(n)\big] - 2w^T(n)E\big[d(n)x^T(n)\big] + w^T(n)E\big[x(n)x^T(n)\big]w(n) \qquad (2.2)$$

$$= E\big[d^2(n)\big] - 2w^T\mathbf{P} + w^T\mathbf{R}w \qquad (2.3)$$

where **P** is the cross-correlation between the desired signal and the input signal and
R is input-signal correlation matrix and are defined as

$$\mathbf{P} = E\big[d(n)x^T(n)\big] \quad \text{and} \quad \mathbf{R} = E\big[x(n)x^T(n)\big] \qquad (2.4)$$

The immediate change in value of error function with reference to adaptive filter
weight vector is defined as

$$\nabla_w\xi(n) = -2\mathbf{P} + 2\mathbf{R}w \qquad (2.5)$$

where ∇_w is called gradient vector of the MSE function.

To minimize the MSE cost function, equating the gradient vector in Eq. (2.5) to
zero and taking **R** is non-singular, then the Wiener solution [15] $\mathbf{w}_0$ and its value is
defined as

$$\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{P} \tag{2.6}$$

The filter coefficient vector $\mathbf{w}(n)$ iteratively adjusting at each time instant n, and defined as

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \frac{\mu}{2}\nabla_w\xi(n) \tag{2.7}$$

where μ is called convergence factor, which adjusts the step size between two successive coefficient vectors $\mathbf{w}(n)$ and $\mathbf{w}(n+1)$.

2.2 Adaptive Algorithms

A variety of algorithms are presented to find the best possible solution of the adaptive filter for the error functions. Each method developed an adaptation algorithm that modifies the coefficients of the adaptive filter in order to reduce the amount of error. There are three types of the algorithms used; these are the LMS, the DR, and the Recursive Least Square (RLS). Each type has specific features to compute the design complexity and manage the speed of convergence. Among all these algorithms, generally, we have used LMS algorithms because of its simplicity and easy implementation of hardware realization. Further, after some modification, verities of LMS algorithms such as normalized LMS (NLMS), delay-LMS (DLMS), data-reusing LMS (DR-LMS), and variable step-size LMS are given in literature [16–23]. Here, we have briefly discussed the LMS algorithms.

The LMS Algorithms for 1-D is defined as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu e(n)\mathbf{x}(n) \tag{2.8}$$

and

$$e(n) = d(n) - \mathbf{w}^{\mathrm{T}}(n-1)\mathbf{x}(n) \tag{2.9}$$

The very important parameter, the convergence factor μ [1, 17, 24] plays an significant preamble in the convergence characteristics of the algorithm as well as in its stability position. And the values of μ will be

$$0 < \mu < \frac{2}{\mathrm{tr}[\mathbf{R}]} \tag{2.10}$$

where $\mathrm{tr}[\mathbf{R}]$ is trace operator of the matrix.

The LMS algorithms for 2-D adaptive filter [10] are given below.

Let us consider the desire image D, the corrupted image X and weight matrix of filter coefficients $\mathbf{w}(n)$ of dimension $M \times M$, $M \times M$ and $N \times N$ of 2-D adaptive FIR filter respectively then the output equation is defined as

$$y(m, n) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x(m - i, n - j)\, w_k(i, j), \qquad (3)$$

where $x(m, n)$, $w_k(i, j)$ and $y(m, n)$ are the current processed corrupted-input pixel, weight of the filter-coefficient at kth iteration and the output pixel, respectively. Iteration number k can be given as $k = mM + n$, where $0 \leq m \leq M - 1$ and $0 \leq n \leq M - 1$. The weight of the filter coefficient w_k in each iteration may be updated by the TDLMS algorithm. If $e(k)$ represents the error in the output pixel $y(m, n)$ with respect to the desired-pixel $d(m, n)$ then the error signal $e(k)$ at k^{th} iteration can be given as,

$$e(k) = d(m, n) - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x(m - i, n - j)\, w_k(i, j). \qquad (4.1)$$

Mean Square Error (MSE) of $e(k)$ can be calculated as,

$$\text{MSE} = \text{E}\{e(k)^2\}. \qquad (4.2)$$

Now from (4.1) into (4.2), we get

$$\text{E}\{e(k)^2\} = \text{E}\left\{ d^2(m, n) + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} x(m - i, n - j) x(m - p, n - q) \right.$$
$$\left. w_k(i, j) w_k(p, q) - 2d(m, n) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x(m - i, n - j)\, w_k(i, j) \right\} \quad (4.3)$$

weight of the filter coefficient corresponding to the least value of $\text{E}\{e(k)^2\}$ may be calculated using the maxima-minima theorem. After simplification, the adaptive weight-updation algorithm from [10] can be expressed as,

$$w_{k+1}(i, j) = w_k(i, j) + 2\mu e(k) x(m - i, n - j) \qquad (4.4)$$

The convergence factor μ for 2-D adaptive filter is given as.

$$0 < \mu \leq \frac{1}{\lambda_{\text{max}}} \qquad (4.5)$$

where

$$\lambda = E\|x(m, n) x(m - i, n - i)\|. \qquad (4.6)$$

3   A Brief Survey

In this segment, we will consider concisely of the available literature survey of hardware-based structure of 2-D adaptive filters. A number of papers have appeared in the literature in which the hardware realization of the 2-D adaptive filter has been constructed using different adaption algorithms. Hadhoud and Thomas [10] have presented the mathematical implementation of 2-D adaptive filter algorithm which is a capaciousness of the popular LMS adaptive algorithm of Widrow [16]. The TDLMS algorithm is very useful for the application of digital signal processing such as image and video processing, data compression and image enhancement. Soni et al. [25] describe image enhancement using the TDLMS algorithm. Youlal et al. [26], has proposed a 2-D joint process lattice algorithm to tackle the convergence factor problem in 2-D adaptive digital filters and this algorithm has the higher convergence rate than the TDLMS algorithm. The lattice algorithm for 2-D adaptive filters is first evaluated. Convergence properties of the algorithm are included for the 2-D adaptive lattice LMS (TDAL-LMS) case. At the cost of the slightly complex structure, the 2-D normalized LMS algorithm (TDAL-NLMS) gives faster converging to that of TDAL-LMS. Cho et al. [27] has suggested an automatically step-size adjustment method for the TDLMS algorithm to improve the convergence rate as well as estimation error in 2-D adaptive digital filters.

The 2-D block adaptive filtering algorithms are presented by Mikhael and Ghosh [28, 29] Wang and Wang [30] and by Tan and Chen [31] for IIR filtering. The block adaptive filtering algorithms tend to decrease the computational time with good convergence at the cost of the increase in hardware. Cho and Priemer [32] have suggested a new concept which automatically adjusts the convergence factor of LMS to renovate image enhancement as the change of contrast variation over an entire image. The direct form implementations of 2-D adaptive FIR filters have the very long critical path due to inner product calculation for the weight updation and the computing the output. This long critical-path problem may be tackled by the pipelined based architecture as suggested by [8, 9, 33–37]. The pipelined architecture increases the latency in output computation but it also increases the throughput of the architecture. Harada et al. [35] has suggested a pipelined architecture for the 2-D normalized LMS (TDNLMS) adaptive digital filters. In this architecture, it is shown that a stable and a small critical path can be achieved without producing output latency.

Systolic pipelined architectures of 2-D adaptive FIR filters are suggested by [36, 37]. In this suggested architectures, they have reduced the critical-path delay using the multistage pipelining. Van and Feng [38] again proposed a new technique of an efficient systolic structure for the DLMS adaptive FIR digital filter. In this technique, a processing element is designed for optimizing tree-level rule. It has a higher convergence rate and throughput in the word-level than that of the structure of conventional DLMS can be found without compromising the properties of the systolic-array structure. Santhaa and Vaidehi [39] suggested a new architecture for 1-D and 2-D FIR adaptive filter. The architecture is independent to the length of the filter and

has a higher speed up and the convergence performance compares with the conventional LMS algorithm. Dutta et al. [40] have proposed a semi-automatic mapping methodology for the 2-D adaptive filter which is applicable in image processing. In this technique, array architectures of the parallel processor can be used as hardware accelerators for the application of plenty of data-flows dominant. In recent time, many authors [41–47] gives the efficient hardware realization of 1-D and 2-D FIR, IIR and adaptive digital filters all these architecture have shown the momentous reduction in power, area and critical-path delay compare to other existing structures.

4 Conclusions

The beauty of adaptive filter is that the values of filter coefficients are updated automatically according to the desired output signals. The adaptive filter structure is a unified structure of filtering and weight updating block. The filtering block may be FIR or IIR and weight updating block updates their filter coefficients according to which types of adaptive algorithm is used. The LMS algorithm is very popular and efficient for hardware realization. The concept of the most of the adaptive algorithms is inspired by the LMS as discussed above. The 2-D adaptive filters give high throughput and reduction in latency and useful for high-speed- and real-time applications. The concluding remark on this concise paper is that the hardware structure of 2-D adaptive filter has better scope for researchers because only a few works of literatures are given.

References

1. Haykin, S.: Adaptive Filter Theory. Prentice-Hall, Englewood Cliffs, NJ (1986)
2. Sid-Ahmed, M.A.: Image Processing: Theory, Algorithms and Architectures. McGraw-Hill, New York (1995)
3. Lu, W.S., Antoniou, A.: Two-Dimensional Digital Filters. Marcel Dekker, New York (1992)
4. Madisetti, V.K., Williams, D.B.: The Digital Signal Processing Handbook. IEEE Press, New York (1998)
5. Long, G., Ling, F., Proakis, J.G.: The LMS algorithm with delayed coefficient adaptation. IEEE Trans. Acoust., Speech, Signal Process. ASSP **37**(9), 1397–1405 (1989)
6. Herzberg, H., Haimi-Cohen, R., Be'ery, Y.: A systolic array realization of an LMS adaptive filter and the effects of delayed adaptation. IEEE Trans. Signal Process. **40**(11), 2799–2803 (1992)
7. Meyer, M.D., Agrawal, D.P.: A high sampling rate delayed LMS filter architecture. IEEE Trans. Circuits Syst.-II CAS **40**(11), 727–729 (1993)
8. Douglas, S.C., Zhu, Q., Smith, K.F.: A pipelined LMS adaptive FIR filter architecture without adaptation delay. IEEE Trans., Signal Process. **46**(3), 775–779 (1998)
9. Matsubara, K., Nishikawa, K., Kiya, H.: Pipelined LMS adaptive filter using a new look-ahead transformation. IEEE Trans. Circuits Syst., 11, CAS **46**(1), 51–55 (1999)
10. Hadhoud, M.M., Thomas, D.W.: The two-dimensional adaptive LMS (TDLMS) algorithm. IEEE Trans. Circuits Syst. **35**, 485–494 (1988)

11. Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Doodlin, R.C., Zeidler, J.R., Hearn, R.H., Dong, E.: Adaptive noise cancelling: principles and applications. Proc. IEEE **63**, 1692–1716 (1975)
12. Tan, A.C., Chen, S.T.: Two-dimensional adaptive LMS IIR algorithm. In: 1993 IEEE International Symposium on Circuits and Systems, vol. 1, pp. 299–302, May (1993)
13. Wu, F.H.: The optimum adaptive algorithms and applications for noise cancelation. M.S. thesis, West Virginia University, Morgantown, WV, May (1984)
14. Mikhael, W.B., Wu, F.H., Kazovsky, L.G., Kang, G.S., Fransen, L.L.: Adaptive filters with individual adaption of parameters. IEEE Trans. Circuits Syst., CAS **33**, 677–686 (1986)
15. Wiener, N.: Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications. Wiley, New York (1949)
16. Widrow, B., Hoff, M.E.: Adaptive switching circuits. In: Proceedings of WESCON Convention Record, part 4, pp. 96–140 (1960)
17. Widrow, B., Stearns, S.D.: Adaptive Signal Processing. Prentice-Hall (1985)
18. Sayed, A.H.: Fundamentals of Adaptive Filtering. Wiley, Hoboken, NJ (2003)
19. Treichler, J.R., Johnson Jr., C.R., Larimore, M.G.: Theory and Design of Adaptive Filters. Wiley, NewYork, NY (1987)
20. Macchi, O.: Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission. Wiley (1995)
21. Soraghan, J.J., Stewart, R.W., Durrani, T.S.: Comparative analysis between non-canonical LMS and LMS adaptive filtering. Electron. Lett. **27**(11), 947–950 (1991) (Department of Electrical & Electronic Engineering, Strathclyde University, Glasgow, UK)
22. Harris, R.W., Chabries, D.M., Bishop, F.A.: A variable step (VS) adaptive filter algorithm. IEEE Trans. Acoust., Speech, Signal Process., ASSP **34**, 309–316 (1986)
23. Nagal1, R., Kumar, P., Bansal, P.: A survey with emphasis on adaptive filter, structure, LMS and NLMS adaptive algorithm for adaptive noise cancellation system. J. Eng. Sci. Technol. Rev. **10**(2), 150–160 (2017)
24. Diniz, P.S.R.: Adaptive Filtering: Algorithms and Practical Implementation, 3rd edn. Springer, New York, NY, USA (2008)
25. Soni, T., Zeidler, J.R., Ku, W.H.: Performance Evaluation of 2-D adaptive prediction filters for detection of small objects in image data. IEEE Trans. Image Process. **2**(3), 327–340 (1993)
26. Youlal, H., Janati, M., Najim, M.: Two-dimensional joint process lattice for adaptive restoration of images. IEEE Trans. Image Process. **1**, 366–378 (1992), ISSN 1057-7149
27. Cho, H., Priemer, R.: Automatic step size adjustment of the two-dimensional LMS algorithm. In: Proceedings of the 37th Midwest Symposium on Circuits and Systems 1994, vol. 2, pp. 864–867 (1994)
28. Mikhael, W.B., Ghosh, S.M.: Two dimensional block adaptive filtering algorithms. In: Proceedings of ISCAS'gB, pp. 1219–1222, May 1992
29. Mikhael, W.B., Ghosh, S.M.: Two-dimensional block adaptive filtering algorithms with optimum convergence factors. IEEE Trans. Circuit Syst.-11: Analog. Digit. Signal Process. **42**(8), 505–515 (1995)
30. Wang, T., Wang, C.-L.: A new two-dimensional block adaptive FIR filtering algorithm and its application to image restoration. IEEE Trans. Image Process. **7**, 238–246 (1998). ISSN 1057-7149
31. Tan, A., Chen, S.T.: Image enhancement with 2-D adaptive LMS-based recursive filters. In: Proceedings of International Conference on Signal Processing, Applications and Technology, Boston, MA, vol. 2, pp 1068–1072, Oct. 24–26 1995
32. Cho, H., Priemer, R.: Automatic step size adjustment of the two-dimensional LMS algorithm. IEEE conference, 0-7803-2428-5/95, pp. 864–867
33. Matsubara, K., Nishikawa, K., Kiya, H.: 2-D pipelined adaptive filters based on 2-D delayed LMS algorithm. IEICE Trans. E80 **A**(6), 1009–1014 (1997)
34. Kimijima, T., Nishikawa, K., Kiya, H.: Pipelining of 2-dimensional adaptive filters based on the LDLMS algorithm. In: Proceedings of the 1998 IEEE International Symposium on Circuits and Systems 1998. ISCAS '98, vol. 5, pp. 190–193 (1998)

35. Harada, A., Nishikawa, K., Kiya, H.: A pipelined architecture for the normalized LMS adaptive digital filters. In: IEEE Asia-Pacific Conference on Circuits and Systems, Chiangmai, Thailand, 24–27 Nov. 1998
36. Van, L.D., Tenqchen, S., Chang, C.H., Feng, W.S.: A tree-systolic array of DLMS adaptive filter. IEEE J. Accoustics, Speech Signal Process. **3**, 1253–1256 (1999)
37. Van, L.D., Feng, W.S.: Efficient systolic architectures for 1-D and 2-D DLMS adaptive digital filters. In: Proceedings of IEEE Asia Pacific Conference on Circuits Systems, Tianjin, China, Dec. 2000, pp. 399–402
38. Van, L.D., Feng, W.S.: An efficient systolic architecture for the DLMS adaptive filter and its applications. IEEE Trans. Circuits Systems—II: Analog. Digit. Signal Process. **48**(4) (2001)
39. Santhaa, K.R., Vaidehi, V.: Design of efficient architectures for 1-D and 2-D DLMS adaptive filters. Sciencedirect VLSI J. **40**, 209–225 (2007)
40. Dutta, H., Hannig, F., Teich, J., Heigl, B., Hornegger, H.: A design methodology for hardware acceleration of adaptive filter algorithms in image processing. In: The IEEE Computer Society, Application-specific Systems, Architectures and Processors (ASAP '06), 0-7695-2682-9/06
41. Meher, P., Park, S.: High-throughput pipelined realization of adaptive FIR filter based on distributed arithmetic. In: VLSI Symposium on Technology Digital, pp. 428–433 (2011)
42. Meher, P., Park, S.: Critical-path analysis and low-complexity implementation of the LMS adaptive algorithm. IEEE Trans. Circuit Syst.-I, Regul. Pap. **61**(3), 778–788 (2014)
43. Mohanty, B., Meher, P., Singhal, S., Swamy, M.: A high performance VLSI architecture for reconfigurable FIR using distributive arithmetic. J. Integr., VLSI J. **54**, 37–46 (2016)
44. Park, S., Meher, P.: Efficient FPGA and ASIC realization of a DA-based reconfigurable FIR digital filter. IEEE Trans. Circuits Syst. II. Express Briefs **61**(7), 511–515 (2014)
45. Tiwari, A., Kumar, P., Tiwari, M.: High throughput adaptive block FIR filter using distributed arithmetic. In: 2016 1st India International Conference on Information Processing (IICIP), Delhi, 2016, pp. 1–6. https://doi.org/10.1109/iicip.2016.7975385
46. Kumar, P., Shrivastava, P.C., Tiwari, M., Dhawan, A.: ASIC Implementation of area-efficient, high-throughput 2-D IIR filter using distributed arithmetic. Circuits Syst. Signal Process. **37**, 2934 (2018). https://doi.org/10.1007/s00034-017-0698-z
47. Kumar, P., Shrivastava, P.C., Tiwari, M., Mishra, G.R.: High-throughput, area-efficient architecture of 2-D block FIR filter using distributed arithmetic algorithm. Circuits Syst. Signal Process., 1–15 (2018). https://doi.org/10.1007/s00034-018-0897-2

A Survey on H_∞ Control Techniques

Abhay Vidyarthi and Manish Tiwari

Abstract The stability investigation of 2-D discrete systems has been one of the most explored areas for the researchers due to its wide variety of practical and diversified applications. The 2-D discrete systems stability analysis using H_∞ control technique has always been an active field of research. The H_∞ control techniques play a vital role in the analysis and design work of control and signal processing systems. A detailed survey report of 2-D discrete systems has been presented on the stability analysis using H_∞ control techniques.

Keywords H_∞ control · 2-D discrete systems · Static state feedback controller · Bounded real lemma · Robust stability

1 Introduction

In signal processing field, for the system performance analysis H_2 and H_∞ norms are of vital importance. Let us try to differentiate them in view of filter design.

H_2 or Kalman filtering presumes the knowledge of the dynamic model as well as knowledge of the spectral content of incoming noises or disturbances. If there is uncertainty in the system model, then Kalman filter type of estimators may not be robust enough [1] and this has motivated the development of H_∞ filtering approach, which provides an upper bound on the maximum energy gain between the filtering error and the noise input. H_∞ filtering does not require any statistical assumptions on the exogenous signals which is surely an advantage over Kalman filtering.

The H_2 norm for a stable and causal system $F(z)$ is defined by

$$\|F\|_2 := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| F(e^{j\omega}) \right|^2 d\omega \right)^{1/2} \tag{2}$$

A. Vidyarthi (✉) · M. Tiwari
ITM College, Gwalior, MP, India
e-mail: vidyarthi.abhay@gmail.com

The H_∞ norm for such a system is defined by

$$\|F\|_\infty := \sup_{\omega \in (-\pi,\pi)} \left|F(e^{j\omega})\right| \tag{3}$$

Let us differentiating the two by considering the following inverse-filtering problem.

In inverse-filtering problem if we have two filters $H(z)$, and $K(z)$ then $H(z) K(z) \approx 1$.

Using H_2 norm, stable and causal $K(z)$ can be determined that minimizes $\| H(z) K(z) - 1\|_2$ [2, 3].

On the other hand in case of H_∞ norm, stable and causal $K(z)$ can be determined that minimizes $\| H(z) K(z) - 1\|_\infty$ [4].

The difference is that H_2 tries to minimize the average magnitude of the error system $H(z) K(z) - 1$ while H_∞ tries to minimize the maximum magnitude. The following picture shows an example of the two designs (Fig. 1).

This comparison of frequency versus error using H_2 and H_∞ control approach, clarifies the difference. H_2 designed filter is giving better results for almost all frequencies but is *fragile* at the mid-value of the frequency. Hence, the performance of H_2 is better if the frequency components of the input signals does not lie at the mid value of the frequency. While the filter designed using, H_∞ filtering approach guarantees a certain error level for *all* frequencies. That is, optimization using H_∞ filtering approach leads to *robustness* against uncertainty for the whole frequency band of the incoming exogenous input signals.

Consider a system as represented by this block diagram (Fig. 2).

Fig. 1 Comparison of frequency versus error using H_2 and H_∞ control approach
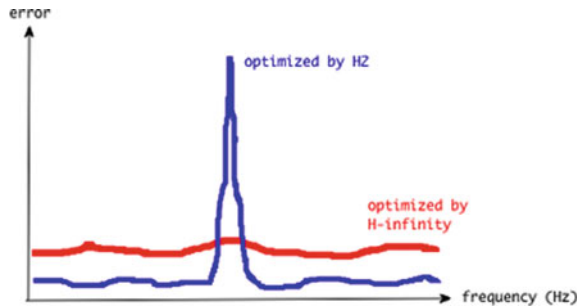
Fig. 2 The block diagram of generalized plant
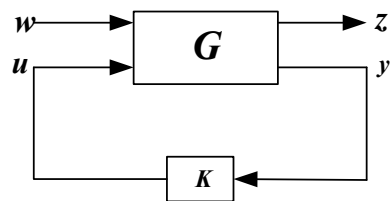
Here **K** is the controller and **G** is the generalized plant. The generalized plant can be a representation of any linear time invariant system represented either in continuous time or in discrete time. In the block diagram we are showing two different types of inputs that are acting on the system and also the system has two types of outputs. The disturbance that is acting on the system is represented by the symbol **w**. The controlled output **z** is the output of the system. The other output **y** is the measured output, which we calculate from the measurement we make on the system and which helps us to choose the input **u**. The choice of **u** works as a tool which minimizes the effect of **w** on **z**.

Signal **w** contains all external inputs, including disturbances, sensor noises and commands, signal **z** represents a controlled output, signal **y** represents the measured output and **u** is the control input.

The resulting closed loop transfer function from **w** to **z** is denoted by T_{zw} The basic theme in dealing with H_∞ control problem is that the H_∞ norm [5–7] of the closed loop transfer function is less a scalar number γ i.e. $\|T_{zw}\|_\infty < \gamma$.

1.1 Finite Horizon H_∞ Control Problem

The basic aim of H_∞ control problem is to minimize the H_∞ norm. The H_∞ norm is the maximum value of all the disturbances **w** that is going into the system and coming out of that system. Typically in a H_∞ control problem the measurement of this energy is carried out over an infinite time interval $[0, \infty)$. However, practically this does not seem realistic. Due to this reason the infinite horizon H_∞ control problem can be reduced to a finite horizon H_∞ control problem where also the prime objective is to minimize the same norm but in this case the energy is measured over a finite time interval $[0, T]$ for some given $T > 0$.

In optimal H_∞ controller there is an effort to find all admissible controllers K such that $\|T_{zw}\|_\infty$ is minimized [8]. However as per [9], finding such an optimal H_∞ controller is both numerically and theoretically involved, which is unlikely to the standard H_2 theory in which finding an optimal controller can be simply achieved by solving two Riccati equations without iterations [10]. Finding an optimal H_∞ controller can be achieved through a chain of successive solutions of H_∞ suboptimal problems [11] with γ approaching the $\gamma_{\min}$ value. Optimal H_∞ norm is although useful theoretically however in practice, it is not essentially required to design an optimal H_∞ controller [12]. With respect to norm sense, the suboptimal controllers are quite close to the optimal controllers and are cheaper to design. The added advantage in designing a suboptimal controller is that the bandwidth requirement is also less.

For designing a suboptimal controller, a scalar number $\gamma > 0$ is given and an effort is made to find all admissible controllers K such that $\|T_{zw}\|_\infty < \gamma$. In the presented work, effort has been made in making an extensive survey report of H_∞

control approach for the stability analysis of 2-D discrete systems, which has not been done since long and this motivates us to compile all those findings and present a survey report of the same. To the best of author's knowledge no such report has been published so far in the literature.

2   A Brief Survey Report

Thus with the development in the H_∞ control theory there has been a rapid advancement in the analysis and design of robust 1-D control systems [13, 14].

However it is not possible to extend all the well-established results of H_∞ control techniques of linear 1-D control systems to 2-D systems due to their analytical and structural complexities [15, 16]. As the 2-D systems are also to be dealt with disturbances and modeling uncertainties. H_∞ control based performance analysis is examined to be more suitable than any other performance analysis [15], particularly when the system is subjected to disturbances and modeling uncertainties. It is therefore a requirement to apply robust H_∞ control techniques

The 2-D H_∞ problem was first addressed in [17], in which the problem of robust H_∞ control for 2-D systems has been discussed using polynomial approach (using spectral factorization). Similar to the 1-D systems, the bounded real lemma for linear 2-D discrete systems is developed with the help of Riccati inequality approach. From the derived lemma finite horizon and infinite horizon H_∞ filtering problem based on the linear matrix inequality (LMI) approach is addressed in [18].

The main advantage of using H_∞ control based technique is that the performance analysis of the system under consideration is carried out depending upon the worst value of the system energy gain [15]. Even if the performance optimality is achieved, using H_∞ noise performance analysis, it could not guarantee the robust stability of the system. Therefore, the robust stability constraint is to be examined along with the H_∞ noise performance analysis.

Many notable results have been given in the literature for 1-D and 2-D systems [15, 18–30], based on this idea. In [18], a 2-D bounded real lemma for FM second model has been established. Using the bounded real lemma, the finite horizon and infinite horizon H_∞ filtering problems has been solved. In [15], using Riccati equation based approach the bounded real lemma for 2-D discrete systems have been established by Du and Xie. Using this concept, many researchers have investigated the H_∞ control of 2-D discrete systems [15, 18, 24, 30–35].

The H_∞ control analysis using state feedback techniques are extended to various 2-D discrete state delayed systems in [36–38].

A new improved version of bounded real lemma is derived in [33], which provides a flexibility in the scaling matrices and thus give less constraint based conditions on the H_∞ norm of the system as compared to the bounded real lemma of [18]. Using this improved bounded real lemma a 2-D deconvolution filter [33] under the performance specification for the FM second model is designed, with the aim to reconstruct the input signal given to the 2-D system in the presence of the measurement noise. The

result has been further extended for the design and computation procedure for 2-D deconvolution filter subjected to polytopic modeling uncertainties and quantization errors.

There has been a considerable attention on the model reduction for 2-D systems [39, 40] due to their extensive and important applications in 2-D signal and image processing. Modeling often leads to a high-order mathematical model, which is problematic for system analysis and synthesis. This initiated the development of model reduction techniques. Hankel norm approximation, balanced truncation techniques, pseudo-balanced, structurally balanced, quasi-balanced, and weighted balanced [40, 41] are some of the model reduction techniques. The basis of model reduction is to approximate the higher order system to its equivalent lower order model without significant errors. Using H_∞ model reduction, the objective is to find a low order model, such that the H_∞ norm of the difference between the high-order (Original) model and the reduced order model is small. The system is subjected to uncertain parameters that appear affinely in the model matrices. The filter design method employs a Lyapunov function and scaling matrices with affine dependence on the system uncertain parameters. State feedback H_∞ stabilization problem for 2-D switched systems as represented by the Roesser model has been carried out in [42]. Sufficient conditions for ensuring asymptotic stability and guarantee H_∞ disturbance attenuation for the system under consideration are derived via common Lyapunov function approach and multiple Lyapunov function approach.

H_∞ control based techniques is also used to design the output feedback controller. In general, formulation of H_∞ output feedback control may result in the formulation of bilinear matrix inequality (BMI) problem, which is non-convex. To handle BMI problems, sufficient LMI based conditions are usually preferred [43–45] due to its computational simplicity. For such LMI formulations, the implementation of slack variable technique is used in [31, 32, 35] and the change of variable methods is adopted in [30]. In [30], the authors have derived a bounded real lemma using Lyapunov-based approach [46], which ensures the stability of the dynamic output feedback controllers. Finally H_∞ control based robust stabilization for uncertain FM second model is done by computing a dynamic output feedback H_∞ controller. In [35] performance analysis of 2-D discrete systems as described by the Roesser model via 2-D output feedback controller is done using the bounded real lemma of [15].

In a H_∞ control problem, the goal is to design a control law such that the l_2 gain of the mapping from the exogenous input to the controlled output must not exceed the prescribed level of attenuation level, which is expressed by a positive scalar number. A new $l_2 - l_\infty$ based stability criterion has been proposed in [47] which ensures the asymptotic stability of the 2-D discrete systems as defined by the Roesser model under the presence of saturation arithmetic, when there is no external interferences. Further it ensures the reduction of the effect of disturbances to a desired level, when external interference occurs in this 2-D discrete digital filter as represented by the Roesser model. This suggested approach is also LMI solvable, hence it is computationally efficient.

3 Conclusions

The survey report on 2-D discrete systems stability analysis using H_∞ control techniques has been presented. In this report the detailed study has been done on the application of various H_∞ control techniques for the stability analysis of 2-D discrete systems. All the different approaches that has been carried out using H_∞ control techniques for the stability analysis of 2-D discrete systems has been presented in a compiled form in this paper.

References

1. Shaked, U., Theodor, Y.: H_∞-optimal estimation: a tutorial. In: 31st Conference on Decision Control, pp. 2278–2286 (1992)
2. Zames, G.: Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses. IEEE Trans. Automat. Contr. **26**(2), 301–320 (1981)
3. Xie, L., Soh, Y.C., de Souza, C.E.: Robust Kalman filtering for uncertain discrete-time systems. IEEE Trans. Automat. Control **39**, 1310–1314 (1994)
4. Glover, K.: All optimal Hankel-norm approximations of linear multivariable systems and their L_∞ error bounds. Int. J. Contr. **39**, 1115–1193 (1984)
5. Ball, J.A., Cohen, N.: Sensitivity minimization in an H_∞ norm: parametrization of all suboptimal solutions. Int. J. Contr. **46**, 785–816 (1987)
6. Boyd, S., Balakrishnan, V., Kabamba, P.: On computing the H_∞ norm of a transfer matrix. Math. Contr. Signals, Syst. (1998)
7. Doyle, J.C.: Lecture Notes in Advances in Multivariable Control: ONR/Honeywell Workshop. Minneapolis, MN (1984)
8. Zhou, K., Doyle, J.C., Keith, G.: Robust and Optimal Control. Prentice-Hall, NJ (1996)
9. Glover, K., Doyle, J.C.: A state space approach to H_∞ optimal control. Three Decad. Math. Syst. Theory **135**(1), 179–218 (1989)
10. Kabamba, P.T., Boyd, S.P.: On parametric H_∞ optimization. In: Proceedings of the 27th Conference on Decision and Control, Austin, Texas, pp. 1354–1355 (1988)
11. Francis, B.A., Helton, J.W., Zames, G.: H∞ optimal feedback controllers for linear multivariable systems. IEEE Trans. Automat. Contr. **29**(10), 888–900 (1984)
12. Zhou, K., Doyle, J.C., Glover, K.: Robust and Optimal Control. Prentice-Hall, NJ (1996)
13. Khargonekar, P.P., Rotea, M.A.: Mixed H_2/H_∞ control: a convex optimization approach. IEEE Trans. Automat. Contr. **36**(7), 824–837 (1991)
14. Khargonekar, P.P., Petersen, I.R., Zhou, K.: Robust stabilization of uncertain linear systems: quadratic stabilizability and H_∞ control theory. IEEE Trans. Automat. Contr. **35**(3), 356–361 (1990)
15. Du, C., Xie, L.: H_∞ Control and Filtering of Two-Dimensional Systems, vol. 278. Lecture Notes in Control and Information Sciences. Springer, Berlin (2002)
16. Galkowski, K.: State-Space Realisation of Linear 2-D Systems with Extensions to the General n D case, vol. 263. Lecture Notes in Control and Information Sciences. Springer, London (2001)
17. Sebek, M.: H_∞ problem of 2-D systems. In: Proceeding of the European Control Conference, Groningen, Netherlands, pp. 1476–1479 (1993)
18. Du, C., Xie, L., Soh, Y.C.: H_∞ filtering of 2-D discrete systems. IEEE Trans. Signal Process. **48**, 1760–1768 (2000)
19. Petersen, I.: Disturbance attenuation and H_∞ optimization: a design method based on the algebraic Riccati equation. IEEE Trans. Automat. Contr. **32**(5), 427–429 (1987)

20. Cao, J., Sivasamy, R., Rakkiyappan, R.: Sampled data H_∞ synchronization of chaotic Lur'e systems with Time delay. Circuits, Syst. Signal Process. **35**(3), 811–835 (2016)
21. Cao, J., Rakkiyappan, R., Maheswari, K., Chandrasekar, A.: Exponential H_∞ filtering analysis for discrete time switched neural networks with random delays using sojourn probabilities. Sci. China Technol. Sci. **59**(3), 387–402 (2016)
22. De Souza, C.E., Fu, M., Xie, L.: H_∞ analysis and synthesis of discrete-time systems with time-varying uncertainty. IEEE Trans. Automat. Control **38**, 459–462 (1993)
23. Dharani, S., Rakkiyappan, R., Cao, J.: Robust stochastic sampled data H_∞ control for a class of mechanical systems with uncertainties. ASME J. Dyn. Syst., Meas. Control. **137**(10), 1–14 (2015)
24. Du, C., Xie, L., Zhang, C.: H_∞ control and robust stabilization of two-dimensional systems in Roesser models. Automatica **37**, 205–211 (2001)
25. Farges, C., Peaucelle, D., Arzelier, D., Daafouz, J.: Robust H_∞ performance analysis and synthesis of linear polytopic discrete-time periodic systems via LMIs. Syst. Control. Lett. **56**, 159–166 (2007)
26. Francis, B.A., Doyle, J.C.: Linear control theory with an H_∞ optimality criterion. SIAM J. Control Optim. **25**(4), 815–844 (1987)
27. Gao, H., Lam, J., Xu, S., Wang, C.: Stabilization and H_∞ control of two dimensional Markovian jump systems. IMA J. Math. Control. Inf. **21**, 377–392 (2004)
28. Khargonekar, P.P., Petersen, I.R., Rotea, M.A.: H_∞ optimal control with state-feedback. IEEE Trans. Automat. Contr. **33**(8), 786–788 (1988)
29. Paszke, W., Galkowski, K., Rogers, E., Owens, D.H.: H_∞ and guaranteed cost control of discrete linear repetitive processes. Linear Algebr. Appl. **412**, 93–131 (2006)
30. Xie, L., Du, C., Soh, Y.C., Zhang, C.: H_∞ and robust control of 2-D systems in FM second model. Multidimens. Syst. Signal Process. **13**(3), 265–287 (2002)
31. Feng, Z.Y., Xu, L., Anazawa, Y.: Sufficient LMI conditions for H_∞ static output feedback control of 2-D systems. In: 11th International Conference on Control, Automation, Robotics and Vision. Singapore, pp. 57–60 (2010)
32. Feng, Z.Y., Xu, L., Wu, M., She, J.H.: H_∞ static output feedback control of 2-D discrete systems in FM second model. Asian J. Control **14**(6), 1505–1513 (2012)
33. Xie, L., Du, C., Zhang, C., Soh, Y.C.: H_∞ deconvolution filtering of 2-D digital systems. IEEE Trans. Signal Process. **50**, 2319–2331 (2002)
34. Xu, H., Zou, Y., Xu, S., Guo, L.: Robust H_∞ control for uncertain two-dimensional discrete systems described by the general model via output feedback controllers. Int. J. Contr. Automation Syst. **6**(5), 785–791 (2008)
35. Yang, R., Xie, L., Zhang, C.: H_2 and mixed H_2/H_∞ control of two-dimensional systems in Roesser model. Automatica **42**(9), 1507–1514 (2006)
36. Xu, J., Yu, L.: H_∞ control of 2-D discrete state delay systems. Int. J. Control., Autom. Syst. **4**, 516–523 (2006)
37. Xu, J., Yu, L.: H_∞ control for 2-D discrete state delayed systems in the second FM model. Acta Automatica Sinicia **34**, 809–813 (2008)
38. Xu, J., Yu, L.: Delay-dependent H_∞ control for 2-D discrete state delay systems in the second FM model. Multidim. Syst. Sig. Process. **20**, 333–349 (2009)
39. Lu, W.S., Luo, H., Antoniou, A.: Recent results on model reduction methods for 2-D discrete systems and Systems. In: ISCAS, pp. 348–351 (1996)
40. Luo, H., Lu, W.S., Antoniou, A.: A weighted balanced approximation for 2-D discrete systems and its application to model reduction. IEEE Trans. Circuits Syst. I Fundam. Theory Appl. **42**(8), 419–429 (1995)
41. Zhou, K., Aravena, J.L., Gu, G., Xiong, D.: 2-D model reduction by quasi-balanced truncation and singular perturbation. IEEE Trans. Circuits Syst. II Analog Digit. Signal Process. **41**(9), 593–602 (1994)
42. Duan, Z., Xiang, Z.: State feedback H_∞ control for discrete 2-D switched systems. J. Franklin Institute. **350**, 1513–1530 (2013)

43. Crusius, C.A.R., Trofino, A.: Sufficient LMI conditions for output feedback control problems. IEEE Trans. Automat. Control **44**, 1053–1057 (1999)
44. Lee, K.H., Lee, J.H., Kwon, W.H.: Sufficient LMI conditions for H_∞ output feedback stabilization of linear discrete-time systems. **51**(4), 675–680 (2006)
45. Yu, L., Gao, F.: Output feedback guaranteed cost control for uncertain discrete-time systems using linear matrix inequalities. J. Optim. Theory Appl., **113**, 621–634 (2002)
46. Lu, W.S.: On a Lyapunov approach to stability analysis of 2-D digital filters. IEEE Trans. Circuits Syst. I **41**, 665–669 (1994)
47. Ahn, C.K.: $l_2 - l_\infty$ Elimination of overflow oscillations in 2-D digital filters described by the Roesser model with external interference. IEEE Trans. Circuits Syst II. **60**(6), 361–365 (2013)
48. Doyle, J.C., Glover, K., Khargonekar, P.P., Francis, B.A.: State space solutions to standard H_2 and H_∞ control problems. IEEE Trans. Automat. Contr. **34**(8), 831–847 (1989)
49. Hassibi, B., Sayed, A.H., Kailath, T.: Indefinite Quadratic Estimation and Control a Unified Approach to H$_2$/H$_\infty$ Theories. Siam, Philadelphia (1999)
50. Francis, B.A.: A Course in H$_\infty$ Control Theory. Lecture Notes in Control and Information Sciences. Springer, London (1987)
51. Glover, K., Doyle, J.C.: State-space formulae for all stabilizing controllers that satisfy an H_∞-norm bound and relations to risk sensitivity. Syst. Control Lett. **11**(3), 167–172 (1988)
52. Doyle, J.C.: Guaranteed margins for LQG regulators. IEEE Trans. Automat. Control **26**(4), 756–757 (1978)
53. He, Y., Wu, M., She, J.H.: Improved bounded-real lemma representation and H_∞ control of systems with polytopic uncertainties. IEEE Trans. Circuits Syst. II **52**, 380–383 (2005)

An Efficient High-Speed CORDIC Algorithm Using Parallel-Prefix Adders (PPA)

Vutukuri Venkatesh, Balaji Yeswanth, Repala Akhil and Ravi Kumar Jatoth

Abstract Delay and Area ceases the actual potential of the modern gadgets. Although, human has sophisticated devices around him yet yearns to save time and space. So, this paper centers on the highly efficient CORDIC algorithm, known for its low-cost implementation in DSP algorithms. In an effort, to improve the algorithm further in terms of area and speed, comparative analysis has been done by replacing Ripple carry adder with Parallel-Prefix adders, namely, Brent-Kung adder, Han-Carlson adder and Kogge-Stone Adder. The algorithm was designed in VHDL using XILINX ISE 14.7 design suite and implemented in XILINX Spartan 6e FPGA. Obviously, Parallel-Prefix adders have shown improved performance.

Keywords CORDIC algorithm · Ripple carry adder · Carry look ahead adder · Kogge-Stone adder · Brent-Kung adder · Han-Carlson adder

1 Introduction

Modules like adders, multipliers, comparators are involved in building up a digital circuit. Researchers have been trying to improve the speed and performance of these devices. CORDIC (Coordinate Rotation Digital Computer) algorithm was initially developed by E. Volder in 1959. It was improvised by Walther for several other mathematical operations. The adding operations are done using Ripple carry adder.

V. Venkatesh (✉) · B. Yeswanth · R. Akhil
Department of ECE, National Institute of Technology, Andhra Pradesh, Tadepalligudem, AP, India
e-mail: venkynany62@gmail.com

B. Yeswanth
e-mail: yeshbalaji16@gmail.com

R. Akhil
e-mail: akhilrepala@gmail.com

R. K. Jatoth
Department of ECE, National Institute of Technology, Warangal, Warangal, Telangana, India
e-mail: ravikumar@nit.ac.in

As ripple carry adder, remains to be a time killer, using Parallel-prefix adders can be considered as the best alternative, enabling CORDIC algorithm to compute results at a faster rate, neglecting unnecessary wire logic delays.

2 CORDIC Algorithm for Trigonometric Functions

For computing sine and cosine values, CORDIC algorithm is operated in Rotation mode where the vector coordinates along with the angle to be rotated are mentioned as input and the coordinates of the vector after the rotation of the vector are obtained as output, where x_n and y_n represent the cosine and sine value respectively. CORDIC algorithm acts as the backbone for various transforms, which involves operations like convolutions, etc.

2.1 Mathematical Analysis of CORDIC Algorithm

The Givens rotation transform gives the equation for the 2-D rotation vector (Fig. 1).

$$x_{out} = x_{in} \cos \theta - y_{in} \sin \theta,$$

$$y_{out} = x_{in} \sin \theta - y_{in} \cos \theta,$$

where (x_{in}, y_{in}) and (x_{out}, y_{out}) are the coordinates of the vector (Fig. 2).

$$x_{i+1} = x_i - \sigma_i y_i \rho^{-i},$$
$$y_{i+1} = \sigma_i x_i \rho^{-i} + y_i,$$
$$z_{i+1} = z_i - \sigma_i \alpha_i.$$

where n = number of micro-rotations, α_i = elementary angle for iteration, these values are stored in the form of LOOK UP table.

Fig. 1 Representing vector rotation
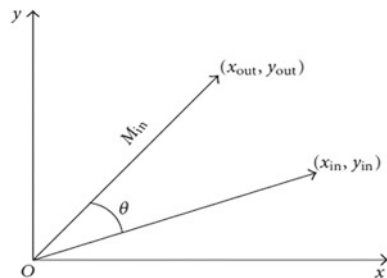
Fig. 2 Representing vector micro-rotations
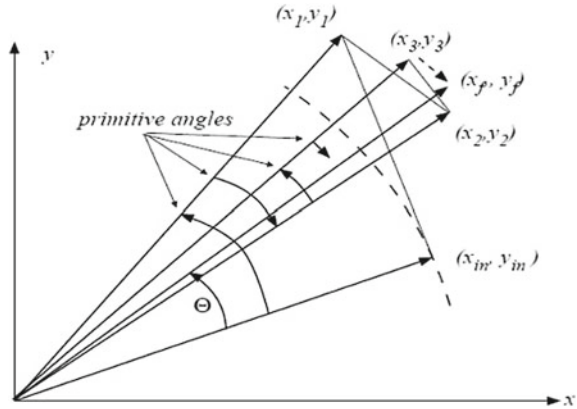
The input angle for rotation is given as angle accumulator, z_0. The σ (direction of micro-rotation) is calculated in every iteration, thus reducing the angle in z_0.

Vector coordinates after n micro-rotations are

$$x_n = K(x_{in} \cos \theta - y_{in} \sin \theta),$$
$$y_n = K(x_{in} \sin \theta - y_{in} \cos \theta),$$
$$z_n \to 0.$$

$$
\begin{aligned}
x_0 &= 1/K_m & x_n &= \cos \theta \\
y_0 &= 0 & y_n &= \cos \theta \\
z_0 &= \theta & z_n &= 0
\end{aligned}
$$

This Fig. 3 shows the equations for implementing trigonometric function where θ refers to the angle to be rotated.

Fig. 3 Look up table

i	β	tan(β)	
0	45.0°	1	$= 2^{-0}$
1	26.6°	0.5	$= 2^{-1}$
2	14.0°	0.25	$= 2^{-2}$
3	7.1°	0.125	$= 2^{-3}$
4	3.6°	0.0625	$= 2^{-4}$
5	1.8°	0.03125	$= 2^{-5}$
6	0.9°	0.015625	$= 2^{-6}$
7	0.4°	0.0078125	$= 2^{-7}$
8	0.2°	0.00390625	$= 2^{-8}$
9	0.1°	0.001953125	$= 2^{-9}$

Fig. 4 Diagram for n stages
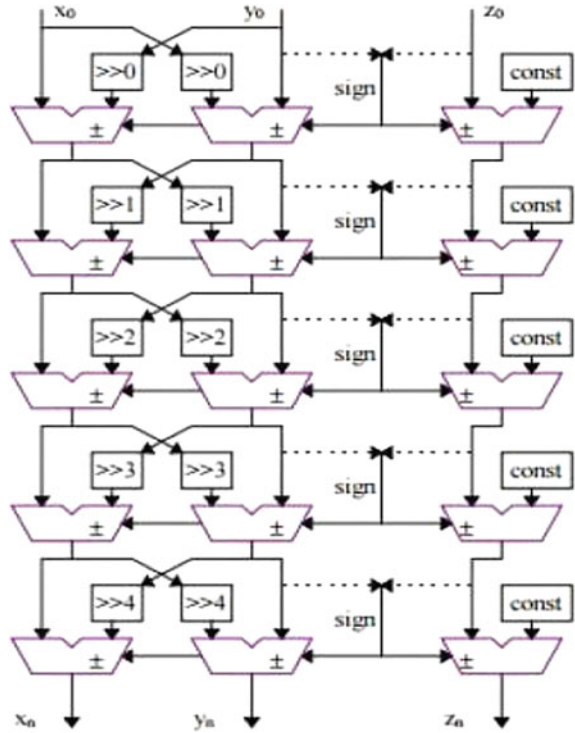
Figure. 4 is the pipeline architecture for implementing CORDIC trigonometric functions. There are n stages where each stage consists of an adder shifter and multiplexer and a small LUT table.

2.2 Optimization

As CORDIC algorithm involves adder as a main component, to optimize its working, usage of efficient adders is mandatory in lieu of normal ripple carry adder which causes delay to provide output.

3 Adders

Adders act as building block for implementing digital algorithms. Researchers have been trying to adopt new techniques to obtain faster results. Ripple carry adder consumes more time in generating output. Carry look ahead adder calculates result

faster compared to ripple carry adder but has a linear architecture. Parallel—prefix adders have logarithmic delays which signifies faster output.

Parallel-Prefix adders are tree-based adders which performs parallel addition. It uses prefix addition operator (o) for its operation.

$$(g_L, p_L) \, o \, (g_L, p_R) = (g_L + p_L \cdot g_R, p_L \cdot p_R)$$

3.1 Parallel-Prefix Adder Mechanism

1. Computation of carry generation and carry propagation signals.
2. Calculation of all carries in parallel.
3. Calculate the final sum and carry signals.
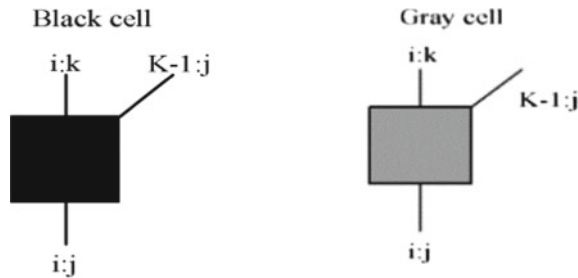
Here, the black cells reckon both generate and propagate signal. The gray cells merely computes generate signal which are required in computation of sum signal.

3.2 Kogge-Stone Adder

Kogge-Stone adder was modeled by Peter M. Kogge and Harold S. Stone. Kogge-Stone adder has minimal fanout and minimum logic depth. This adder has the best performance in VLSI implementations and produces carry signal in O ($\log_2 n$) time. Kogge-Stone adder has n [($\log_2 n$) − n + 1] computational nodes. As the intermediate stages do parallel advancement, carry is obtained in less time. Among all 16-bit adders, Kogge-Stone adder has less delay but at the cost of area and wire complexity at the intermediate stages. Figure 5 shows the Propagate-Generate network of Kogge-Stone adder. Buffer compensates the loading effect due to wires.

Fig. 5 16-bit Kogge-Stone adder
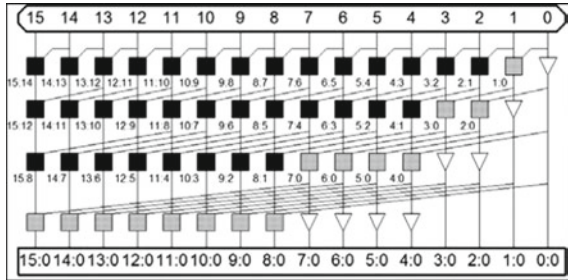
Fig. 6 16-bit Brent-Kung adder
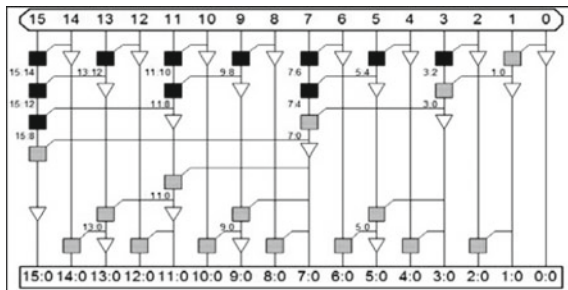
3.3 Brent-Kung Adder

Brent-Kung adder produces carry signal in $[2 * \log_2 n - 2]$ time which indicates that the delay increased when compared to Kogge-Stone adder due to its maximum logic depth. Brent-Kung adder has $[2n - 2 - \log_2 n]$ computation nodes. Brent-Kung adder occupies less area due to its less computation nodes. Brent-Kung adder requires $[2\log_2 n - 1]$ stages. Fanout of this adder is 2. To reduce the fanout buffers are used. The wire complexity of Brent-Kung adder is much less than Kogge-Stone adder. Figure 6 shows the Propagate-Generate network of Brent-Kung adder.

3.4 Han-Carlson Adder

Han-Carlson is the hybrid design of Brent-Kung adder and Kogge-Stone adder. Han-Carlson comprises of $[\log_2 n + 1]$ stages. For a 16-Bit Han-Carlson adder, there are 5 stages. Out of these five stages in prefix graph, the first stage is similar with Brent-Kung adder and the next three stages are similar with Kogge-Stone adder. The wire complexity is less compared to Kogge-Stone adder. Han-Carlson adder has $[(n\log_2 n)/2]$ computation nodes. Han-Carlson adder has minimum fanout but at the cost of increased logic depth. Figure 7 shows the Propagate-Generate network of Han-Carlson adder (Table 1).

Fig. 7 16-bit Han-Carlson adder
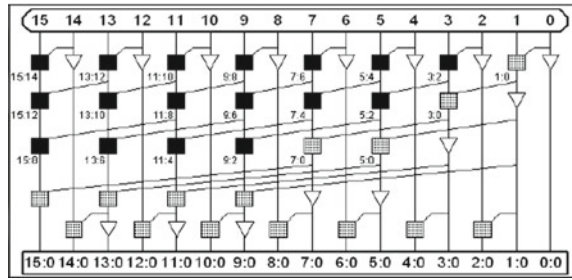
Table 1 The comparison of parallel prefix adders

Architecture	Logic levels	Max fanout	Cells	Features
Kogge-Stone	$\log_2 n$	2	N/2	High speed because of reduced number of logic levels. Power consumption and wire complexity is high
Han-Carlson	$\log_2 n + 1$	2	N/4	It has less number of cells and wire tracks as compared to Kogge-Stone at the cost of one extra logic level
Brent-Kung	$2\log_2 n - 1$	2	1	Less wire congestion and takes less area to implement and speed is less compared to Kogge-Stone adder

4 Implementation Results

ISE Xilinx Simulator tool was used to implement CORDIC Algorithm. Below figures represents the RTL schematic block and the Output simulation results of the 16-bit CORDIC trigonometric block.

The CORDIC block output in Fig. 8 signifies x_n and y_n where the signals represent cosine and sine functions respectively and z_n represents the error in the process of calculation which is to be finally zero. A clock input is given, as the shifter used in CORDIC block need to be synchronized with clock. In Fig. 9, x0 and y0 refers to the initial inputs. As y0 is given as zero and x0 is given with the value of one upon K such that x16 and y16 gives cosine and sine values respectively in 2QN format

Fig. 8 CORDIC block
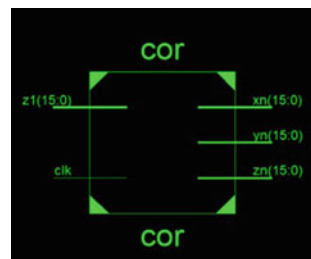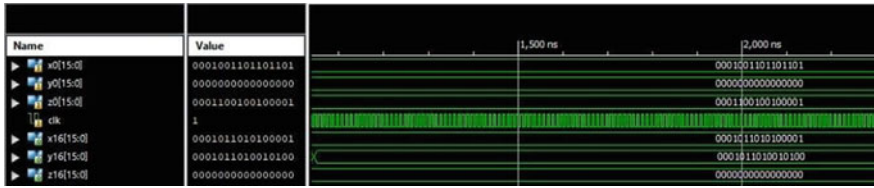
Fig. 9 Simulation results

depending on the z0 value which is the required angle of rotation in radians. The delay of CORDIC algorithm has been verified by considering the example of 45° where the 2QN format of z0 represents the 45° in radians. The values of x16 and y16 come out to be 0.707 which is as expected.

5   Results

Above results show that high speed CORDIC algorithm is obtained using Kogge-Stone adder which comes out to be 84.24 ns. While, Han-Carlson adder proves to be area efficient. Based on customer's priority, CORDIC algorithm must include either Kogge-Stone adder for speed or Han-Carlson for area (Tables 2 and 3).

Table 2   The propagation delay to obtain results by using different adders

CORDIC algorithm using different adders	Delay to obtain the result of trigonometric functions (ns)	Delay of 16-bit adders solely (ns)
Ripple carry adder	102.615	15.236
Carry look ahead adder	99.974	11.822
Kogge-Stone adder	84.24	9.4
Han-Carlson adder	96.438	10.916
Brent-Kung adder	95.214	15.905

Table 3   The area used in terms of LUT-FF pairs

CORDIC algorithm using different adders	Area used in terms of LUT
Ripple carry adder	156
Carry look ahead adder	141
Kogge-Stone adder	145
Han-Carlson adder	107
Brent-Kung adder	144

6 Conclusion

CORDIC Algorithm is applied in various areas like Multimedia for image processing, Communication for signal processing, Digital Electronics to design calculators, etc. In this paper, CORDIC algorithm for computing sine and cosine functions was successfully designed and implemented in Spartan 6e board. Moreover, to increase the efficiency of such useful algorithm, analysis was done by using different adders, among which Kogge-Stone adder proved to be efficient in terms of speed and Han-Carlson proved to be efficient in terms of area. As electronic gadgets have revolutionized the world, speed and area are very important factors which are to be still worked upon. Thus, CORDIC algorithm has mammoth importance making it an inevitable algorithm.

References

1. Anas, M.M., Padiyar, S., Boban, A.S.: Implementation of CORDIC algorithm and design of high speed CORDIC algorithm. In: International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)
2. Rani, G., Kumar, S.: Delay analysis of parallel prefix adders. Int. J. Sci. Res. (IJSR)
3. Weinberger, Smith, J.: A logic for high-speed addition. Natl. Bur. Stand., No. Circ. **591**, 3–12, 195
4. Basha, M.M., Ramanaiah, K.V., Reddy, P.R., Reddy, B.L.: An efficient model for design of 64-bit high speed parallel prefix VLSI adder. Int. J. Mod. Eng. Res. (IJMER) **3**(5) (2013)
5. Kogge, P., Stone, H.: A parallel algorithm for the efficient solutions of a general class of recurrence relations. IEEE Trans. Comput. C **22**(8) (1973)
6. Pullarao, B., Kumar, J.P.: Design of high speed based on parallel prefix adders using in FPGA. Int. J. Eng. Sci. & Res. Technol. (IJESRT) (2013)
7. Han-Carlson: Fast area-efficient vlsi adders, IEEE (1987)
8. Vishal, R., Naik, S.K.: Design of a carry tree adder. International Journal of Pure and Applied Research in Engineering and Technology (IJPRET) **2**(9), 413–424 (2014)
9. Lakshmi, B., Dhar, A.S.: CORDIC architectures: a survey
10. Hoe, D.H.K., Martinez, C., Vundavalli, S.J.: Design and characterization of parallel prefix adders using FPGAs

Stockwell Transform Based Time-Frequency Analysis of the ECG Features for Assessment of Risk of Left Ventricular Hypertrophy in Hypertension Patients

Raghuvendra Pratap Tripathi, Ankita Tiwari, Sristi Jha, Rohini Srivastava, Nitin Sahai, Sudip Paul, Basant Kumar, T. K. Sinha and Dinesh Bhatia

Abstract Hypertension is a major disease that affects millions of people world-wide. If hypertension remains untreated for long time it could give rise to an enlarged heart condition called Left Ventricular Hypertrophy (LVH). However, recent research in medical domain proposes that left ventricular diastolic dysfunction (LVDD) may be considered as a preceding indicator of LVH. The analysis of ECG signal has shown significant results in identifying LVDD condition that can lead to LVH, however the analysis method till date have remain limited to the manual examination of time domain features of ECG parameters by the experts which is time consuming

R. P. Tripathi · A. Tiwari · S. Jha · N. Sahai · S. Paul · T. K. Sinha · D. Bhatia (✉)
Department of Biomedical Engineering, School of Technology, North Eastern Hill University,
793022 Shillong, Meghalaya, India
e-mail: bhatiadinesh@rediffmail.com

R. P. Tripathi
e-mail: raghvendra5936@gmail.com

A. Tiwari
e-mail: ankitaiit15@gmail.com

S. Jha
e-mail: sristi.jha91@gmail.com

N. Sahai
e-mail: nitinbiomedical@gmail.com

S. Paul
e-mail: sudip.paul.bhu@gmail.com

T. K. Sinha
e-mail: tksinha001@gmail.com

R. Srivastava · B. Kumar
Department of Biomedical Engineering, School of Technology, North Eastern Hill University,
211004 Prayagraj, U.P., India
e-mail: srivastava.rohini14@gmail.com

B. Kumar
e-mail: singhbasant@mnnit.ac.in

and cumbersome. Since the time-frequency analysis of the ECG signal have shown more promising results in diagnosis of any abnormality related to the cardiac system, therefore the application of this method is employed in the patients suffering from hypertension and for assessment of the future risk(s) in developing LVH. In the study, we have proposed a Stockwell Transform based time-frequency analysis method of the ECG features (QRS Complex, P-Wave and T-Wave) for accessing the preceding stage of the myocardial remodeling phase. To perform the study, ECG features of the 60 subjects recorded from hospital comprising of 30 controlled and 30 hypertension cases were studied. At the location of the QRS interval a spreading in the power is observed in diseased patients, which signifies increased ventricular activation time, also the power level of the P-Wave and T-Wave have shown significant changes. Increased Ventricular activation time and P-Wave dispersion observed in the frequency domain along with P-wave terminal force, can be used as an indicator of associated risk of developing Left Ventricular Hypertrophy.

Keywords Left ventricular diastolic dysfunction (LVDD) · Left ventricular hypertrophy (LVH) · Stockwell transform · Time-Frequency analysis

1 Introduction

Hypertension is a critical health issue that needs a quick attention. The long term Hypertension condition can lead to the several cardio-vascular diseases (CVDs) or chronic kidney disease. However the management of hypertension condition is quite easy if it is identified at an early stage, this is the reason due to which the condition is considered to be as a leading avoidable risk factor for premature deaths. A study from American Heart Journals estimated that in the year 2000, 26.4% of the total adult population around the world was affected from hypertension condition [1]. Some other reports from national healthcare agencies suggest that the number is much higher and continuously increasing in the low- or middle-income countries. Also the prolonged hypertension condition is considered to be as a high risk condition for many cardio-vascular diseases (CVDs) such as Stroke, Coronary Artery disease Congestive Heart Failure, Left Ventricular Hypertrophy, Atrial Fibrillation, etc. [2]. These facts motivate researchers to investigate and develop reliable, non-invasive and low cost methods for diagnosing hypertension or associated diseases.

The heart muscles of the patients suffering from hypertension needs to do harder effort as compared to the normal subjects. Most of the harder efforts in hypertension condition are performed by the left ventricular muscles, due to the increased efforts and workload the tissues in the ventricular muscles of Heart chamber wall gets thicken, in some cases the enlargement in the heart chambers is also a common effect. These thickened and enlarged heart muscles are having less elasticity as compared to the normal heart muscles and may be fail to pump the blood from required amount of force [3].

The above stated remodeling of the left ventricular muscles is known as the Left Ventricular Hypertrophy (LVH). However, there many other reasons such as Aortic Valve Stenosis, can also cause the LVH [4].

Several investigations in the medical domain suggest that the analysis of the ECG signals in hypertension patients can give us a reliable way to identify the preceding stage of the above stated remodeling of the cardiac muscles. The left Ventricular Diastolic Dysfunction (LVDD) is the earliest possible condition that can be detected and leads the LVH. Till date the well-known methods of detecting LVH condition at an early stage are several ECG criteria such as Cornell voltage product (50% sensitivity), Sokolow-Lyon (29% sensitivity), Romhilt-Estes (22% sensitivity) ECG criteria etc. The poor sensitivity of the above discussed methods limits the accurate diagnosis of the condition [5, 6, 7].

Recently Usama et al. (2015) has presented a study that identifies the early changes in the Electrocardiogram signals of hypertension patients. The major finding in the study were long QR interval as an measure for increased Ventricular Activation Time, Dispersed P-Wave and increased P-Wave terminal force. The authors reported that the above ECG features can be used as early markers in the evaluation of the hypertension condition [8].

The authors of this communication has extended the study performed by Usama et al. and presented a Stockwell Transform based Time-Frequency Method to studying the ECG features in hypertension patients. The ECG signals of the 60 subjects including 30 controlled and 30 diseased are studied and the observation is presented in the paper. Since the ECG features have shown the significant changes in hypertension condition, therefore can be used for diagnosis of the associated condition. The assessment of the associated risk of LVH is estimated based on the study presented by Usama et al. and a more improved visualization of the early changes detected by the authors are presented by Spectrogram representation of the signals [8, 9].

The Fast Discrete Stockwell Transform (FDST) based time-frequency analysis is a new approach and provides higher time and frequency resolution compared with Short-Time Fourier Transform (STFT) and Wavelet Transform (WT) [10, 11, 12, 13]. We applied the FDST to obtain the spectrogram of the ECG signals, and based on the spectrogram we quantitatively calculated the distribution of the power within the frequency ranges of the QRS Interval, P-wave and T-wave in the Hypertension and controlled patients. We have also presented the time domain changes in the ECG features to evaluate the risk of the LVH.

The advantage of the study is that the FDST based time-frequency analysis of the ECG features provides a higher degree of resolution as compared to the existing wavelet and STFT methods [11], and also the quantitative measurement of the time-frequency power contained in the ECG features provides a more reliable information about the cardiac cycle activities [14].
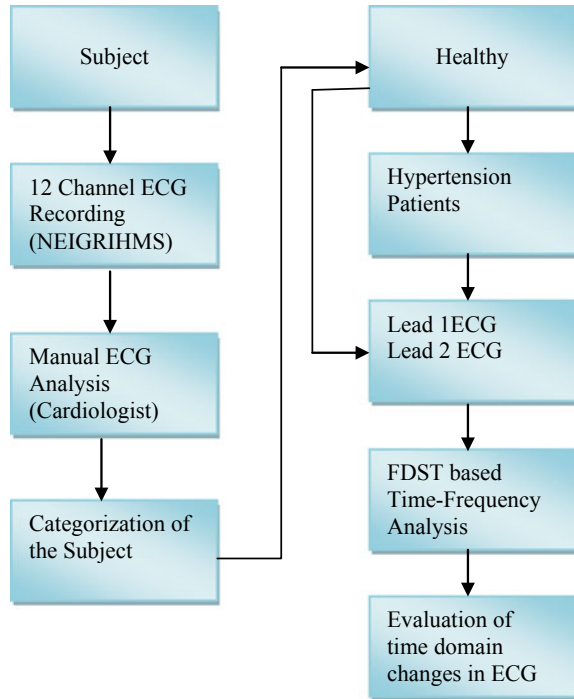
2 Methodology

The steps used in the study are presented with the help of the following flow diagram shown in the Fig. 1. The study starts from the reporting of the subject at the OPD of the North Eastern Indira Gandhi Regional Institute of Medical and Health Sciences (NEIGRIHMS). After the physical examination and Blood Pressure check-up, the patient is directed to the ECG room of the Hospital and a 12-Channel ECG is recorded. A cardiologist evaluates the condition of the patient based on ECG recording and physical examination results. The subjects falling in the Hypertension and Controlled category are considered for the further Evaluation. The characteristics of the subjects participated in the study is presented in the Table 1. To establish a relationship in

Table 1 Characteristics of the subjects included in the

Subject characteristics	Controlled	Hypertension	Total/Avg.
Number of subjects	30	30	60
Age (Year)	48.64 ± 7.83	53.64 ± 11.93	–
Sex (M/F)	19/11	23/7	42/18
H.R. (AVG)	76.38 ± 9.26	71.2 ± 11.95	–

ECG parameters and risk of LVH in Hypertension patients, three types of studies on of ECG parameters are carried out, these are FDST-based spectrogram, histogram analysis of ECG features and Boxplot analysis of ECG features.

The statistical features analysis is done to confirm the early changes in the ECG signals of hypertension patients. Based on the results obtained in the statistical analysis it is concluded that the hypertension patients whose ECG features are varying as per the study presented by Usama et al. are considered to be at risk of LVDD that may proceed into LVH [8] which is a major factor for cardiac diseases [15].

The ECG signals of the hypertension patients satisfying the statistical changes as suggested by Usama et al., are studied using FDST-based spectrogram analysis. The comparison between the spectrogram representation of controlled subjects and Hypertension subjects is made [8, 16].

2.1 Data Acquisition System for Recording of Lead 1 and Lead 2 ECG

To record the ECG signals of the subjects categorized in the above mention two categories we used Equivital EQ02 Life Monitor. The EQ02 device is shown in EQ02 records 2 channel of ECG data by wearing a chest strap as shown in Fig. 2. EQ02 device records ECG signals at 256 Hz with a high resolution of 16 bit. The recording analysis of the ECG signals is done by interfacing the EQ02 to the Labchart software via Bluetooth connectivity [17].

2.2 Labchart-Based ECG Analysis

The ECG signals recorded using the Equivital device is studied using Labchart software. Labchart Software is a sophisticated signal analysis software developed by the

Fig. 2 Equivital EQ02 life monitor used for 2 channels ECG recording and placing of strap sensor [23]

AD Instruments and has inbuilt advanced analysis methods. The Labchart software provides us an easy way of extracting the ECG features which includes RR intervals, heart rate, PR interval, P duration, QRS interval, QT, QTc, T peak, T end, JT interval. Various ECG components such as QRS interval, T-Wave, R-amplitude etc. are extracted from the ECG recording using inbuilt tools incorporated in the software. The total 10 ECG features are extracted from the recorded ECG lead data. After recording and analysis of the ECG signals the text files of the raw ECG data and extracted feature are exported to the PC for further studies [18].

2.3   *Statistical Analysis of ECG Features*

Out of the 10 ECG features extracted from the ECG recordings using Labchart and exported as text file are imported into the MATLAB software to perform the statistical analysis. Two types of statistical representation, Boxplot and Histogram are used for comparing the ECG features.

In Histogram representation of the ECG features, the X-axis is divided into 10 bins, which separates the whole distribution of ECG Features into 10 distinct intervals. On Y-axis the instances of the features are plotted. The instance provides the information regarding the number of patients diagnosed with the particular range of an ECG features. In this way we can represent that how the distribution of a particular ECG feature has been occurred [19].

The second statistical representation of the ECG features is done using boxplot methods. The boxplot method is seemed to be very useful approach in comparing the of the ECG features of different classes. Using the Boxplot representation we convey a lot information about changes in the ECG features of the controlled subjects as compared to the hypertension patients. Basically the four important statistical quantities median, variability, shape of the distribution and information of the outliers related to ECG features are obtained using boxplot [20]. Based on the changes in the above four quantities of the ECG features along with observations from histogram analysis the estimation of the cardiac remodeling phase is done.

Since the QRS complex in ECG signals represent the electrical activity of the ventricular activation therefore the wider QRS complex with normal sinus rhythm may be related to the increased ventricular activation time. The dispersion of the P-Wave and tall T-Wave are the other ECG features that are used for estimating the risk of LVH. As per the Medical domain the disperse P-Wave is a symptom of increased Terminal force when appears along with Tall T-Wave [8].

2.4   *FDST-Based Time-Frequency Analysis*

Stockwell Transform is a novel method for time-frequency analysis of the biosignals and it was proposed by the Stockwell et al. [13]. The benefits of the representation

were the improved time-frequency resolution as compared to the existing methods like wavelet and STFT. The basis of the Stockwell transform is the wavelets transform only, however use of a unique frequency dependent wavelet makes the resolution dependent upon the frequency of the signal to be transformed. Therefore it results in a more improved resolution as compared to the wavelet where the wavelet is independent of the frequency of the signal. In the earlier years it was avoided in use due to the computational cost required in calculating the transform of the signal. But the recent years a Fast Discrete Stockwell Transform (FDST) algorithm have been developed to decrease the computational cost, still the use of the transform was to the applications like image processing, audio processing, etc. In 2008 Robert A. Brown in their communication encouraged researchers to use the FDST in biomedical signal processing due to the associated advantages [2].

Authors of communication used the FDST to obtain the spectrogram of the ECG signals in controlled and hypertension patients. The spectrogram is obtained as a unique representation in each of the above mentioned case. For the patients those were at risk LVH, as per the statistical analysis of the ECG features suggested by Usama et al. a dispersed spectrogram is obtained. The distribution of the power in QRS Interval was quantitative much wider as compared to the controlled subjects; also an increased power level at location of P-Wave and T-Wave indicates the increased P-Wave terminal force that may cause the remodeling of the Cardiac muscles or LVDD. The condition progresses and converts in a disease called LVH. The unique spectrogram representation in controlled and hypertension condition with increased terminal force and QRS interval may be used for screening of the patients at risk of cardiac remodeling phase [8] or developing LVH.

3   Result and Discussion

The analysis of the ECG features is done to estimate the variation in several factors such as increased ventricular estimation time, P-Wave terminal force, tall T-Wave etc. To compare the ventricular activation time of hypertension patients with controlled subject analysis of QRS interval is performed.

Figure 3 shows the boxplot of QRS interval for hypertension and controlled subjects. The plot is drawn using the average QRS interval extracted from approximately 2-min long ECG recording of each patient. The analysis of the figure indicated that the QRS interval in case of hypertension condition is appeared to be more random and wider as compared to the controlled subjects. However the QRS interval of more than 25% hypertension patients is greater than that of controlled subjects. The median value of the QRS interval of hypertension is also higher than the QRS interval of controlled subjects. The observation from study indicates that an increase in the QRS interval is observed for hypertension patients. To have a much clear picture about the variation of QRS interval, the histogram representation of the QRS interval is done. The resulting histogram plots are shown in Fig. 4, in case of controlled the maximum value of QRS interval is appeared to be 0.095 s, however in hypertension condition 8
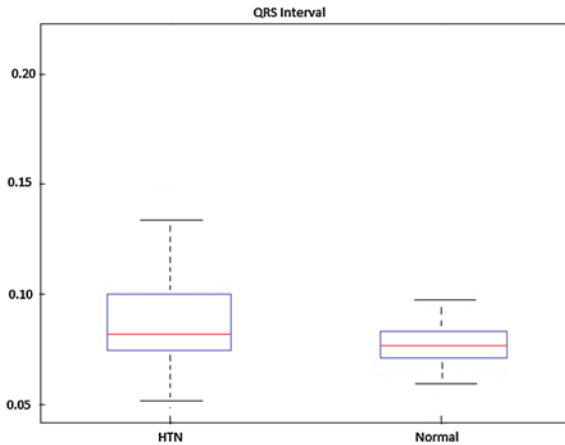
Fig. 3 Comparison of QRS interval to estimate the variation in ventricular activation time in hypertension and controlled subjects
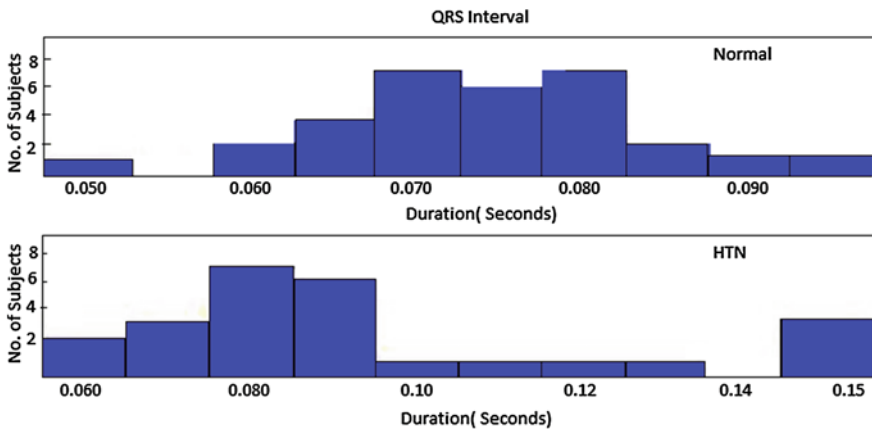
Fig. 4 Analysis of QRS interval in hypertension and controlled subjects using histogram plot

out of 30 patients were diagnosed with much wider QRS interval. The histogram plot shown in Fig. 4, shows that in controlled case 26 patents were having the QRS interval less than or equal to the 0.08 s where as in case of hypertension patients 23 patients were having the interval greater than or equal 0.08 s. The observation from study is that the QRS complex in hypertension patients was much higher or closer to the higher value, also 4 out of these 30 patients hypertension patients were having QRS interval of 0.15 s, as higher QRS interval is correspond to the increased ventricular activation time [21], hence the remodeling of the cardiac muscles is suspected in these 4 patients. Further analysis of P-Wave and T-Wave is done to obtain more information about the cardiac events.

Fig. 5 Comparison of
P-Wave duration in
controlled and hyper-tension
patients
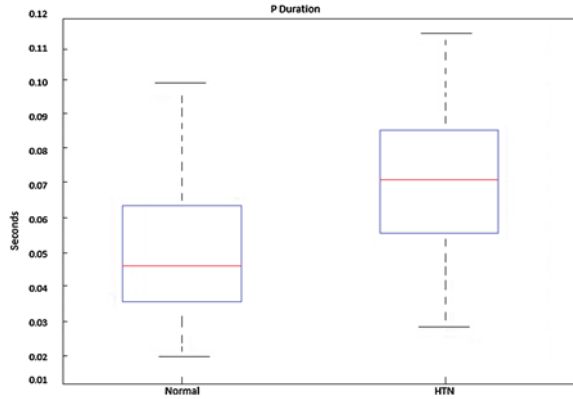
Figure 5 shows the boxplot of P-Wave duration of controlled and hypertension patents. The P-Wave is seemed to be more disperse in Hypertension condition as more than 50% patients of hypertension were having the wider P-Wave. Analysis of the histogram plot showing the distribution P-Wave duration in subjects indicate that, in case of 30 controlled subjects only 3 subjects were having the P-Wave duration greater than 0.07 s whereas in case of hypertension 13 patients were diagnosed with P-Wave duration greater than 0.07 s. Also the distribution is skewed toward the higher P-Wave duration in hypertension patients. The combined analysis of the both figures for P-Wave duration signifies that the hypertension patients are having the dispersion in P-Wave, since the dispersion in P-Wave is related to the increased P-Wave terminal force that means the excretion of extra force over heart muscles. The increased force over the heart muscles may lead to the remodeling of the muscles or dysfunctionality of the heart [8] (Fig. 6).

Another ECG feature (Figs. 7 and 8) that is evaluated for screening the patients at risk of developing LVH is T-Wave. In process of supplying blood to various body
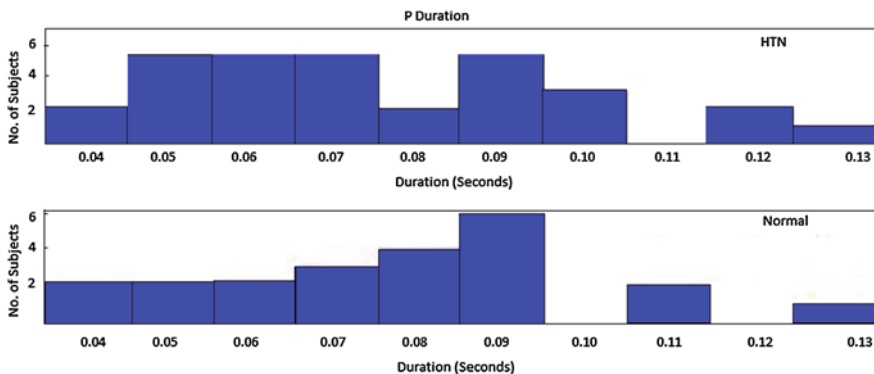
Fig. 6 Analysis of P-Wave duration in hypertension and controlled subjects using histogram plot
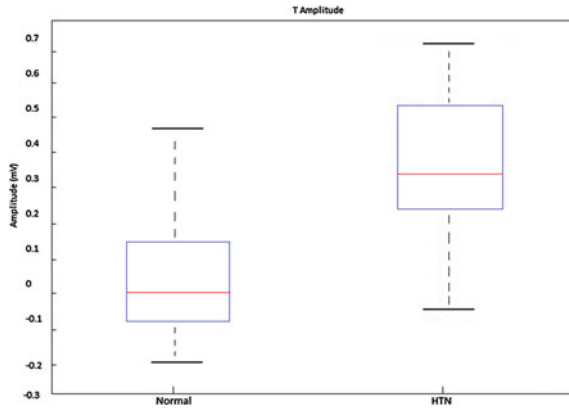
Fig. 7 Comparison of T-Wave amplitude in controlled and hypertension patients
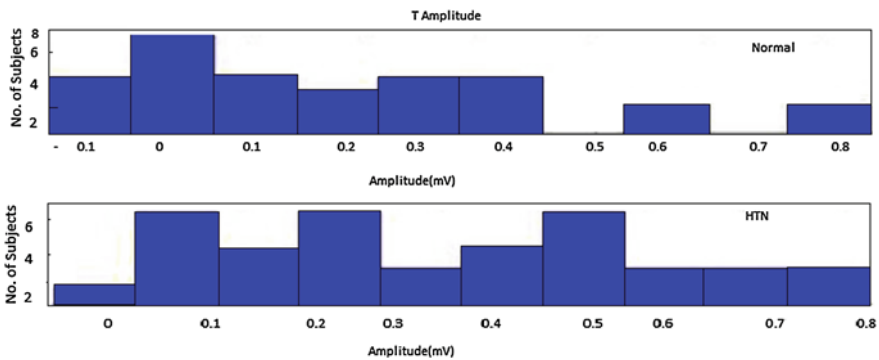
Fig. 8 Analysis of T-Wave amplitude in hypertension and controlled subjects using histogram plot

organs tallness in T-Wave amplitude is taken as a measure of extra effort put by the ventricular muscles, more than 25% of hypertension patents are diagnosed with T-Wave amplitude greater than that of maximum of controlled subjects, the histogram representation of T-Wave amplitude shows that 27 out 30 controlled subjects are having the T-Wave amplitude less than or equal to 0.3 mV whereas in case of hypertension patients 11 out of 30 patients are diagnosed with T-amplitude greater than 0. mV, this signifies that in the T-Wave is appeared as taller in hypertension condition. The increase in T-Wave amplitude is considered as an effect of extra effort put by the ventricular muscles. This extra effort could make the heart muscles harder and may lead to the enlargement of the heart and develop the LVH [22].

The study signifies that the variation in the ECG features can be useful in diagnosing the stage of remodeling of cardiac muscles and increased muscle effort in supplying the blood to body organs and the variation in ECG is manually calculated by trained professionals via analyzing the time domain analysis of ECG signals. To
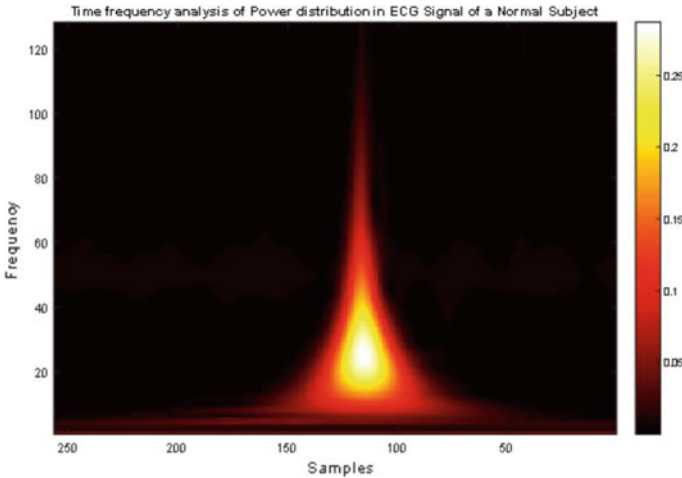
Fig. 9 FDST based analysis of ECG signals and resulting spectrogram for controlled subjects

make the analysis easy and present a better visualization of the variation in ECG features a FDST based spectrogram method is proposed.

Figure 9 shows the spectrogram for a controlled subject, most of the power in the signal is cantered at the location of QRS, spectrogram shows that the distribution of the power in the controlled subject is narrower in time domain and wider in frequency domain that means the higher frequency (20–60 Hz) are present in case of controlled subjects, the presence of higher frequencies making the spectrogram to appear as sharp peak. Another observation from spectrogram is strength of ECG signal at the location of P-Wave and T-Wave, the strength of ECG signals in controlled condition is quite low as compared to the hypertension patients. FDST based spectrogram representation for ECG signals of a hypertension patient screened for risk of LVH using the statistical analysis is depicted in Fig. 10. The spectrogram is appeared to be quite disperse in time domain and also having the presence of lower frequencies typically in the range of 2–20 Hz, also at location of P-Wave and T-Wave the strength of the signal is little bit higher as compared to that of controlled subjects. The quantitive analysis of the spectrogram for 30 controlled and 30 hypertension patients is presented in Table 2 provided below, the 30 hypertension patients are further analyzed into two sub groups of 26 patients without any indication of remodeling of cardiac muscles and 4 subjects screened with indication of remodeling phase of cardiac muscles. The patients who have shown the changes in ECG feature as explained by Usama et al. are suspected for developing the LVH. The spectrogram of both the conditions is depicted in Figs. 9 and 10 respectively, the spectrogram for each condition are unique and easily distinguishable there they have capability to be used for assessing the risk of developing LVH in hypertension condition. The results presented in the paper are only primary and detailed investigation is required before adapting.
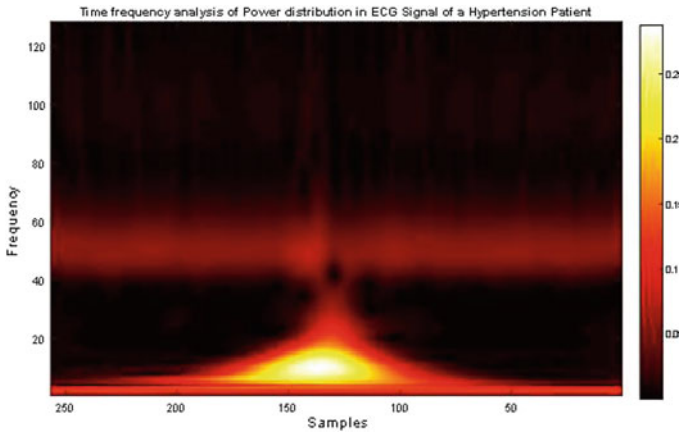
Fig. 10 FDST based time-frequency analysis of ECG signals and resulting spectrogram for controlled subjects

Table 2 Results from quantitive analysis of ECG features using FDST based spectrogram

Category analysis	Controlled	HTN	HTN (Risked of LVH)
QRS (frequency range Hz)	0–60	0–35	0–20
QRS interval (ms)	55–95	60–120	120–150
P-Wave duration (ms)	40–110	30–90	90–120
T-Wave amplitude (mV)	−0.1–0.6	−0.2–0.6	0.4–0.6

4 Conclusion

The article presents the analysis of ECG signals and compares the variation in ECG features. Based on the study presented by other researchers variation in ECG features are correlated with the remodeling of the cardiac muscles, the patients whose ECG signal changes indicate toward the changes in their cardiac muscles are identified as patients with risk of developing LVH. The Time-Frequency analysis of the ECG features is done using FDST approach and quantitive results for 60 patients are presented. The spectrogram representation using FDST approach is proposed for identifying the patient at risk of developing LVH. The unique and distinguishable spectrogram for are obtained for each different condition. The divergent spectrograms may be used for screening the patients at risk of developing LVH. The primary results using the method are promising in identifying the LVH, however further analysis on the proposed method is still required before adaption. Upon successful adaption the

method can be used as a primary diagnostic tool for assessing the risk of LVH in hypertension patients.

Conflict of Interest There is no conflict of interest associated with publication of this research article.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Ethics Committee (IEC) of the North Eastern Hill University and North Eastern Indira Gandhi Regional Institute of Health and Medical Sciences, Shillong Meghalaya.

Informed Consent Written consent was informed and obtained from all individual participants included in the study.

References

1. Sowers, James R., Epstein, Murray, Frohlich, Edward D.: Diabetes, hypertension, and cardiovascular disease: an update. Hypertension **37**(4), 1053–1059 (2001)
2. Chockalingam, A., Campbell, N.R., Fodor, J.G.: Worldwide epidemic of hypertension. Can. J. Cardiol. **22**(7), 553–555 (2006)
3. Levy, Daniel: Left ventricular hypertrophy. Drugs **35**(5), 1–5 (1988)
4. Lorell, B.H., Carabello, B.A.: Left ventricular hypertrophy: pathogenesis, detection, and prognosis. Circulation **102**(4), 470–479 (2000)
5. Ishikawa, J., Ishikawa, S., Kabutoya, T., Gotoh, T., Kayaba, K., Schwartz, J.E., Kario, K.: Cornell product left ventricular hypertrophy in electrocardiogram and the risk of stroke in a general population. Hypertension **53**(1), 28–34 (2009)
6. Schröder, J., Nuding, S., Müller-Werdan, U., Werdan, K., Kluttig, A., Russ, M., Medenwald, D.: Performance of Sokolow-Lyon index in detection of echocardiographically diagnosed left ventricular hypertrophy in a normal Eastern German population-results of the CARLA study. BMC Cardiovasc. Disord. **15**(1), 69 (2015)
7. Cabezas, M., Comellas, A., Ramón, J.G., López, L.G., Casal, H., Carrillo, N., Castillo, R.: Comparison of the sensitivity and specificity of the electrocardiography criteria for left ventricular hypertrophy according to the methods of Romhilt-Estes, Sokolow-Lyon, Cornell and Rodríguez Padial. Rev. Espanola Cardiol. **50**(1), 31–35 (1997)
8. Boles, U., Enriquez, A., Ghabra, W.A., Abdollah, H., Michael, K.A.: Early changes on the electrocardiogram in hypertension E-J. Cardiol. Pract. **13**(30) (2015)
9. Tsutsumi, T., Okamoto, Y., Kubota-Takano, N., Wakatsuki, D., Suzuki, H., Sezaki, K., Nakajima, T.: Time–frequency analysis of the QRS complex in patients with is-chemic cardiomyopathy and myocardial infarction. IJC Hear. Vessel. **4**, 177–187 (2014)
10. Biswal, B.: ECG signal analysis using modified S-transform. Healthc. Technol. Lett. **4**(2), 68 (2017)
11. Saritha, C., Sukanya, V., Murthy, Y.N.: ECG signal analysis using wavelet transforms. Bulg. J. Phys. **35**(1), 68–77 (2008)
12. Odinaka, I., Lai, P.H., Kaplan, A.D., O'Sullivan, J.A., Sirevaag, E.J., Kristjansson, S.D., Rohrbaugh, J.W.: ECG biometrics: a robust short-time frequency analysis. In: 2010 IEEE International Workshop on Information Forensics and Security (Wifs), pp 1–6. IEEE (2010)
13. Boashash, B.: Time-Frequency Signal Analysis and Processing: A Comprehensive Reference. Academic Press (2015)

14. Khadra, L., Al-Fahoum, A.S., Binajjaj, S.: A quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques. IEEE Trans. Biomed. Eng. **52**(11), 1840–1845 (2005)
15. Gosse, P.: Left ventricular hypertrophy as a predictor of cardiovascular risk. J. Hypertens. **23**, S27–S33 (2005)
16. Wang, Y., Orchard, J.: Fast discrete orthonormal Stockwell transform. SIAM J. Sci. Comput. **31**(5), 4000–4012 (2009)
17. Neubert, S., Arndt, D., Thurow, K., Stoll, R.: Mobile real-time data acquisition system for application in preventive medicine. Telemed. E-Health **16**(4), 504–509 (2010)
18. Bolkhovsky, J.B., Scully, C.G., Chon, K.H.: Statistical analysis of heart rate and heart rate variability monitoring through the use of smart phone cameras. In: 2012 Annual International Conference of the IEEE, Engineering in Medicine and Biology Society (EMBC), pp. 1610–1613. IEEE (2012)
19. Johnson, P., Stemple, C.: Math, art and technology, a cross roads. In: Society for Information Technology and Teacher Education International Conference, pp. 2921–2924. Association for the Advancement of Computing in Education (AACE) (2005)
20. Ieva, F., Paganoni, A.M.: Depth measures for multivariate functional da-ta. Commun. Stat. Theory Methods **42**(7), 1265–1276 (2013)
21. Gosse, P., Jan, E., Coulon, P., Cremer, A., Papaioannou, G., Yeim, S.: ECG detection of left ventricular hypertrophy: the simpler, the better? J. Hypertens. **30**(5), 990–996 (2012)
22. Boles, U., Almuntaser, I., Brown, A., Murphy, R.R., Mahmud, A., Feely, J.: Ventricular activation time as a marker for diastolic dysfunction in early hyperten-sion. Am. J. Hypertens. **23**(7), 781–785 (2010)
23. Equivital, EQ02.: LifeMonitor Sensor Electronic Module Datasheet (2012)

Secure Image Restoration and Contrast Enhancement Using Wavelet Transform and Thresholding Technique

Kumari Suniti Singh, Yogesh Kumar Mishra and Harsh Vikram Singh

Abstract The paper represents a methodology which involves the use of Satellite images taken from Satellite Imaging Corporation and Google Earth. This technique proposes a satellite image enhancement system in two steps which consists of restoration and contrast enhancement. The work represents a methodology for enhancement of Satellite images using wavelet transform such as Discrete wavelet transform (DWT) and Stationary wavelet transform (SWT), thresholding and contrast enhancement using equalization based on absolute difference. The result of the evaluation indicates that the technique obtains good PSNR value of satellite images. The results also strengthen that the technique fares well in preserving the information content of the satellite images and also improve the visual quality of images.

Keywords Satellite images · DWT · SWT · DWT with SWT · Thresholding and equalization

1 Introduction

Satellite images play a significant role in providing geographical information. Satellite image sensing provides quantitative and qualitative information which is used for reducing the complexity of field work and time of study and analysis. Satellite remote sensing agencies such as NASA, ISRO, etc., collects data and images at regular intervals which are used in various applications. The amount of data received at data centres is very huge and it is increasing rapidly as the technology is changing at high speed and timely. Also the volume of data has been growing rapidly. So we require successful and efficient techniques to extricate the data and provide information of satellite pictures taken from monstrous satellites.

Satellite image enhancement procedures mostly used to remove the effects which are caused by catching devices and to get the quality upgraded image for additional utilization. Satellite pictures are captured from a very long distance, so they contain

K. S. Singh (✉) · Y. K. Mishra · H. V. Singh
Department of Electronics Engineering, KNIT Sultanpur, Sultanpur, India
e-mail: sunitisingh1990@gmail.com

too much noise and distortion because of barometric boundaries. After capturing the image, some radiometric and geometric amendments are completed on it, yet they are not sufficient for one of the applications. It is very important to improve the restored image before using it [1]. The most common issues occur in satellite images need to be addressed such as noise, artifacts, grey scale/colour image, resolution, distortion, high frequency content, etc.

The quality of satellite images can be acquired by increasing the contrast related content of images. Resolution enhancement is one of the techniques to conquer the drawbacks of satellite images. The change of satellite images might be done in both spatial domain and in frequency domain. In frequency domain enhancement is done using Fourier transform and where as in spatial domain it specifically operates on the pixels of the image. Utilization of both methods, i.e., spatial and frequency domain may utilize the benefits of both methods.

To improve extensive number of details of an image a band adaptive contrast modification is used and also amplification of noise has been done by using various filters. Transform domain methods are observed to be the best because in this appropriate transform can be determined that can be utilized for the better improvement of satellite images. In addition to transform techniques there are different interpolation techniques available which are used for enhancement. Wavelet transform plays a very important role in satellite image enhancement.

In addition to wavelet techniques different types of wavelet filters are used for satellite image enhancement. For image enhancement different wavelet families such as Haar wavelet, Daubechies wavelet, Coiflets wavelet, Symlets wavelets, etc., are used. All the wavelet family has its own specific property.

This paper is composed as follows:

Section 1. Introduction. Section 2. Overview of the proposed Scheme. Section 3. The Proposed Method. Section 4. Results and Discussion. Section 5. Conclusion.

2 Overview of the Proposed Method

In this paper, Discrete wavelet transform (DWT), Stationary Wavelet Transform (SWT), DWT with SWT transform are used; which are well-known techniques utilized in image enhancement. Enhancement is achieved in two parts of the proposed technique.

2.1 Discrete Wavelet Transform

If the functions capacity being expanded is a sequence of numbers, like samples of a continuous function f(x), the subsequent coefficients are known as the discrete wavelet transform (DWT) of f(x). The DWT transform match/pair can be given as [2].

$$W_\varphi(j_0, k) = \frac{1}{\sqrt{M}} \sum_x f(x) \varphi_{j0,k}(x) \tag{1}$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_x f(x) \varphi_{j,k}(x) \tag{2}$$

For j ≥ j_0 and

$$F(x) = \frac{1}{\sqrt{M}} \sum_K W_\varphi(j_0, k) \varphi_{j0,k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j0}^{\infty} \sum_k W_\varphi(j, k) \varphi_{j,k}(x) \tag{3}$$

Here f(x), $\varphi_{j0,k}(x)$, and $\psi_{j,k}(x)$ are elements of the discrete variable x = 0, 1, 2, …, M − 1.

The high recurrence components are saved by using DWT. The input image is decomposed into four sub groups that are low-low (LL), low-high (LH), high-low (HL) and high-high (HH) [3, 4]. In this 2D wavelet decomposition has been performed along the rows and then result is decomposed along the columns of an image. Discrete wavelet transform can likewise be utilized for simple and quick noise removal of an image. If we take only a predetermined number of highest coefficients of the DWT spectrum, and we perform an inverse transform (IDWT) (with the same wavelet premise) we can acquire more or less denoised image. There are different ways how to pick the coefficients that will be kept [5–7].

2.2 Stationary Wavelet Transform

The Stationary Wavelet Transform algorithm (SWT) is a wavelet change algorithm which overcomes the absence of translation-invariance of the DWT [8]. Translation-invariance in SWT is accomplished by expelling the Downsamplers in the DWT. SWT basically Upsamples the coefficients of filter by a factor $2^{(j-1)}$ in the jth level of algorithm. The SWT is naturally redundant type of plan, as the yield of each level of SWT contains indistinguishable number of samples from contained in the input. So after a decomposition of N levels wavelet coefficients have also a redundancy of N [9]. Aside from (DWT), Stationary Wavelet Transform (SWT) is one of the adaptable tools which are utilized presently in image processing. This strategy is utilized on different image processing applications such as super resolution, multiple depiction coding, video upgrading, facial remaking, etc. High frequency components of satellite images are saved by using this method. Though, losses are minimized and smoothened the images which are provided by SWT yet it has greater complexity with compared to DWT.

2.3   Thresholding Technique

To suppress the noise and restore the image various filters are used in image processing. In addition to filters several thresholding techniques can also be used to restore the image such as hard and soft thresholding.

Hard and soft thresholding with a specific threshold can be given as follows [10]:
The hard thresholding operator is characterized as

$$D(U, \lambda) = U \ for \ all \ |U| > \lambda$$
$$= 0 \ otherwise. \tag{4}$$

The soft thresholding operator can be is given as

$$D(U, \lambda) = sgn(U)\max(0, |U| - \lambda) \tag{5}$$

where input data is standardized to a range between $[-1, 1]$ and level of threshold was put to $\lambda = 0.5$ in the Figs. 1 and 2.

Hard thresholding is a keep or kill technique and is more intuitive. Hard thresholding preserves the edges and reduces noise at some extent. Though in soft thresholding, coefficients above the threshold value shrinks in the absolute value. Soft thresholding removes noise more efficiently but does not preserve edges. While at the main sight hard thresholding seems like natural but the congruity of soft thresholding gives some favourable advantages [10].

Fig. 1 Hard thresholding
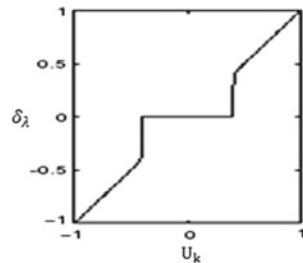
Fig. 2 Soft thresholding
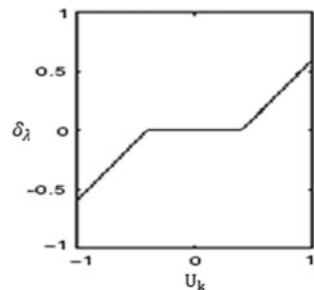
A basic method of determining λ is by calculating percentage of coefficients maxima. There are distinctive ways of picking the appropriate threshold value in accordance to the noise level which is calculated by using the wavelet coefficients. Noise level is basically variance calculated by using wavelet coefficients.

2.4 Threshold Selection

1. Universal Threshold: Soft thresholding with random Gaussian noise can be resolved in a favourable manner for by using Universal threshold. Implementation of Universal thresholding scheme is easy however it furnishes a threshold level which is larger than with other threshold decision criteria, hence it results in smoother recreated information. Likewise this type of estimation does not consider the content of the information, but only relies upon the data size [11].

$$\lambda = \sigma\sqrt{2\log(n)} \qquad (6)$$

   where n is the size of image and σ is the standard deviation of noise.
2. Minimax Threshold: Minimax threshold rule works such that estimation error which represents maximum risk of it across all locations of the data/information has been minimized. This level of threshold relies on the relationship of noise and signal in the input image/information [11].

$$\lambda = \sigma\lambda n \qquad (7)$$

   where λn is determined by the minimax rule & σ is standard deviation of noise.
3. Spatial Adaptive Threshold: Spatial adaptive threshold is based on the local variance of the input signal, which can be evaluated utilizing a local window by moving around the data/information or, more efficiently, by using a context-based clustering algorithm [11].

$$\lambda = \sigma^2/\sigma U \qquad (8)$$

   σU is the local variance of the input image.

2.5 Noise Variance Calculation

Calculate the standard deviation (σ) of the noise from the wavelet coefficients using a common estimation formula given by [11],

$$\sigma = \frac{MAD}{0.6745} \tag{9}$$

where MAD denotes median of the absolute values of the wavelet coefficients. Here median of the absolute values is calculated using cD coefficient of the wavelet.

3 The Proposed Method

The overall implementation of the proposed technique can be understood easily by the following steps (Fig. 3):

1. DWT and SWT are applied on the input Satellite image using advanced db4 wavelet. Which decomposes the input image into Approximation coefficient matrix cA and Detail coefficient matrixes cV, cH and cD. SWT is used to overcome the lack of translational-invariance of the DWT.
2. Wavelet decomposition has been performed at a level 2.
3. Standard deviation of noise has been calculated by using Eq. (9) using the high level detail coefficient.
4. Now threshold value has been estimated using minimax threshold rule.
5. After that hard thresholding technique is used to remove the noise (partially) and to preserve the edges.
6. Inverse wavelet transform has been applied on both approximation and detail coefficient to reconstruct the image.
7. Equalization has been performed by using absolute difference between error per pixel after wavelet reconstruction and the input image pixels to get the enhanced output image (Fig. 3).

4 Results and Discussion

In this section, simulation results are shown, which are carried out to evaluate the performance of the proposed methodology. Figures 4a and 5c shows Original low contrast image taken from Satellite Imaging Corporation. In this paper, Satellite Image Enhancement has been compared using three different standard wavelet transform technique as given in Tables 1 and 2.

Commonly used parameters considered for evaluation of quality enhancement of Satellite images are PSNR, MSE, SNR and SSIM.

1. PSNR (in dB): PSNR gives peak signal to noise ratio of the image. It can be calculated by the formula given as [11],

$$PSNR = 10 \log_{10} \frac{(R^2)}{MSE} \tag{10}$$
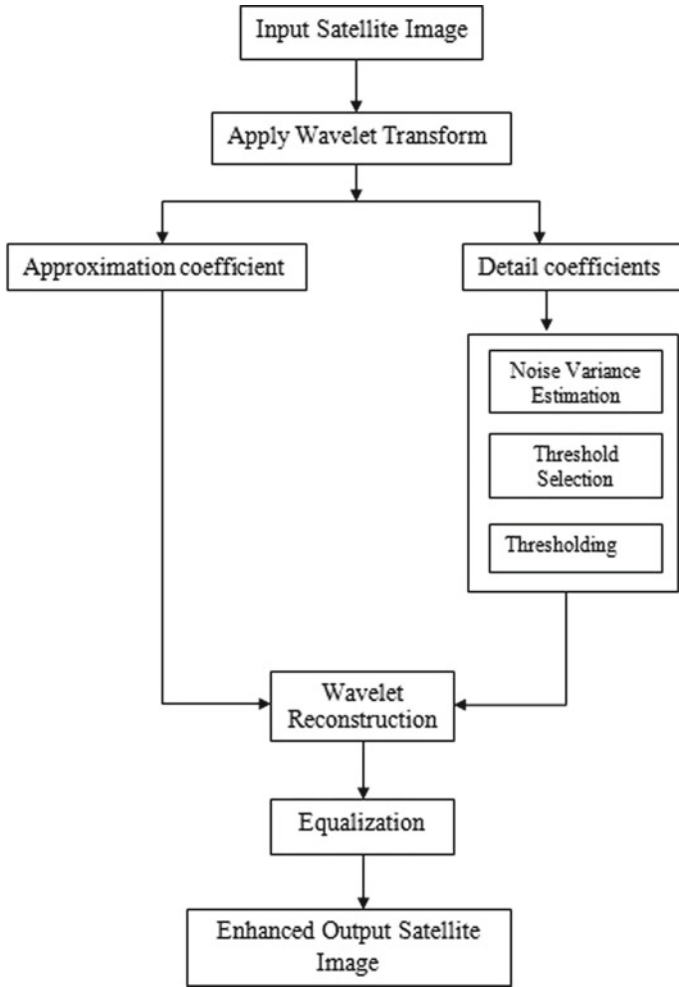
Fig. 3 Proposed methodology for satellite image enhancement

where R denotes the maximum fluctuation in the input image/data.

2. Mean Square Error (MSE): MSE represents the Mean square error between the given input image/data I_{in} and the Output enhanced image I_{org} [11].

$$MSE = \frac{\sum_{i,j}(I_{in(i,j)} - I_{org(i,j)})^2}{MXN} \tag{11}$$

3. Signal-to-Noise Ratio (SNR): SNR can be given as [12],
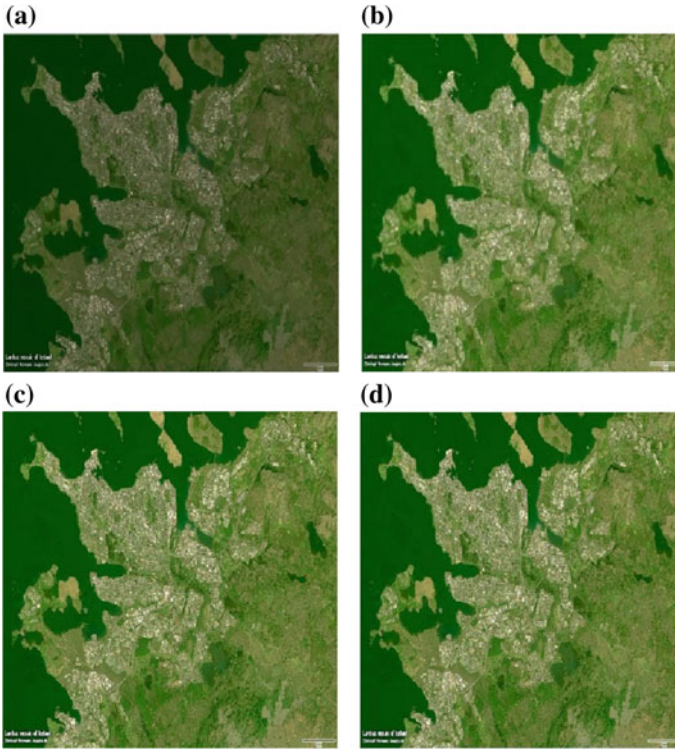
Fig. 4 **a** Original low contrast satellite image. **b** Image enhanced by DWT technique. **c** Image enhanced by SWT technique. **d** Image enhanced by DWTSWT technique

$$SNR_{rms} = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_{org}^{(i,j)^2}}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I_{org} - I_{in}]^2} \qquad (12)$$

4. Structural Similarity Index (SSIM): Structural Similarity Index is given as,

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \qquad (13)$$

where c_1 and c_2 are the constants. Its values are $c_1 = (0.01 * L)^2$ and $c_2 = (0.03 * L)^2$, L is dynamic range and its value is 255 for 8-bit grayscale image (Table 2).
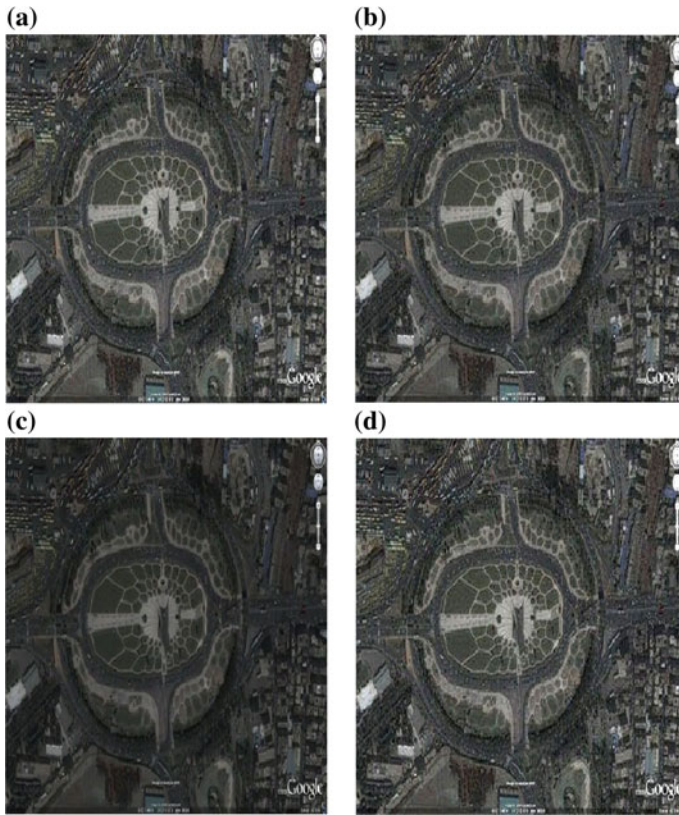
Fig. 5 **a** Original low contrast satellite image. **b** Image enhanced by DWT technique. **c** Image enhanced by SWT technique. **d** Image enhanced by DWT-SWT technique

Table 1 Comparison of results using different wavelet transforms

Parameters	Image 4			Image 5		
	DWT	SWT	DWT-SWT	DWT	SWT	DWT-SWT
PSNR	51.9401	51.9074	51.9896	55.1431	55.1784	55.4879
MSE	360.8700	362.0531	359.0893	261.9655	261.0437	253.0889
SNR	20.4509	20.0875	21.0633	15.5232	15.7597	18.6655
SSIM	0.8349	0.8334	0.8373	0.7793	0.7824	0.8141
Threshold	10.1617	10.6403	9.4147	10.1518	9.8502	7.0340

Table 2 Threshold optimization for satellite images

Image 4	Threshold	9.4147	5.4147	2.4147	**0.4147**	0.04147	10.4147	15.4147
	PSNR	51.9896	52.2139	52.3074	**52.3236**	52.3137	51.9227	51.5607
	SSIM	0.8373	0.8487	0.8541	**0.8556**	0.8554	0.8341	0.8188
Image 5	Threshold	7.0340	5.0340	2.0340	**0.0340**	0.00340	8.0340	10.0340
	PSNR	55.4879	55.6593	55.8069	**55.8091**	55.8020	55.3829	55.1569
	SSIM	0.8141	0.8368	0.8615	**0.8640**	0.8632	0.8024	0.7805

5   Conclusion

In this paper, we considered contrast enhancement technique that can produce resultant satellite image that looks better subjectively by restoring the image and contrast enhancement by equalizing the pixel intensities. The experimental results are compared with three different transform techniques. The algorithm was applied on different images taken from satellite. It has been observed that the DWT with SWT technique provides better result compare to DWT and SWT technique because it utilizes the translational invariance property of SWT. The transform domain Satellite image enhancement algorithm on Satellite images, with thresholding and equalization method demonstrates the robustness of the proposed algorithm. The robustness of the wavelet transform methods has been quantitatively measured by comparing the PSNR, MSE, SNR and SSIM values with the original satellite images. In future, the proposed transform domain algorithm can be combined with other transform algorithms to obtain better enhancement of satellite images. This work can be extended by using combination of spatial domain and wavelet transform (WT) domain for the enhancement of satellite images. Also this algorithm can be combined with other algorithm for both resolution and contrast enhancement of satellite images simultaneously. As we decrease the value of threshold PSNR is increased and also visual quality of image increases. Threshold can also be optimized by using other transform technique for a particular satellite image, for increasing the quality of image.

References

1. Ahuja, S.N., Biday, S.: A survey of satellite image enhancement techniques. Int. J. Adv. Innov. Res. (IJAIR) **2**(8) (2013). ISSN: 2278-7844
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice-Hall, Englewood Cliffs, NJ (2007)
3. Demirel, H., Ozcinar, C., Anbarjafari, G.: Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition. IEEE G and RS Lett. **7**, 333–337 (2010)
4. Rathod, M., Khanapuri, J.: Satellite image resolution enhancement using SWT and DWT with SWT. In: IEEE International Conference on Nascent Technologies in the Engineering Field (ICNTE), 978-1-5090-2794-1 (2017)
5. Demirel, H., Anbarjafari, G.: Discrete wavelet transform-based satellite image resolution enhancement. IEEE Trans. Geosci. Remote. Sens. **49**(6), 1997–2004 (2011)

6. Bhandaria, A.K., Sonia, V., Kumara, A., Singh, G.K.: Artificial bee colony-based satellite image contrast and brightness enhancement technique using DWT-SVD. Int. J. Remote. Sens. **35**(5), 1601–1624 (2014)
7. Shamna, K.S.: Satellite image resolution and brightness enhancement using discrete, stationary wavelet and singular value decomposition. In: IEEE International Conference on Power, Signals, Controls and Computation (EPSCICON), 978-1-4799-3612-0 (2014)
8. Suganya, P., Mohanapriya, N., Vanitha, A.: Survey on image resolution techniques for satellite images. Int. J. Comput. Sci. Inf. Technol. **4**(6), 835–838 (2013). ISSN: 0975-9646
9. Nason, G.P., Silverman, B.W.: The Stationary Wavelet Transform and Some Statistical Applications. Springer, New York (1995)
10. Paul, O.I., Lu, Y.: Image denoising using wavelet thresholding techniques. Int. J. Educ. Res. **2**(2) (2014). ISSN: 2201-6740
11. Bouchouareb, R., Benatia, D.: Comparative study between wavelet thresholding techniques (hard, soft and invariant-translation) in ultrasound images. Int. J. Bio-Sci. Bio-Technol. **6**(6), 29–38 (2014). ISSN: 2233-7849
12. Sajid, M., Khurshid, K.: Satellite image restoration using RLS adaptive filter and enhancement by image processing techniques. IEEE transactions, 978-1-4673-8240-3 (2015)

An Efficient Image Watermarking Technique Based on IWT-DCT-SVD

Priyank Khare and **Vinay Kumar Srivastava**

Abstract Confidentiality of multimedia data such as image must be protected from illicit users. Thus, a novel image watermarking technique using Integer Wavelet Transform (IWT), Discrete Cosine Transform (DCT), and Singular Value Decomposition (SVD) is proposed in this paper. LL (low-low) subband is selected after applying one-level IWT on host image, which is further processed with DCT. Embedding of watermark is performed directly into Singular Values (SVs) of transformed image. Robustness of proposed technique is comprehensively determined under various set of attacks such as scaling and histogram equalization. The comparative study is also done among proposed technique and other prevailing techniques and experimental results clearly indicate effectiveness of this proposed method.

Keywords Image watermarking · Integer wavelet transform · Discrete cosine transform · Singular value decomposition · Robustness

1 Introduction

Digital technology has rapidly advanced in previous years due to Internet. The Internet has become a convenient medium for exchanging multimedia contents such as images, video, and audio etc. among different users in a very short duration of time. This exchange of information must be secured from various illicit users which generally manipulates, tampers the data easily. So, there is a requirement of technique which can provide copyright protection. Image watermarking [1] is a method which ensures security against these malicious attacks. In this technique, generally, confidential information such as the logo of the organization, is made hidden inside the original host image in such a manner that it becomes imperceptible. This information

P. Khare (✉) · V. K. Srivastava
ECED, MNNIT Allahabad, Allahabad, India
e-mail: rel1606@mnnit.ac.in

V. K. Srivastava
e-mail: vinay@mnnit.ac.in

can be later extracted by authorized users at receiver side. Salient features of watermarking technique are imperceptibility, robustness, capacity, and security. Imperceptibility signifies the perceptual invisibility among original and watermarked images. Robustness indicates strength of algorithm against various attacks such as filtering and translation etc. whereas capacity defines amount of bits embedded in watermark. Lastly, security ensures copyright protection. Mainly watermarking techniques are categorized into two domains; first is spatial domain and second is transform domain. Embedding takes place by just varying pixel values of image [2], whereas in transform domain, transform coefficients are altered for embedding of watermark. Various transforms employed are DCT, Discrete Wavelet Transform (DWT), etc. This method is superior over spatial domain in terms of robustness. Recently researchers started using the fusion of several transforms such as DCT-SVD [3], DWT-SVD [4], DCT-DWT-SVD [5], etc. to strengthen robustness and imperceptibility.

Sakthivel and Sankar [6] developed an approach using IWT and SVD. LL_2 subband of IWT decomposed cover image is used for embedding of the watermark. This method is also tested using Verilog for hardware implementation. In [7], the authors suggested a method in which original image is decomposed using DWT in various subbands followed by SVD. SVD is performed on watermark image. Finally, SVs of watermark are embedded with SVs of DWT transformed image. This results in a non-blind method. Authors in [8] developed a SVD-based approach which directly embeds watermark into SVs of host image. Though this method achieves low robustness. In [9], researchers investigated method using IWT-SVD-Arnold Transform (AT). In their approach, IWT decomposes cover image resulting in various subbands, subsequently, SVD of each subband is performed. Watermark is scrambled using AT to provide security. Scrambled watermark is directly embedded into SVs of transformed host image. This method shows good degree of robustness. Lagzian et al. [10] developed method similar to [7]. The only difference is that Redundant Discrete Wavelet Transform (RDWT) is used instead of DWT. Researchers in [11] suggested a method based on significant difference of IWT coefficients. Largest coefficients of randomized HL_3 subband are chosen for embedding of watermark.

The remaining portion of this paper is arranged as follows: Preliminaries about IWT, DCT, and SVD are discussed in Sect. 2 whereas Sect. 3 discusses the proposed watermarking technique. Section 4 deals with experimental results. Conclusions are presented in Sect. 5.

2 Preliminaries

2.1 Integer Wavelet Transform

Sweldens [12] introduces Lifting Wavelet Transform (LWT) which overcomes limitations of DWT. In LWT, mapping is performed between integers to integers which completely eliminates truncation. LWT also requires lesser memory requirement.

Fig. 1 Schematic diagram
for LWT
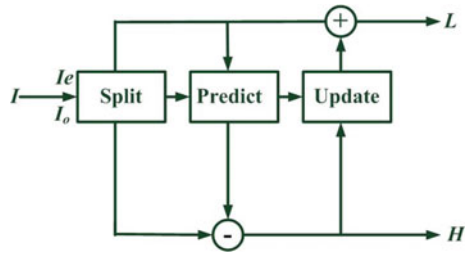
IWT uses this lifting scheme, as this process is reversible. Hence, perfect reconstruction is achieved. The steps for lifting scheme are summarized below:

Split: Input signal "I" is splitted in even "I_e" and odd samples "I_o".
Predict: In this phase, the predictor is used to determine the odd samples from even samples.
Update: Predicted sample output is used to generate new even samples with the help of original even samples.

Figure 1 presents the implementation of the lifting scheme where "H" and "L" denotes output of predictor and update stages.

2.2 Discrete Cosine Transform (DCT)

DCT is an image transform used for converting image to frequency domain from spatial domain [13]. Application of DCT on the image can be represented as weighted sum of sinusoids oscillating at various frequencies. DCT is a real and orthogonal transform which is having excellent energy compaction and data decorrelation properties.

$$I(U, V) = \frac{2}{\sqrt{MN}} \beta(U)\beta(V) \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} \left(i(x, y) \cos\left(\frac{U(2x+1)\pi}{2M}\right) \cos\left(\frac{V(2y+1)\pi}{2N}\right) \right) \quad (1)$$

Equation (1) denotes DCT of an image "$i(x, y)$", here $U = 0, 1, 2 \ldots M - 1$ and $V = 0, 1, 2 \ldots N - 1$. $\beta(U)$ and $\beta(V)$ are constants.

2.3 Singular Value Decomposition (SVD)

In linear algebra, diagonalization of any real matrix can be easily performed with SVD. SVD [14] decomposes a real matrix "R" into three matrices as orthogonal matrices U, V, and diagonal matrix "S" which contains singular values arranged in decreasing manner. SVD is effectively used due to fact that it resists slight variations

on image; hence, providing excellent stability to SVs which are used for embedding and extraction of watermark.

$$R = U_R S_R V_R^T \tag{2}$$

3 Proposed Algorithm

This section presents the watermark embedding and extraction algorithms in detail. The schematic diagram of proposed embedding and extraction are shown in Fig. 2 and Fig. 3, respectively.

3.1 Embedding Algorithm

Watermark embedding is explained below in the following steps:

- Host image "F" is decomposed in various subbands (LL, HL, LH, and HH) using 1-level IWT.
- Now LL subband is chosen for embedding and DCT is applied on it resulting in "F_{DCT}".
- SVs of "F_{DCT}" are obtained using SVD.
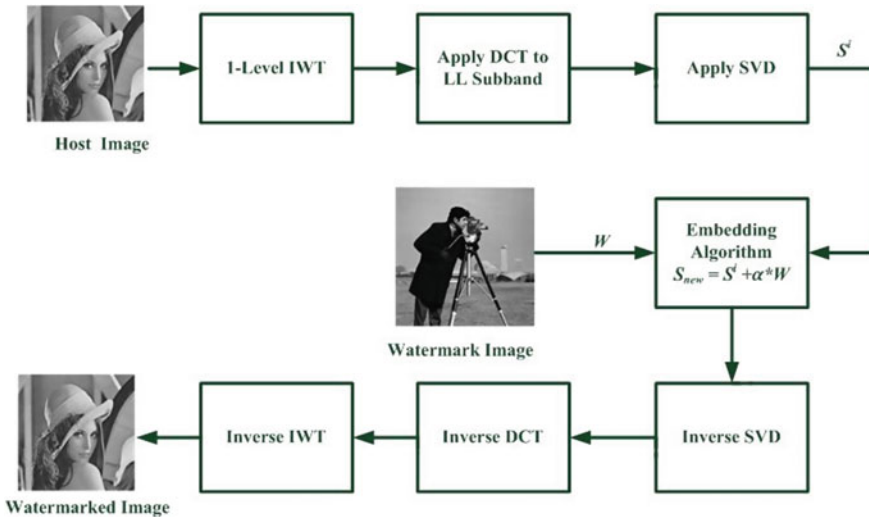
Fig. 2 Schematic diagram for embedding process

Watermarked Image

Extracted Watermark Image
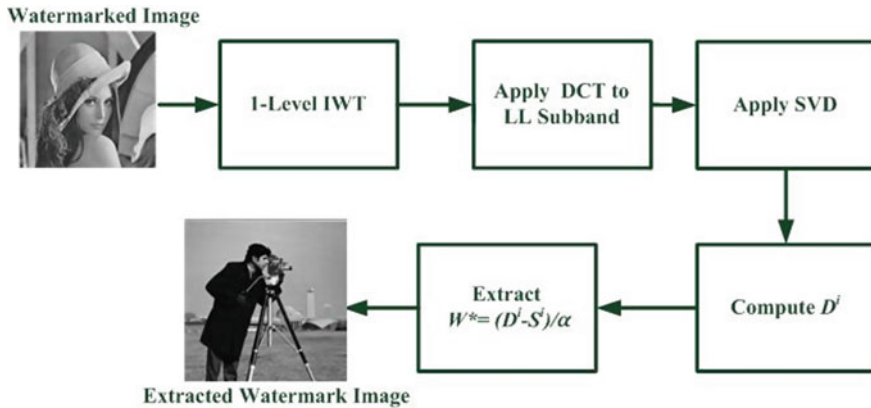
Fig. 3 Schematic diagram for extraction process

$$F_{DCT} = U^i S^i V^{iT} \tag{3}$$

- Watermark is directly embedded into SVs "S^i" of "F_{DCT}" according to Eq. (4).

$$S_{new} = S^i + \alpha * W \tag{4}$$

here "α" is scaling factor whose value is taken as 0.05.

- New SVs "S_{new}" are again modified using SVD.

$$(S_{new}) = U^i_p S^i_p V^{iT}_p \tag{5}$$

- New modified coefficients of IWT for LL subband are obtained as "F^{new}".

$$F^{new} = U^i S^i_p V^{iT} \tag{6}$$

- Lastly, inverse DCT followed by inverse IWT is applied on "F^{new}" to obtain "F_w" as watermarked image.

3.2 Extraction Algorithm

This subsection discusses the extraction of watermark from watermarked image "F_w".

- 1-Level IWT is performed on "F_w" to split into various frequencies subbands.
- Now, DCT of LL subband is performed for obtaining "F_{DCT}^*".
- SVs of "F_{DCT}^*" are obtained using SVD.

$$F_{DCT}^* = U^{*i} S^{*i} V^{*iT} \qquad (7)$$

- Calculate "D^i" according to Eq. (8).

$$D^i = U_p^i S^{*i} V_p^{iT} \qquad (8)$$

- Watermark "W^*" is extracted as

$$W^* = \left(D^i - S^i\right)/\alpha \qquad (9)$$

4 Experimental Results

This section highlights various simulation results of the proposed method. The proposed algorithm is tested on three standard grayscale images which are chosen as "Lena", "Boat", and "Lake" of 512 × 512, whereas "Cameraman" image of 256 × 256 is taken as watermark as shown in Fig. 4 whereas Fig. 5 demonstrates different watermarked images and extracted watermark. MATLAB is used to implement the proposed method.

Fig. 4 Different host images **a** Lena. **b** Boat. **c** Lake. **d** Watermark image

Fig. 5 Different watermarked images **a** Lena. **b** Boat. **c** Lake. **d** Extracted watermark image

Proposed technique performance is evaluated with Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Normalized Correlation Coefficient (NCC). PSNR and SSIM are parameters used to measure imperceptibility of scheme. Imperceptibility means visually both original and watermarked images should be analogous. PSNR above 25 dB is acceptable. Mathematically PSNR is given as

$$PSNR = 10 \times \log_{10}\left(\frac{255}{MSE}^2\right) \tag{10}$$

where MSE is mean square error evaluated between original and watermarked image. SSIM [15] gives information about the structural similarity between original and watermarked images.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{11}$$

here μ_x, μ_y are mean of x and y. σ_x, σ_y are the standard deviation of x and y, whereas σ_{xy} is covariance between x and y, respectively.

NCC is another parameter used to measure the robustness of the proposed scheme against signal processing attacks like sharpening, rotation, etc. NCC values should lie close to $+1$ for acceptable recovery of watermark. NCC is given by Eq. (12)

$$NCC = \frac{\sum_{a=1}^{c}\sum_{b=1}^{d}[W(a, b)W^*(a, b)]}{\sqrt{\sum_{a=1}^{c}\sum_{b=1}^{d}[W(a, b)]^2}\sqrt{\sum_{a=1}^{c}\sum_{b=1}^{d}[W^*(a, b)]^2}} \tag{12}$$

where W, W^* are original and extracted watermark, respectively.

Table 1 PSNR, SSIM and NCC values for various images under no attacks

Image	PSNR (dB)	NCC	SSIM
Lena	47.4396	1.0	0.9980
Boat	51.9673	1.0	0.9985
Lake	52.5517	1.0	0.9988

Table 1 presents the value of PSNR, SSIM and NCC for various host images. It can be concluded from results that the proposed scheme attains high value of performance parameters under no attacks.

Table 2 tabulates NCC values of the extracted watermark from different watermarked images under numerous attacks. Proposed technique shows good robustness toward these attacks.

The comparative study among proposed scheme and existing techniques such as [7, 9] is carried out in Table 3. Clearly, the proposed scheme surpasses the other existing methods.

Table 2 Robustness of proposed scheme under various attacks

Attacks	Lena	Boat	Lake
Salt and pepper noise (0.001)	1.00	1.00	1.00
Gaussian noise (0, 0.3)	0.9904	0.9977	0.9878
Histogram equalization	0.9831	0.9881	0.9937
Gaussian filter (3 × 3)	0.999	1.00	1.00
Median filtering (3 × 3)	0.9996	1.00	0.9989
JPEG compression (Q = 50)	1.00	1.00	1.00
Rotation (50°)	0.8873	0.9066	0.8846
Scaling (0.5, 2)	0.9996	1.00	0.9995
Cut (20 rows)	0.9977	0.9962	0.9965
Translation [10 10]	0.9996	0.9992	0.9995

Table 3 Robustness comparison among proposed scheme and existing schemes

Attacks	Ganic and Eskicioglu [7]	Makbol and Khoo [9]	Proposed method
Gaussian noise (0, 0.3)	0.865	0.77	0.9904
Histogram equalization	0.586	0.986	0.9831
Median filtering (3 × 3)	–	0.981	0.9996
JPEG compression(Q = 30)	0.993	0.994	1.00
Scaling (0.5, 2)	0.940	0.971	0.9996
Cut (20 rows)	–	0.99	0.9977
Sharpening	0.528	0.950	1.00

5   Conclusions

A novel approach of image watermarking is proposed in this paper which uses a combination of IWT-DCT-SVD. IWT makes scheme faster since it directly maps integer to integer. A fair amount of robustness is achieved by using LL subband for embedding the watermark. Direct insertion of watermark makes the scheme blind. It can be noticed from analysis of experimental results that the proposed scheme attains high values of PSNR and NCC against attacks. On comparison with prevailing schemes [7, 9] present in the literature, the proposed method is found to be more robust. Hence, an effective technique of watermarking is developed in this paper.

References

1. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital watermarking and steganography, 2nd edn. Morgan kaufmann (2007)
2. Nikolaidis, N., Pitas, I.: Robust image watermarking in the spatial domain. Sig. Process. **66**(3), 385–403 (1998)
3. Ali, M., Ahn, C.W., Pant, M.: A robust image watermarking technique using SVD and differential evolution in DCT domain. Opt.-Int. J. Light. Electron Opt. **125**(1), 428–434 (2014)
4. Lai, C.C., Tsai, C.C.: Digital image watermarking using discrete wavelet transform and singular value decomposition. IEEE Trans. Instrum. Meas. **59**(11), 3060–3063 (2010)
5. Al-Haj, A.: Combined DWT-DCT digital image watermarking. J. Comput. Sci. **3**(9), 740–746 (2007)
6. Sakthivel, S.M., Sankar, A.R.: Real time watermarking of grayscale images using integer DWT transform. In: Proceedings of the International Conference on VLSI Systems, Architecture, Technology and Applications (2016)
7. Ganic, E., Eskicioglu, A.M.: Robust embedding of visual watermarks using discrete wavelet transform and singular value decomposition. J. Electron. Imaging **14**(4), 043004 (2005)
8. Liu, R., Tan, T.: An SVD-based watermarking scheme for protecting rightful ownership. IEEE Trans. Multimedia **4**(1), 121–128 (2002)
9. Makbol, N.M., Khoo, B.E.: A hybrid robust image watermarking scheme using integer wavelet transform, singular value decomposition and arnold transform. In: International Visual Informatics Conference, pp. 36–47 (2013)
10. Lagzian, S., Soryani, M., Fathy, M.: A new robust watermarking scheme based on RDWT-SVD. Int. J. Intell. Inf. Process. **2**(1), 22–29 (2011)
11. Verma, V.S., Jha, R.K.: Improved watermarking technique based on significant difference of lifting wavelet coefficients. Signal, Image Video Process. **9**(6), 1443–1450 (2015)
12. Sweldens, W.: The lifting scheme: a construction of second generation wavelets. SIAM journal on mathematical analysis **29**(2), 511–546 (1998)
13. Khayam, S.A.: The Discrete Cosine Transform (DCT): Theory and Application, vol. 114. Michigan State University (2003)
14. Chung, K.L., Yang, W.N., Huang, Y.H., Wu, S.T., Hsu, Y.C.: On SVD-based watermarking algorithm. Appl. Math. Comput. **188**(1), 54–57 (2007)
15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

Sparse Matrix Completion for Effective Recommendation System

Vivek Kumar Singh, Anubhav Shivhare and Manish Kumar

Abstract   In this era of information retrieval, the revenue of the e-commerce system mainly rely upon how intelligently information is being processed and decision are being made. In this paper, the problem of unavailability of complete information is targeted for information processing and decision-making in e-commerce domain. The existing system which works on a similar phenomenon is known as recommendation system which becomes the most evolving subject in the area of electronically operated markets. However, the decision-making capability and suggestive nature of these techniques have improved the overall output of the electronic market, but lacking behind where sufficient information is not available. In literature, most of the existing schemes are designed using collaborative filtering and content-based recommendations with KNN and K-means. But, huge increment in the number of online users and their varied pattern of purchasing goods increase the sparsity in information matrix due to which neighbor selection is getting more problematic. The proposed recommended system embeds the collaborative filtering method to complete the obtained incomplete matrix which lifts the aforementioned problem of high sparsity in data. The proposed scheme is verified and validated over different datasets and achieve significant results over existing schemes.

Keywords Recommendation system · Matrix completion · Collaborative filtering · Sparse matrix

V. K. Singh (✉) · A. Shivhare · M. Kumar
Indian Institute of Information Technology, Allahabad, India
e-mail: vivekshikarvar@gmail.com

A. Shivhare
e-mail: PCL2015002@iiita.ac.in

M. Kumar
e-mail: manish@iiita.ac.in

1   Introduction

Recently, the recommendation system becomes a game-changing framework for not only e-commerce industries like Flipkart, Amazon, etc., but also for many other online business platforms such as Facebook and YouTube. [1, 2]. Every day, this platform comes with various choices and options like what to eat? where to go? what to read? What to wear? What movie to rent? What stock to buy? What blog post to read? The availability of these online platforms in such a highly competitive environment presents a wide variety of same products with different pricings which leads buyers or subscribers in confusion in terms of quality and assurance. In order to avoid such confusion among the customers, bigger companies like Amazon, Flipkart, Netflix, and Ebay have adopted the technique of recommendation based on the past user's experience and their needs. The commercial success of these bigger firms have been studied in recent years which states that the user satisfaction and the ease of post-purchase support impact the performance in terms of revenue. The indirect nature of marketing of these companies impose more difficulty in attracting the customers, the only way is recommendation based on past experiences and user's rating.

The reason behind the continuous research is the relevancy and uniqueness of products being recommended is a big question in today's competitive environment as size of data available over Internet mislead these systems and poorly perform in real time. For example, Netflix has over 17,000 movies in its selection, and Amazon has over 410,000 titles in its Kindle store alone. Discovery in information spaces of this magnitude is a known challenge in this field. Even simple decision what movie should I see this weekend? can be difficult without prior direct knowledge of the candidates and with this growing technology, people want everything to be on point even with the decision. So, this is a simple scenario where the recommendation system comes in the picture. Computer-based systems provide the opportunity to expand the set of people from whom users can obtain recommendations. This system also enables the analyst to mine user's history and stated preferences for patterns. However, many schemes exist which provide more or less some predictions or recommendations by recognizing the buying pattern of the customers and matching that pattern with other customer's buying pattern. But, it must interact with the user for both to learn the user's preferences and provide recommendations. To achieve relevancy and user satisfaction for these recommendations, systems must have accurate and relevant data from which to compute their recommendations and preferences, leading to work on how to collect reliable data and reduce the noise in user preference datasets. Multiple techniques have been deployed by researchers for fast and accurate prediction based on user preferences like PCA-GAKM is utilized in [3] to search in dense movie space and performance results have been validated using Movielens dataset. Similarly, PCA-SOM has been refined in [4] based on the local data distribution stored in covariance matrix such that no prior information of the principal subspace is required. Similarly, K-means (KM), GA-based KM (GAKM), and PCA-KM results for prediction performances have been analyzed in many research works like [5]. In this work, the problem of sparsity in the obtained information from different users

is handled with the matrix completion approach and the recommendation is done using a collaborative filtering method over obtained complete matrix. The proposed scheme works in two phases: The first phase illustrates how the incomplete user's information can be completed based on their interest and similarity. In the second phase, the collaborative filtering uses the Pearson Correlation coefficient to compute the similarity in two rows of the obtained matrix. Based on the similarity index, the system suggests recommendations to the users. The main contributions of this work are

- Sparsity in incomplete information matrix obtained through similar users is taken care of using matrix completion approach.
- The obtained complete matrix is then used for getting recommendations using collaborative filtering which significantly increases the performance of the system.

The rest of the paper is organized as, Sect. 2 presents the existing work done in the same domain, Sect. 3 describes the problem statement, the proposed idea is discussed in Sect. 4, Sect. 5 explains the results obtained and Sect. 6 gives the concluding remarks.

2 Related Works

In recommendation systems, Top-N recommender is a widely researched application by industry and academia alike. In [6], authors maintain a low-rank attribute matrix and also keep original information to improve the prediction performance. Further, reliable prediction for unrated items is proposed in [7] using low-rank and sparse matrix completion (LRSMC) algorithm which derives rating matrix to improve the prediction. In the last decade, matrix factorization has been used to solve the problem of collaborative filtering. The idea was to factorize the rating matrix into a lower dimension based on user dependent vectors either by using SGD [8]; alternating least square methods [9] and probabilistic matrix factorization [10]. Earlier, traditional algorithms for collaborative filtering rely on user-item relational matrix, but these algorithms have limited application in a cold start situation. Hence, specifically for cold start problems, trust networks [11, 12] and social tags [13] have been utilized to collect auxiliary information or user-related attributes. Authors in [14] have leveraged similarity between users and to summation of weighted ratings to simply predict ratings for new user. Recently, due to great achievements in the field of imaging and artificial intelligence, research interest is increased in recommendation systems. Collaborative deep learning is a recent example that uses deep learning methods for recommendation systems. It takes user situation, interaction, and implicit behavior into account to make recommendations. Timestamp and location information in addition to this scheme further increase the performance of recommendation system like TimeSVD++ [15]. This model changes static and latent factors to time-dependent parameters to change the user interaction dynamics. To handle the cold start problem for a new user or a new item, social trust networks [12], tagging system

[13], and interview process have been taken into account. Due to data security and user privacy issues of cyber security paradigms, such approaches face challenges in data collection. Hence, the collection of item-related attributes becomes important especially for cold start problems. Despite great attempts to solve the problem, the following research gaps still exist. First, it is difficult to find the features of a newly introduced item using rough attributes. Second, collecting and using the attribute information through networks, tagging, key words, etc., is time and cost consuming. Lastly, many proposed solutions using probabilistic models for user recommendation system [16] do not take time information into account. In this work, authors try to address these issues by integrating deep learning and collaborative filtering to increase the performance of recommendation system, in particular for a cold start situation.

3 Problem Definition

In the past decade, recommendation systems have improved rapidly and growing day by day. With the rapid growth of information over these online platforms, it becomes much more demanding for effectively extracting useful information. Platforms like Amazon, Netfilx, Wynk, etc., are continuously using such systems so that the customers has less task to search the desired option and are more likely attracted to the platform for future needs. This makes the platform more in demand and likeD by all. But the process of preparing one is a tedious task. A recommendation system can be divided into mainly three categories: neighborhood-based collaborative filtering, model-based collaborative filtering, and ranking-based methods. Although, many schemes are proposed in each category, all work on the fact that the amount of information present is sufficient for inferential task and recommendation to be suggested. Based on these facts, the following problems focused in this work are as follows:

1. **High sparsity in obtained information** The various availability of the products or services on these online platforms provides the user a wide range to compare the quality and cost which consequently leads to sparse entries in the information matrix as the choice of one user may vary from the others. Due to rare availability of information for each item or user category, the sparseness increases and makes this task of inference a tedious one for analyst or system based on which recommendation can be made. So, sparsity is a major issue which must be addressed while designing a recommended system to improve the performance of the system.
2. **No information i.e. cold start problem** The cold start is a potential problem in computer-based information systems which involves a degree of automated data modeling. Specifically, it concerns with the issue that the system cannot draw any inference from the user or item about which it has not gathered sufficient

information. There are basically three kinds of cold start problem which can be classified as

– User cold start: When new users enter in the lobby and there is no history available for recommendations.
– Item cold start: When new item added in the product catalog and there is no history of user's rating available.

So, the problems discussed here not only improve the demand of the platform but also increase the accuracy of the recommendation system.

4 Methodology and Implementation

The general principle of the recommendation system is to identify the similarities among users/items. For example, item-based k-nearest-neighbor (Item KNN) which is a collaborative filtering method first identifies a set of similar items for each of the items that the consumer has purchased, and then recommend Top N items based on those similar items. However, it suffers from low accuracy due to the fact of unavailability of sufficient data, i.e., known as cold start problem. The problem of sparse data availability in both User cold start and Item cold start cases is handled by matrix completion approach of sparse data recovery by using Jaccard distance and cosine similarity index and then the collaborative filtering approach is used to get the recommendations over complete matrix. But, the drawback which comes up with Jaccard distance for similarity checking ignores the rating values and missing values. So, we embed cosine similarity index checking with collaborative filtering.

4.1 Matrix Completion

Jaccard Distance: The Jaccard distance is an statistic to measure the similarity or diversity in the sample set, defined as (Tables 1 and 2)

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{1}$$

Here, $X \cap Y = 1$, $X \cup Y = 5$

Hence, Jaccord Similarity $= \frac{1}{5}$, Jaccard Distance $= \frac{4}{5}$; When $X \cap Y = 2$, $X \cup Y = 4$ Jaccord Similarity $= \frac{2}{4}$, Jaccard Distance$= \frac{1}{2}$.

The problem with the Jaccard index is it only takes the presence of element under consideration and not the magnitude of the element. But, in our case, the magnitude of the element is equally important as the presence is important. Because the presence of an element signifies that the product should be recommended but the magnitude of the element signifies the priority of the product being recommended. In case

of neglecting the magnitude, the relevant products might not be recommended. In that case, the cosine similarity index is used. Although, Jaccard distance is able to recommend but not gives relevancy in the recommended products.

Cosine Similarity: The cosine similarity between user profile x and item profile i can be estimated as

$$U(x, i) = cos() = \frac{(x, i)}{|x||i|} \tag{2}$$

The cosine index of two attributes signifies more similarity in case of smaller angle and shows high dissimilarity between attributes if the angle is large. The cosine of the angle between X and Y is 0.38. The cosine of the angle between X and Z is 0.32. Large cosine value implies a smaller angle and therefore smaller distance. Thus, X is closer to Y as compared to Z. Similarity among these attributes are very close to each other but the problem is that it treats missing rating as negative which is not accepted. So, initially, we assign 0 to empty boxes and it implies that users have not rated the particular item. The scheme calculates the initial cosine index for each entry and then fills up the missing entries with the average value of respective column by normalizing the value. As all the values are between 0 and 1, so normalizing it at it's mean brings the missing entries (i.e., 0) more closer to other similarity indexes. It implies the least prioritized item in the recommended items list with centered cosine index rather than removing it from the list of recommendations.

Centered Cosine Similarity: Normalize rating by subtracting row mean, i.e., we have centered the rating of each user around zero. Zero becomes average rating for every user. The cosine similarity of X and Y now becomes 0.092. The cosine similarity of X and Z now becomes 0.083. Now, similarity(X,Y) is greater than similarity(X,Z)

Table 1 Finding Jaccord Distance

Name	HP1	HP2	HP3	TW	SW1	SW2	SW3
X	4			5	1		
Y	5	5	4				
Z				2	4	5	
W		3					3

Table 2 Finding jaccord distance

Name	HP1	HP2	HP3	TW	SW1	SW2	SW3
A	$\frac{2}{3}$			$\frac{5}{3}$	$\frac{-7}{3}$		
B	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{-2}{3}$				
C				$\frac{-5}{3}$	$\frac{1}{3}$	$\frac{4}{3}$	
D		0					0

and missing rating treated as an average. Thus, on the purchase of item X, priority of item Y in the recommendation list gets higher value in comparison to item Z.

Now, for filtering the noise from the completed matrix, collaborative filtering is applied to get the more relevant similarity matrix. It is applied over normalized matrix obtained through cosine similarity index. The collaborative filtering is based on the user preference that is being generated from the Pearson correlation coefficient. It computes the similarity between $n^t h$ user and other users. In simple words, Pearson correlation computes the statistical correlation among two items and their mutual ratings to determine their similarity. The PCC can be formulated as

$$S(u, v) = \frac{\sum_{i=I_u \cap I_v} (r_u, i - r_u)(r_v, i - r_v)}{\sqrt{\sum_{i=I_u \cap I_v} (r_u, i - r_u)^2} \sqrt{\sum_{i=I_u \cap I_v} (r_v, i - r_v)^2}} \tag{3}$$

The value of $S(u, v)$ ranges from -1 to 1. Here, we observe that if a value equals to -1, then that is said to be a perfect negative indication correlation. It means there is no similarity between the n users and other users and vice versa. For the movie and their users, K-means algorithm aims to partition these users into k groups automatically.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} ||x_i^j - C_j||^2 \tag{4}$$

where $J =$ Objective Function, $k =$ number of clusters, $n =$ number of cases, $C_j =$ Centroid, and $||x_i^j - C_j||^2 =$ Distance Function

We then evaluate the performance of the proposed method using mean absolute errors. We know that mean absolute error(MAE) is a measure of statistical accuracy through which the accuracy of the scheme is defined.

$$MAE = \frac{\sum |P_{ij} - r_{Ij}|}{M} \tag{5}$$

where M is the total number of predicted movies, P_{ij} represents the predicted value for user i on item j, and r_{ij} is the true rating.

5   Results and Evaluation

Results have been put in this section to support the concepts proposed in this scheme. The scheme has been evaluated on two parameters

1. Scheme evaluation on different datasets from UCI libraries to find corresponding MAE value, as shown in (Table 3).
2. Figure 1 shows a comparative analysis of proposed scheme with various existing recommendation system techniques in pictorial format.

Table 3 Table for MAE Based on Datasets

udata20	u.data20	u.data30	u.data40	u.data50	u.data60	u.data70	u.data80	u.data90	u.data100
0.8254	0.8124	0.8758	0.8125	0.8485	0.8958	0.8214	0.8485	0.8412	0.8589

Fig. 1 Comparison of MAE Values Based on Different Techniques
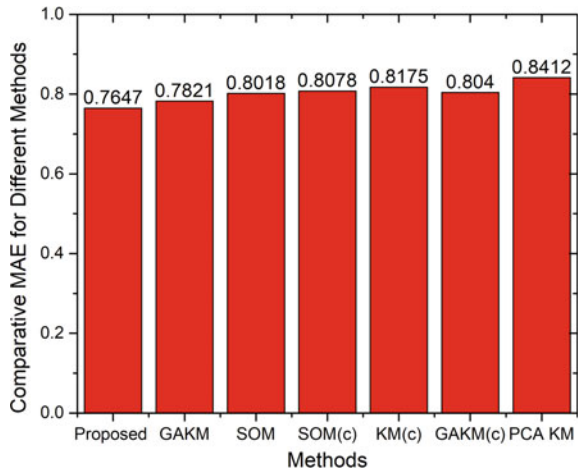
MAE is calculated on different dataset files, namely udata20, u.data20, u.data30, udata40, u.data50, u.data60, u.data70, u.data80, u.data90, u.data100. These files contain a different number of movies described as parameter x. Later, individual MAE for each dataset is combined to get absolute MAE for entire movie dataset. Thus, the model based on collaborative filtering uses concept of matrix completion to create a movie rating and recommendation system with minimal noise. It is clearly observable that MAE over different datasets is minimum for u.data70 and MAE is also minimum for the proposed scheme compared to other existing schemes. This is because, for cold start problems, K-means has been used to find the initial correlation, while Pearson correlation coefficient helps in finding actual correlation among users, items (here, movies), and systems. As a result, the proposed scheme outperforms other schemes in the domain of collaborative filtering because they mostly use traditional classification and clustering algorithms that too at a single-stage level; and thus provide a limited solution to cold start problems.

6   Conclusion

The paper targeted the problem of sparsity in information for recommendation system applicable in various application domains. The approach used in the work is the combination of collaborative filtering with matrix completion approaches which has lifted the problem of sparsity. The sequential working of these techniques has improved the performance of the system and accuracy by 5% in comparison to

existing schemes. Besides this, for future work, we can continue to improve our work approach that possibly can deal with a larger dataset.

References

1. Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S. et al.: The YouTube video recommendation system. In: Proceedings of the Fourth ACM Conference on Recommender Systems, pp. 293–296. ACM (2010)
2. Wang, Z., Liao, J., Cao, Q., Qi, H., Wang, Z.: Friendbook: a semantic-based friend recommendation system for social networks. IEEE Trans. Mob. Comput. **14**(3), 538–551 (2015)
3. Wang, Z., Xue, Y., Feng, N., Wang, Z.: An improved collaborative movie recommendation system using computational intelligence. J. Vis. Lang. Comput. **25**(6), 667–675 (2014)
4. Lu, C., Yuan, H., Tang, Y.: 1247. Bearing performance degradation assessment and prediction based on EMD and PCA-SOM. J. Vibroengineering 16(3) (2014)
5. Hu, S., Chen, Z., Nia, V.P., Chan, L., Geng, Y.: Causal inference and mechanism clustering of a mixture of additive noise models. arXiv:1809.08568 (2018)
6. Kang, Z., Peng, C. and Cheng, Q.: Top-n recommender system via matrix completion. In: AAAI, pp. 179–185 (2016)
7. Zhao, Z.L., Huang, L., Wang, C.D., Lai, J.H. and Philip, S.Y.: Low-rank and sparse matrix completion for recommendation. In: International Conference on Neural Information Processing, pp. 3–13. Springer, Cham (2017)
8. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **8**, 30–37 (2009)
9. Zhou, Y., Wilkinson, D., Schreiber, R., Pan, R.: Large-scale parallel collaborative filtering for the netflix prize. In: International Conference on Algorithmic Applications in Management, pp. 337–348. Springer, Berlin, Heidelberg (2008)
10. Mnih, A., Salakhutdinov, R.R. (2008) Probabilistic matrix factorization. In: Advances in Neural Information Processing Systems, pp. 1257–1264 (2008)
11. Victor, P., Cornelis, C., Teredesai, A.M. and De Cock, M.: Whom should I trust? The impact of key figures on cold start recommendations. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 2014–2018. ACM (2008)
12. Ma, H., King, I., Lyu, M.R.: Learning to recommend with explicit and implicit social relations. ACM Trans. Intell. Syst. Technol. (TIST) **2**(3), 29 (2011)
13. Zhang, Z.-K., Liu, C., Zhang, Y.-C., Zhou, T.: Solving the cold-start problem in recommender systems with social tags. EPL (Europhys. Lett.) **92**(2), 28002 (2010)
14. Yingyuan, X., Pengqiang, A., Ching-Hsien, H., Hongya, W., Jiao, X.: Time-ordered collaborative filtering for news recommendation. China Commun. **12**(12), 53–62 (2015)
15. Koren, Y.: Collaborative filtering with temporal dynamics. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 447–456. ACM (2009)
16. Wang, H., Wang, N., Yeung, D.-Y.: Collaborative deep learning for recommender systems. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1235–1244. ACM (2015)

Realization of Efficient Architectures for Digital Filters: A Survey

Prashant Kumar (ORCID)**, Prabhat Chandra Shrivastava, Manish Tiwari and Amit Dhawan**

Abstract Digital filters are considered as one of the most important parts of signal processing. Due to lots of advancements in VLSI designs, the dedicated hardware design for the 2-D digital filters snatched the attention of researchers. This paper briefly discusses the challenges of 2-D digital filter's architecture designs and presents the review on suggested methodologies and efficient architectures for 2-D digital filters. A discussion on architectures for separable 2-D digital filters is also presented. In many applications, the frequency response of digital filters poses certain symmetries and due to that, the dedicated architectures require a lesser number of multipliers. So, the suggested architectures for these filters are discussed in this paper. Furthermore, a short review of architectures for 2-D block processing is also given. Moreover, this paper also presents the comparative analysis of hardware and time complexities of all suggested architectures..

Keywords 2-D digital filter · Systolic architecture · Symmetries in digital filters · Block filtering

1 Introduction

Two-dimensional (2-D) digital filter plays a vital role in signal processing. It finds its applications in many image and video processing areas including image enhancement [1], image restoration [2], face recognition [3], beamforming [3], seismic data processing [4], satellite communication [4], digital video processing [5], etc. Further, it can be categorized into two parts, i.e., finite impulse response (FIR) digital filter and infinite impulse response (IIR) filter. IIR filter has high computational efficiency. It is always preferred over the FIR digital filter when magnitude response is the main concern. FIR filter has the advantage of simplicity in design and numerical stability. It is widely used when the applications require preserving the linear phase.

P. Kumar (✉) · P. C. Shrivastava · M. Tiwari · A. Dhawan
Motilal Nehru National Institute of Technology, Allahabad, UP, India
e-mail: pkumar.mnnit@gmail.com

These 2-D IIR and 2-D FIR filters can be processed using the general-purpose computer but it may not suitable for processing the large data for real-time applications. Real-time applications require a high-throughput system to process the large data [6]. In the past few decades, lots of advancements take place in the area of VLSI design. So, the dedicated architecture for digital filter design snatched the attention of researchers. For the dedicated architectures, high-throughput, compactness in size, and low-power consumption are the main characteristics. In the recent past, many efficient architectures [7–29] for 2-D digital filters have been suggested by the researchers. Zhang [9] and Venetsanopoulos [10] have drawn the systolic architecture for 2-D digital filter directly from transfer function. In these architectures, authors have not mentioned the input scanning methodologies and also the suggested architectures have very long critical path. Further, Sid-Ahmed in [11] has suggested the systolic architecture for the 2-D IIR filter. The suggested architecture is designed for the raster scanning of inputs. The suggested architecture is a mixture of direct form and transposed form systolic architecture. It requires a large number of delay elements. The number of delay elements requirement in 2-D IIR filter has reduced by Sunder [12]. Later, Shanbhag has suggested the systolic architecture that has zero latency in output computation in [13]. Similar to [13], architecture for 2-D FIR filter can be also drawn in the same manner. From [14–16, 30, 31], it can be observed that global broadcast path in architecture affects the time performance at the circuit level in VLSI design. Global broadcast path in the 2-D digital filter is minimized by the Van [14–16]. Van [15, 16] has also shown that the proposed architectures have lesser storage error than the earlier reported architecture. Further, Khoo [17] has also proposed the new systolic architecture for the 2-D digital filter with the local broadcast path. The proposed architecture requires a lesser number of delay elements than [16]. In [18], a different framework has been suggested by Khoo to reduce the global broadcast path in 2-D digital filters.

Separable architecture for 2-D digital filters are less sensitive toward the quantization errors and in dedicated hardware design, it requires a lesser number of multipliers, storage elements, and adders. Mohanty [19] has suggested the separable architecture for the 2-D FIR filter. In this suggested architecture, they have removed intermediate transposition which is required in the conventional design. It also reduces the latency in output computation. Two versions of separable architectures namely, Type 1 and Type 2 for 2-D IIR filter are suggested by Khoo [17]. In both suggested architectures, they have minimized the global broadcast path. Type 2 version of architecture has lesser critical path delay than the Type-1.

In many filtering applications, 2-D digital filters pose certain kind of symmetries in frequency response, which in result comes with some relationship among the filter coefficients. Reddy [20] has discussed the symmetries in 2-D digital filters in detail. Based on Reddy's discussions, Chen [21] has suggested the systolic architecture for 2-D IIR digital filter and 2-D FIR digital filter for diagonal symmetry. Further, Khoo [17] and Chen [21–24] have suggested the efficient architectures for separable 2-D IIR filter for diagonal symmetry, quadrantal symmetry, fourfold symmetry, and octagonal symmetry. In all these suggested architecture [17, 23, 24], they have applied the symmetries on the separable architectures of 2-D IIR filter, since the separable

structure requires a lesser number of multipliers and due to the symmetries, it can be reduced further.

To meet the high-speed demands, block input processing is a frequently used method in a general computer-based signal processing. The same can be done with dedicated hardware design but in dedicated hardware design, hardware complexities increase linearly with block inputs. Kawan [25] has suggested the systolic architecture block input processing for the 2-D IIR filter and Mohanty [26] has suggested the cell-level systolic architecture for 2-D block FIR filters. Further, Mohanty in [27] has explored the memory sharing in block filtering filter that reduces the number of delay elements in the architecture. In [27], they have also suggested the architecture for separable 2-D block FIR filtering. Recently, Kumar [28] has suggested the improved architectures for the 2-D block FIR filter which further reduces the number of delay elements as well as the delay in critical path.

Organization of this paper is as follows. In the next section, we have briefly discussed the suggested architectures of 2-D FIR and 2-D IIR filter. A brief study of separable 2-D digital filters and their architectures are given in Sect. 3. Section 4 discusses the symmetry in 2-D digital filters. A review on 2-D block digital filters is given in Sect. 5.

2 2-D Digital Filters

Input–output relation for 2-D IIR filter can be given as

$$y(m, n) = \sum_{i=0}^{N} \sum_{j=0}^{N} x(m-i, n-j) \, a_{ij} + \sum_{i=0}^{N} \sum_{\substack{j=0 \\ i+j \neq 0}}^{N} y(m-i, n-j) \, b_{ij}, \quad (1)$$

and for 2-D FIR filter, output expression for 2-D FIR filter can be given as

$$y(m, n) = \sum_{i=0}^{N} \sum_{j=0}^{N} x(m-i, n-j) \, h_{ij}. \quad (2)$$

where $N \, X \, N$ is the order of the filter, a_{ij} and b_{ij} is the filter coefficient of 2-D IIR filter, h_{ij} is the filter coefficients of 2-D FIR filter, and m and n are the signal representation into the two independent directions. From (1) and (2), it can be observed that the mathematical representations for both equations are similar. From the observation point of view, the output's equation of 2-D IIR filter can be said as the combination of two 2-D FIR filter expressions. Therefore, the realization of structures for 2-D IIR and 2-D FIR filter may follow similar patterns. Hence, the architecture realizations for both digital filters are further discussed parallelly. Z-transform for (1) and (2) can be expressed as

$$Y_{fir}(z_1, z_2) = \sum_{i=0}^{N} \sum_{j=0}^{N} X(z_1, z_2) \, h_{ij} \, z_1^{-i} z_2^{-j} \tag{3a}$$

and

$$Y_{iir}(z_1, z_2) = \sum_{i=0}^{N} \sum_{j=0}^{N} a_{ij} X(z_1, z_2) z_2^{-j} z_1^{-i} + \sum_{i=0}^{N} \sum_{\substack{j=0 \\ i+j \neq 0}}^{N} b_{ij} Y_{iir}(z_1, z_2) z_2^{-j} z_1^{-i}, \tag{3b}$$

respectively. To design the systolic form architecture for 2-D FIR filter, Eq. (3a) can be expressed in canonical form as

$$Y_{fir}(z_1, z_2) = ((.\ldots.(\sum_{j=0}^{N} X(z_1, z_2) z_2^{-j} h_{N-1j} z_1^{-i} + \sum_{j=0}^{N} X(z_1, z_2) z_2^{-j} h_{N-2j}) z_1^{-i}$$

$$+ \sum_{j=0}^{N} X(z_1, z_2) z_2^{-j} h_{N-3j}) z_1^{-i} + \cdots + \sum_{j=0}^{N} X(z_1, z_2) z_2^{-j} h_{1j}) z_1^{-i}$$

$$+ \sum_{j=0}^{N} \underbrace{X(z_1, z_2) z_2^{-j}}_{U^j(z)} h_{0j}. \tag{4}$$

Equation (4) can be mapped into systolic architecture by delaying the output of each internal expression in systolic manner with respect to z_1 and from (4), it can be also observed that the sub-expression $U^j(z)$ is common in each sub-expression. Hence, in architecture realization, delay elements with respect to z_2 will be common for each internal expression of (4). Based on this concept, an efficient architecture for 2-D FIR filter has been suggested by Zhang in [9]. The suggested architecture is redrawn in Fig. 1a. In the suggested architecture, systolic arrangement of delay elements D_1 gives the delay in m direction and D_2 in n direction. In a similar manner, architecture for 2-D IIR filter can be designed as depicted in Fig. 1b. The Proposed architecture has a longer critical path. Further, Venetsanopouls [10] has suggested the fully transposed form architecture for 2-D IIR filter. The suggested method reduces critical path delay. In both proposed designs [9, 10], authors have not discussed the input scanning methodologies. In software-based filtering operation, it is very convenient to access the required input samples but in hardware realization, we need to follow the set of patterns for input samples scanning. For an illustration, a M_1 x M_2 size image is shown in Fig. 2. Here, we can access the image samples sequentially in row-wise (horizontal scanning, i.e., raster scanning) or column-wise (Vertical scanning) manner. In raster scanning, input samples appear as $\{x(0, 0), x(0, 1), x(0, 2) \ldots x(0, M_1 - 1), x(1, 0), x(1, 1) \ldots x(1, M_1 - 1) \ldots\}$ and in vertical scanning, input samples appear as $\{x(0, 0), x(1, 0), x(2, 0) \ldots x(M_2 -$
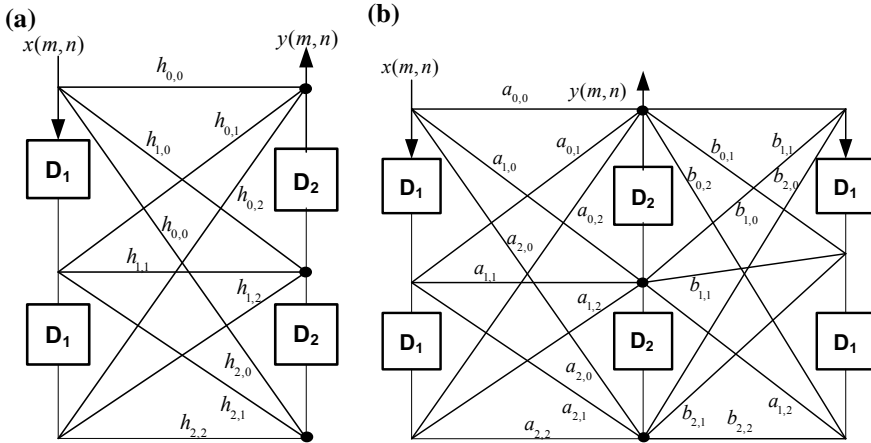
Fig. 1 Architectures for 2-D digital filter suggested in [9]. (**a**) FIR (**b**) IIR

$$x(M_2-1,0)\ x(M_2-1,1)\text{.....................}x(M_2-1,M_1-1)$$

$$\text{..}$$

$$\text{..}$$

$$x(3,0)\ x(3,1)\ x(3,2)\ x(3,3)\,x(3,4)\ x(3,5)\text{.....}x(0,M_1-1)$$

$$x(2,0)\ x(2,1)\ x(2,2)\ x(2,3)\,x(2,4)\ x(2,5)\text{.....}x(0,M_1-1)$$

$$x(1,0)\ x(1,1)\ x(1,2)\ x(1,3)\ x(1,4)\ x(1,5)\text{........}x(0,M_1-1)$$

$$x(0,0)\ x(0,1)\ x(0,2)\ x(0,3)\,x(0,4)\ x(0,5)\text{......}x(0,M_1-1)$$
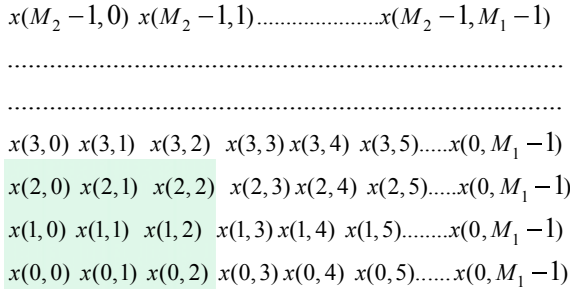
Fig. 2 Input image pixels of size $M_1 X M_2$

1, 1), $x(0, 1)$, $x(1, 1) \ldots x(M_2 - 1, 1) \ldots$}, where $M_1 X M_2$ is the size of the image. Now, if we have to process the current input samples $x(2, 2)$, then for the filtering operation, shaded samples are required as depicted in Fig. 2. But the architectures of [9, 10] do not provide the required samples with such scanning. Input scanning problem has been resolved by Sid-Ahmed in [11]. In the suggested architecture, they have added the $M_1 - 1$ length shift register at each systolic stage. Therefore, z_1^{-1} in (4) can be represented as z^{-M_1} and (4) can be modified accordingly as

$$Y_{fir}(z_1, z_2) = X(z_1, z_2)(\sum_{j=0}^{N} z_2^{-j} h_{0j} + z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} h_{1j} + \ldots \ldots$$

$$+ z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} h_{N-2j} + z^{-M_1} \sum_{j=0}^{N} z_2^{-j} h_{N-1j})) \ldots .)), \qquad (5)$$

and also 2-D IIR filter expression (3b) for raster scanning can be expressed as

$$Y_{iir}(z_1, z_2) = X(z_1, z_2)(\sum_{j=0}^{N} z_2^{-j} a_{0j} + z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} a_{1j} + \cdots$$

$$+ z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} a_{N-2j} + z^{-M_1} \sum_{j=0}^{N} z_2^{-j} a_{N-1j})) \ldots)$$

$$+ Y_{iir}(z_1, z_2)(\sum_{j=1}^{N} z_2^{-j} b_{0j} + z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} b_{1j} + \cdots$$

$$+ z^{-M_1}(\sum_{j=0}^{N} z_2^{-j} b_{N-2j} + z^{-M_1} \sum_{j=0}^{N} z_2^{-j} b_{N-1j})) \ldots)). \quad (6)$$

Based on the systolic arrangement of (6), Sid-Ahmed [11] has realized the architecture for 2-D IIR filter as depicted in Fig. 3 (In Fig., image width $M_1 = M$). In the given architecture, with respect to z_1, input signals are delayed in a systolic manner by direct form architecture and with respect to z_2, transposed form architecture is used for the systolic delay. At the input stage, M_1-length shift register block (SRB) is used at each systolic level for getting the desired delayed-input signals. One-stage pipelining is also used in the architecture to reduce the critical path delay. The suggested structure for (6) by [11] requires large numbers of register due to direct form representation with respect to z_1. Now, for reducing the delay elements, (6) can be again rearranged and expressed as

$$Y = ((\ldots(\sum_{j=0}^{N} (X a_{N-1j} + Y b_{N-1j})z_2^{-j})z^{-M_1} + \sum_{j=0}^{N} (X a_{N-2j} + Y b_{N-2j})z_2^{-j})z^{-M_1}$$
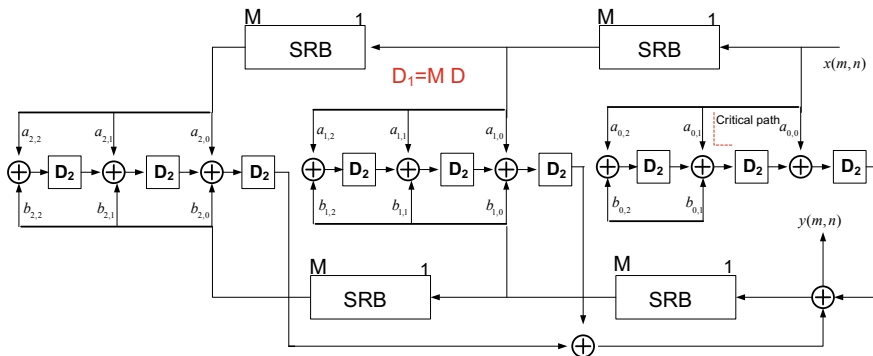
Fig. 3 Architecture for 2-D IIR filter suggested by Sid-Ahmed [11]

$$+ \ldots\ldots + (\sum_{j=0}^{N} (X\, a_{N-1\,j} + Y\, b_{N-1\,j})z_2^{-j})z^{-M_1} + \sum_{j=0}^{N} (X\, a_{0j} + Y\, b_{0j})z_2^{-j},$$

$$(7)$$

where $X = X(z_1, z_2)$ and $Y = Y_{iir}(z_1, z_2)$. Based on the systolic arrangement of (7), Sunder [12] has suggested the new systolic arrangement for 2-D IIR filter. The signal flow graph (SFG) of systolic arrangement with respect to z_2 is similar to the architecture of Sid-Ahmed [11], but with respect to z_1, systolization is achieved by using hybrid of direct form and transposed form systolic arrangement. M_1 length SRB at each systolic level is subdivided into 1 and M–1 length shift registers. One shift register is used at each systolic level in the direct form systolization and $M_1 - 1$ length SRB at systolic stage in the transposed form. The proposed architecture is redrawn and depicted in Fig. 4. It reduces $(N - 1)M_1$ the number of delay elements in realization. It also uses the one-stage pipelining similar to [11] so, critical path delay for both architectures are same. Further, Shanbhag in [13] has suggested the improved systolic architecture for 2-D IIR and 2-D FIR filter. The suggested architecture for 2-D IIR is redrawn and depicted in Fig. 5. In this architecture, with respect to z_2 systolization is achieved by the parallel and systolic transposed form architecture. In each PU (Fig. 5), after each D_2 delay in transposed path two parallel delay elements in input path are used to get the desired delay with respect to z_2 and fully transposed form systolic architecture is used to getting the desired delay with respect to z_1. The suggested architecture has zero latency and also requires the lesser number of registers than earlier reported architecture. In the reported paper, they have also shown the suggested architecture has the lesser storage error (quantization and overflow errors due to fixed width of storage elements) than the earlier reported architecture.

It has been observed in the VLSI design that global broadcast has the significant impact upon speed in the circuit level [14–18, 30, 31]. It can be noticed that input and output of the architectures suggested in [9–13, 30] are globally broadcast. So here,
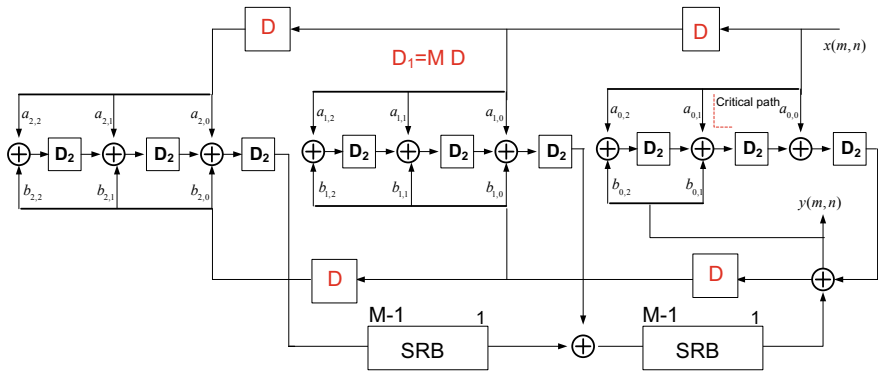
Fig. 4 Architecture for 2-D IIR filter suggested by Sunder [11]
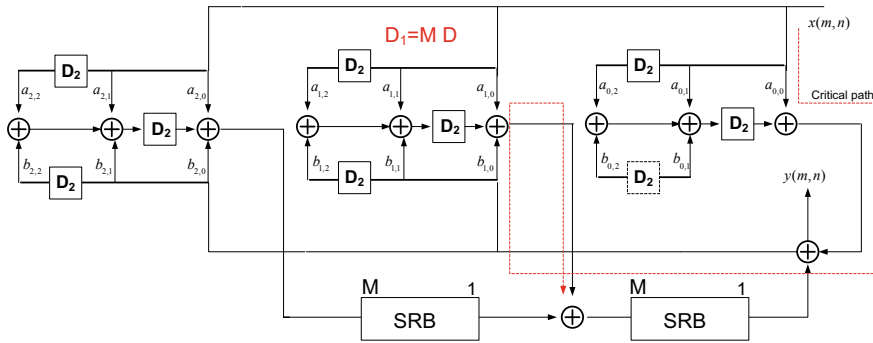
Fig. 5 Architecture for 2-D IIR filter suggested by Shanbhag [13]

local broadcast path architecture will be more advantageous than the architectures of [9–13, 30]. To achieve the local broadcast (7) can be further expressed as,

$$
Y =
$$
$$
((\ldots (\sum_{j=0}^{N} (z^{-(N-1)p} X a_{N-1j} + z^{-(N-1)p} Y b_{N-1j}) z_2^{-j}) z^{-M_1-p} + \sum_{j=0}^{N} (z^{-(N-2)p} X a_{N-2j}
$$
$$
+ z^{-(N-2)p} Y b_{N-2j}) z_2^{-j}) z^{-M_1-p} + \cdots + (\sum_{j=0}^{N} (z^{-p} X a_{N-1j} + z^{-p} Y b_{N-1j}) z_2^{-j}) z^{-M_1}
$$
$$
+ \sum_{j=0}^{N} (X a_{0j} + Y b_{0j}) z_2^{-j} \tag{8}
$$

In (7), z^{-M_1} is expressed in the canonical form to mapped the equation in transposed form systolization. In (8), at each canonical stage, M_1 is subdivided into p and M_{1-P}, where p is a arbitrary number and z^{-M_1-p} is expressed in same canonical manner as given in (7). Further, z^{-p} can be also expressed in canonical form so in realization, input can be delayed by systolic direct form. So, (8) can be further expressed as,

$$
Y = (z^{-p} (\ldots (z^{-p} ((z^{-p} \underbrace{(\sum_{j=0}^{N} (X a_{N-1j} + Y b_{N-1j}) z_2^{-j}))}_{S} z^{-M_1-p} + \sum_{j=0}^{N} (z^{-(N-2)p} X a_{N-2j}
$$
$$
+ z^{-(N-2)p} Y b_{N-2j}) z_2^{-j})) z^{-M_1-p} + \cdots + (\sum_{j=0}^{N} (z^{-p} X a_{N-1j} + z^p Y b_{N-1j}) z_2^{-j})) z^{-M_1}
$$
$$
+ \sum_{j=0}^{N} (X a_{0j} + Y b_{0j}) z_2^{-j}, \tag{9a}
$$
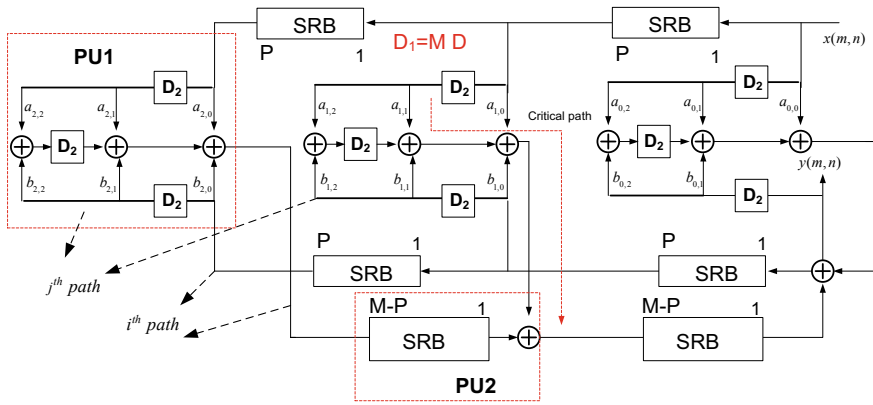
Fig. 6 Architecture for 2-D IIR filter suggested by Van [16]

and sub-expression (S) of (9a, 9b) can be further expressed as

$$S = ((..(z_2^{-1} X\, a_{i\,N-1} + z_2^{-1} Y\, b_{i\,N-1} + X\, a_{i\,N-2} + Y\, b_{i\,N-2})z_2^{-1} +$$
$$(z_2^{-1} X\, a_{i\,N-3} + z_2^{-1} Y\, b_{i\,N-3} + X\, a_{i\,N-4} + Y\, b_{i\,N-4})z_2^{-1} +$$
$$\cdots + (z_2^{-1} X\, a_{i1} + z_2^{-1} Y\, b_{i1} + X\, a_{i0} + Y\, b_{i0}). \tag{9b}$$

Based on (9a, 9b), Van has suggested systolic architectures for 2-D IIR filter in [15, 16]. In [15, 16], with respect to z_1, desired delay is realized by hybrid form of direct form and transposed form architecture and with respect to z_2 desired delay is realized by the hybrid of parallel and systolic arrangement of delay elements as per (9b). Such proposed SFG reduces the global broadcast path in both ith as well as jth broadcast path (Fig. 6). Shanbhag [13] has also used some kind of parallel and systolic architecture to resolve the jth global broadcast path in the architecture but it sacrifices the critical path period. The proposed architecture by Van has resolved the global broadcast in jth-path dimension as well as ith-path without sacrificing the critical path period. In the reported paper, they have also shown that the proposed design has lowest storage error than the architecture of [11–13].

Van has further suggested the architecture for the cascaded form digital filter for the higher order filter. It has been observed from [16] that the cascaded form architectures are less sensitive to the quantization error than the non-cascaded form architecture for the fixed-point operation. Let $H(z_1, z_2)$ be the impulse response of 2-D IIR filter than in cascaded form of second-order filter it can be expressed as
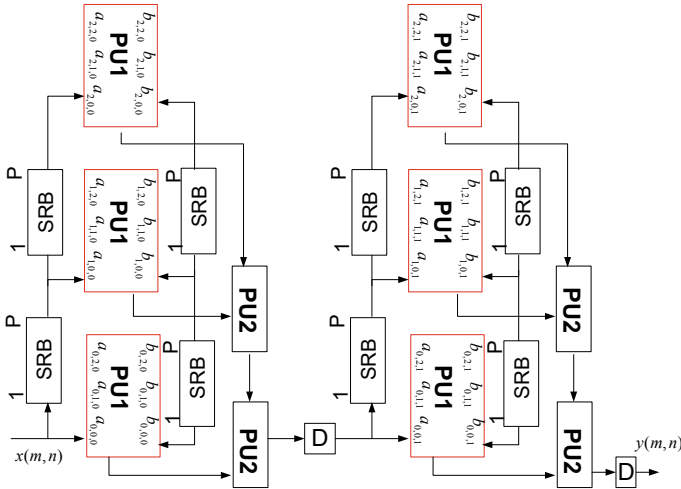
Fig. 7 Cascaded form architecture suggested by Van [16] for higher order filter

$$H_{iir}(z_1, z_2) = \frac{Y(z_2, z_1)}{X(z_2, z_1)} = \prod_{k=1}^{N/2} H_k(z_1, z_2) = \prod_{k=1}^{N/2} \frac{\sum_{i=0}^{2} \sum_{j=0}^{2} a_{i,j,k} z_2^{-j} z_1^{-i}}{1 - \sum_{i=0}^{2} \sum_{j=0}^{2} b_{i,j,k} z_2^{-j} z_1^{-i}}.$$

$$b_{00k} = 0$$

$$(10)$$

To map (10) into the cascaded form, van [16] has used the suggested local broadcast path based second-order architecture (Fig. 7). The output of each cascaded second-order filter is pipelined to limit the critical path delay in the overall architecture. The block diagram for cascaded architecture is shown in the Fig. 8. Further, Khoo [17] has suggested another local broadcast path architecture for 2-D IIR filter (Fig. 9). The suggested architecture is shown in Fig. 10. It requires the fewer number of delay elements than [16] and its critical path delay is the same as [16].
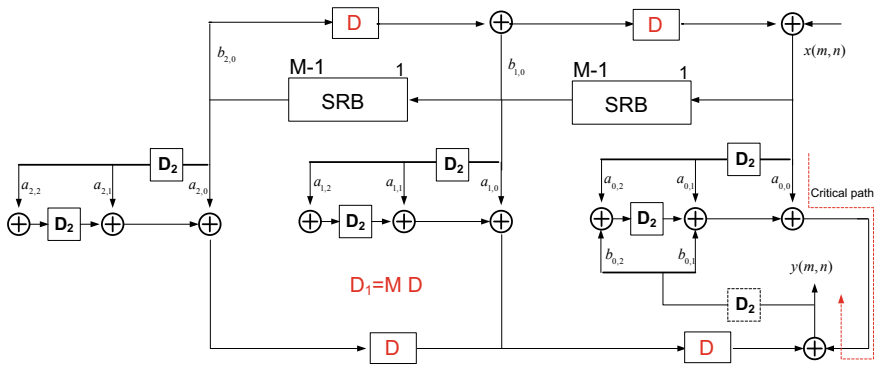
Fig. 8 Architecture suggested by Khoo [17] for separable 2-D IIR filter (Type-1)
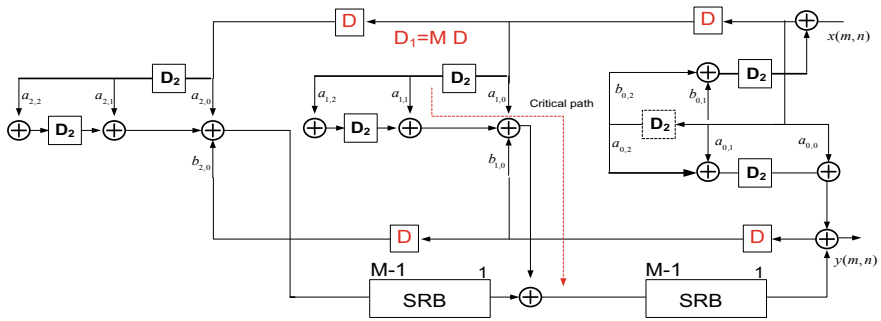
Fig. 9 Architecture suggested by Khoo [17] for separable 2-D IIR filter (Type-2)
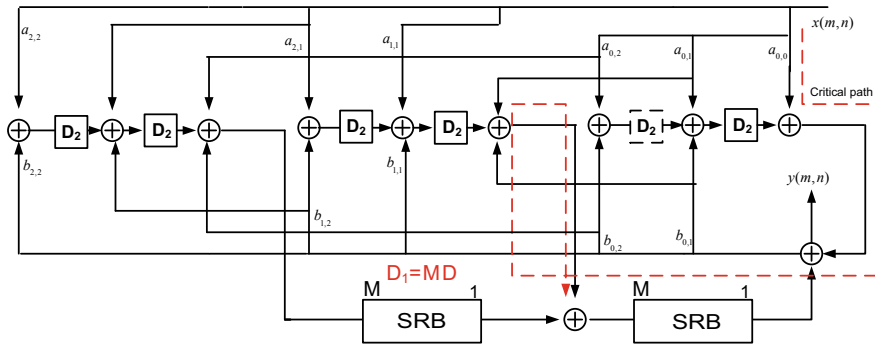
Fig. 10 Architecture suggested by Chen [21] for 2-D IIR filter with diagonal symm.

3 Separable Digital Filters

3.1 Separable Denominator 2-D IIR Filter

A transfer function for separable denominator 2-D IIR filter can be given as,

$$H_{iir}(z_1, z_2) = \frac{Y(z_1, z_2)}{X(z_1, z_2)} = \frac{\sum_{i=0}^{N}\sum_{j=0}^{N} a_{i,j} z_2^{-j} z_1^{-i}}{(1 - \sum_{i=0}^{N} b_{i,0} z_1^{-i})(1 - \sum_{j=0}^{N} b_{0,j} z_2^{-j})} \quad (11)$$

where $b_{0,j} = 0$. Such separable denominator 2-D IIR filter has lots of advantage over the non-separable 2-D IIR filter. Since the denominator is expressed by two 1-D poles so, this will be advantageous in stability checking and also any unstable pole can be replaced easily by their inverse pole without affecting the magnitude response of the filter. Due to the separable denominator, multiplications require for filtering also decreases. Therefore, in realization, fewer numbers of multipliers are required. Now, for the realization of architecture for 2-D IIR filter, (11) can be further expressed as,

$$Y_{iir}(z_1, z_2) = (\underbrace{\frac{X(z_1, z_2)}{(1 - \sum_{i=0}^{N} b_{i,0} z_1^{-i})}}_{Y_1(z_1, z_2)})(\frac{\sum_{i=0}^{N}\sum_{j=0}^{N} a_{i,j} z_2^{-j} z_1^{-i}}{(1 - \sum_{j=0}^{N} b_{0,j} z_2^{-j})}) \quad (12)$$

and
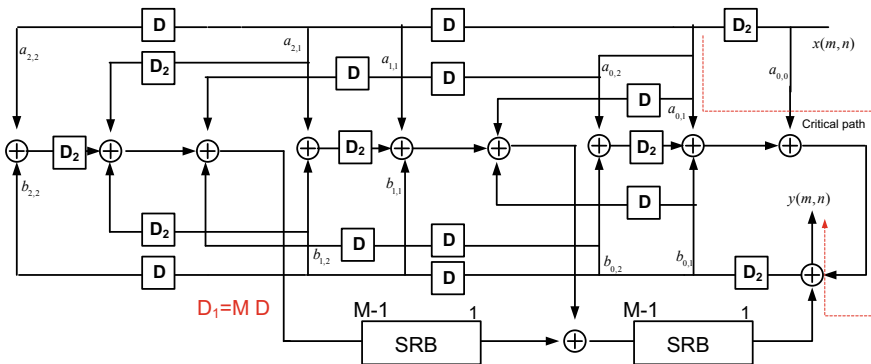
Fig. 11 Architecture suggested by Chen [21] for 2-D IIR filter with diagonal symmetry with local broadcast path

$$Y_1(z_1, z_2) = X(z_1, z_2) + \sum_{i=0}^{N} Y(z_1, z_2) b_{i,0} z_1^{-i} \tag{13}$$

Using (13), (12) can be expressed as,

$$Y_{iir}(z_1, z_2) = \frac{\sum_{i=0}^{N} \sum_{j=0}^{N} Y_1(z_1, z_2) a_{i,j} z_2^{-j} z_1^{-i}}{(1 - \sum_{j=0}^{N} b_{0,j} z_2^{-j})} \tag{14}$$

From (12), (13) and (14), it can be observed that for structure realization for (12), Firstly, we need to design the architecture for (13) i.e. $Y_1(z_1, z_2)$ and $Y_1(z_1, z_2)$ will be input for architecture of (14). Based on (12), (13) and (14) expressions, Khoo [17] has suggested two different architectures (Type-1 and Type-2) for separable denominator 2-D IIR filter. The Type-1 (Fig. 8) architecture requires $N^2 + 2N$ numbers of multipliers and its critical path delay is $T_m + 3T_a$. To reduce the critical path delay, they have again rearranged the (12) as,

$$Y_{iir}(z_1, z_2) = (\underbrace{\frac{X(z_1, z_2)}{(1 - \sum_{j=0}^{N} b_{0,j} z_2^{-j})}}_{Y_2(z_1, z_2)})(\frac{\sum_{i=0}^{N} \sum_{j=0}^{N} a_{i,j} z_2^{-j} z_1^{-i}}{(1 - \sum_{i=0}^{N} b_{i,0} z_1^{-i})}) \tag{15a}$$

and

$$Y_2(z_1, z_2) = X(z_1, z_2) + \sum_{j=0}^{N} Y(z_1, z_2) b_{0,j} z_2^{-j} \tag{15b}$$

Similarly to the realization of (12), for (15a), firstly Khoo [17] has designed the architecture for $Y_2(z_1, z_2)$ and then output $Y_2(z_1, z_2)$ fed as an input to the further structure as depicted in Fig. 9. Its critical path delay is $T_m + 2T_a$ and rest of complexities are same as the structure of (12). The both proposed designs, sub-expressions are rearranged similarly to (9b) to reduce the global broadcast path in the architecture.

3.2 Separable 2-D FIR Filter

For separable 2-D FIR filter, (2) can be expressed as,

$$y(m, n) = \sum_{i=0}^{N-1} a_{i0} \, v(m - i, n),$$ (17a)

where

$$v(m, n) = \sum_{j=0}^{N-1} a_{0j} \, x(m, n - j).$$ (17b)

From (17a, 17b), it can be observed that for separable 2-D FIR, two 1-D FIR filtering in two stages is required to compute the output. Conventionally, first, we need to scan the input in raster scanning format for M rows and in each row up to M column. After M^2 clock cycle, total MXM outputs for $v(m, n)$ can be computed. In the next stage, we need to take the transposed of MXM matrix so in Stage 2 filtering, $v(m, n)$ can be processed in row-wise manner to compute the final output. This method computed the first output after the latency of $M^2 + 2N$ clock-cycles. Further, Mohanty [19] has suggested the architecture for separable 2-D FIR filter which has lower latency than the conventional architectures. In the proposed design, they have not used the raster scan method for input scanning. At the input side, they have used $(M + N–1)$x$(M + N–1)$ size buffer to store the inputs. Each input is buffered in memory with its N-1 past samples and these samples are fed to the first stage 1-D filtering with each clock. Therefore here, desired transposed outputs can be generated directly at the intermediate stage. It reduces the latency in outputs computation but it requires memory elements in huge numbers.

4 Symmetries in 2-D Digital Filters

There are several applications of 2-D digital filters [20] where their frequency response exhibits certain symmetries. Symmetry in frequency response induces the certain relationship among the filter coefficients of digital filters which in result reduces the requirement of multipliers in the realization of these filters. If there is no symmetry in the frequency response of the digital filter then there is a method by which frequency response of digital filter can be decomposed into the desired symmetry [20]. Due to that, in last two decades, many efficient architectures [17, 21–24] for 2-D digital filters have been suggested by the researchers. Symmetry in the 2-D digital filter can be categorized into diagonal symmetry, fourfold symmetry, quadrantal symmetry, and octagonal symmetry. The transfer function for 2-D IIR filter can be expressed as
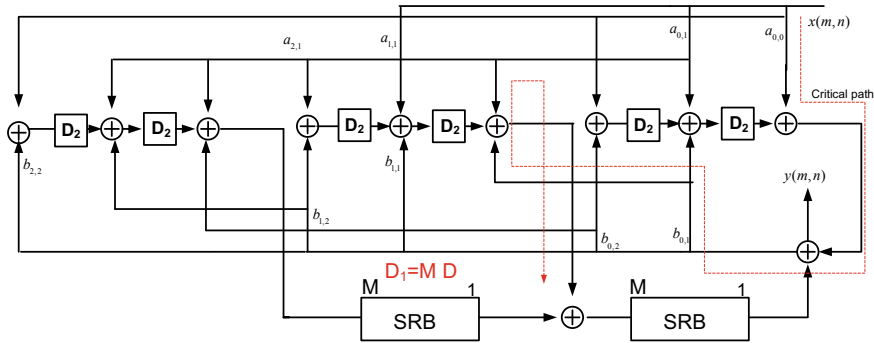
Fig. 12 Architecture suggested by Chen [21] for 2-D IIR filter with fourfold symmetry

$$H_{iir}(z_1, z_2) = \frac{N(z_1, z_2)}{D(z_1, z_2)}, \qquad (18)$$

where $N(z_1, z_2)$ denotes the numerator function and $D(z_1, z_2)$ denotes denominator function. The magnitude response of 2-D digital filter poses the diagonal symmetry if $N(z_1, z_2) = N(z_2, z_1)$ and $D(z_1, z_2) = D(z_2, z_1)$ therefore, the relation among filter coefficients can be given as $a_{i,j} = a_{j,i}$ and $b_{i,j} = b_{j,i}$. For fourfold symmetry, this relationship can be given as $N(z_1, z_2) = N(z_2^{-1}, z_1)z_2^{-N}$ and $D(z_1, z_2) = D(z_2, z_1)$. Consequently, the magnitude response poses such relations only if $a_{i,j} = a_{j,(N-i)}$ and $b_{i,j} = b_{j,i}$. For FIR filter, relation for diagonal symmetry can be given as $H_f(z_1, z_2) = H_f(z_2, z_1)$ so, $h_{i,j} = h_{j,i}$ and for fourfold symmetry such relation can be given as $H_{fir}(z_1, z_2) = H_{fir}(z_2^{-1}, z_1)z_2^{-N}$ hence, $h_{i,j} = h_{j,(N-i)}$ where H_{fir} the transfer function of 2-D FIR filter and h_{ij} is the filter coefficient. Chen [21] and Chen [22] have suggested the architecture for the diagonal symmetry and fourfold symmetry as the architectures are redrawn in Figs. 10 and 12, respectively. In both designs, they have used the fully transposed form architecture and common filter coefficients are shared the same multipliers. They have also suggested the architecture with local broadcast path as shown in Fig. 11. Later, Khoo [17] has suggested 2-D IIR and 2-D FIR filter with diagonal symmetry. For designing the architecture for IIR filter, they have used the separable architecture (Fig. 9). Due to that the suggested architecture requires lesser multipliers and has shorter critical path. Further, based on Khoo's Type-1 (Fig. 10) and Type-2 (Fig. 11) architectures, Chen [23] has suggested the architectures for 2-D IIR filter with diagonal symmetry, Fourfold symmetry, quadrantal symmetry, octagonal symmetry. Since the architectures are based on Khoo's designs so, the suggested architectures do not contain any global broadcast path.

5   2-D Block Digital Filter

In computer-based filtering operations, block processing is a well known approach for achieving high computational speed. It can be also done in the VLSI design but it increases the area requirement and power consumption in large amount. Initially, Kawan [26] has suggested the architecture for 2-D block IIR filter. In the suggested design, hardware complexities increase with block inputs. Recently, Mohanty [27] has suggested the architecture for non-separable (Figs. 13 and 14) and separable 2-D block FIR filter. In the proposed design, they have suggested the memory sharing approach to reduce the registers requirements. By introducing the memory overlapping concept, Kumar [28] has further reduced the memory requirements in the architecture. Kumar has also suggested a method by which shorter critical path can be achieved (Fig. 15).
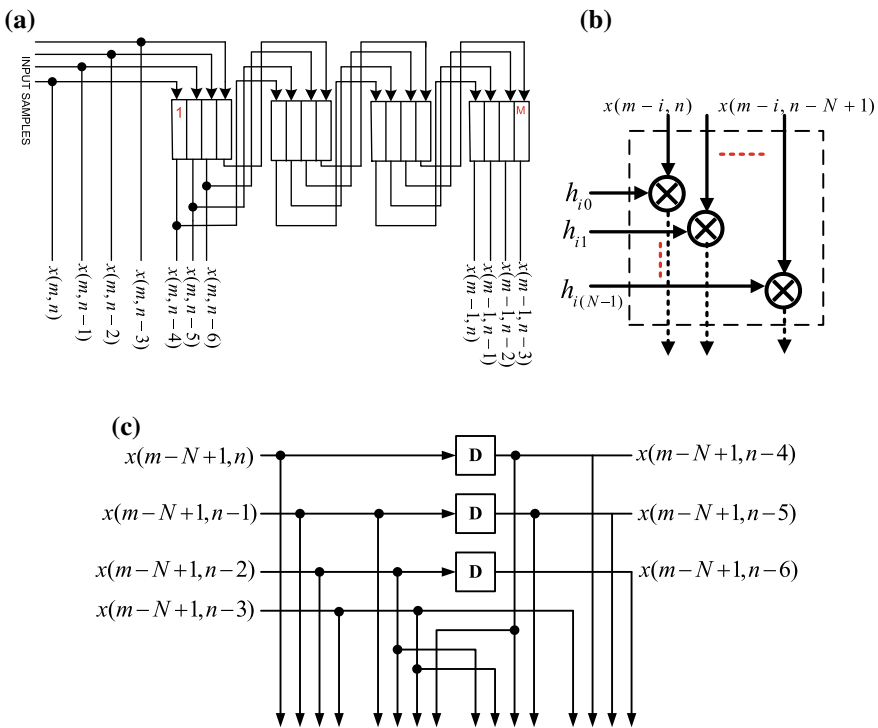
Fig. 13 a Shift registers unit (SRU) structure for block processing **b** Inner product (IP) block. **c** Register Unit (RU) for generating the required delayed-inputs
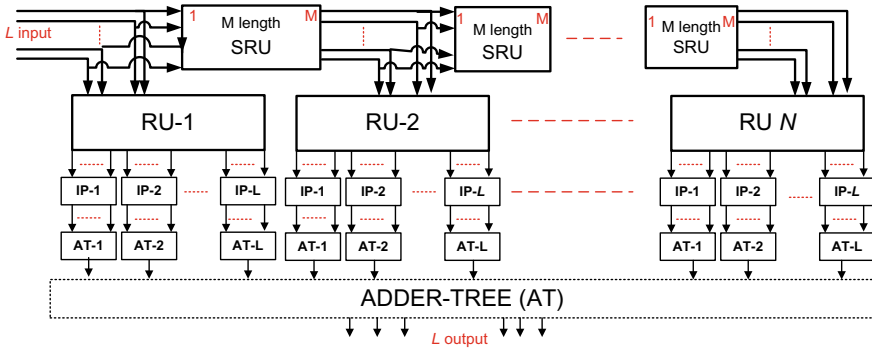
Fig. 14 Architecture for 2-D block FIR filter suggested by Mohanty [27]
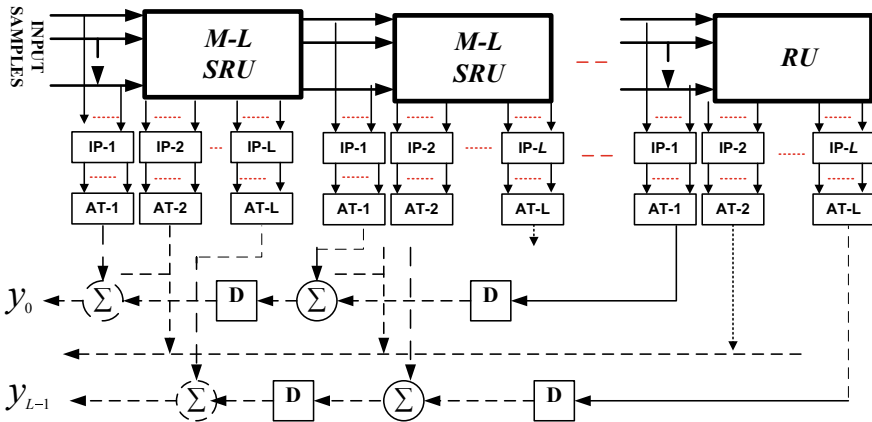
Fig. 15 Modified architecture for 2-D block FIR filter suggested by Kumar [28]

6 Complexities Comparisons

Hardware complexities and Time complexities of all the discussed architectures are listed in Table 1. In Table 2, we have listed the complexities of 2-D IIR filter for filter order $N = 3$. From Table 2, it can be observed that the non-separable architectures [11–16, 31] require 31 multipliers and the separable architectures [17, 18] require only 22 multipliers. Multipliers in 2-D digital filters are further reduced with symmetries as listed in Table 2. Architectures for diagonal symmetry suggested in [20] and for fourfold symmetry suggested in [21] require only 19 and 16 multipliers, respectively. Furthermore, the multipliers requirement can be reduced by realizing the separable digital filter with symmetry. It can be observed from Table 2, the separable architecture with octagonal symmetry [22] requires only 9 multipliers, which is the lowest among all design for $N = 3$. Moreover, it can be also observed that the local broadcast path containing architectures [14–24] require the greater number

Table 1 Hardware complexities and time complexities

Design	Number of multipliers	Number of registers	Critical path delay	Latency	Global broadcast
Sid-Ahmed [11]	$2(N+1)^2 - 1$	$(N+1)^2 + 2MN$	$T_m + (2 + \lfloor \log_2(N+1) \rfloor) T_a$	1	YES
Sunder [12]	$2(N+1)^2 - 1$	$(N+1)^2 + (M+1)N$	$T_m + 2T_a$	1	YES
Shanbhag [13]	$2(N+1)^2 - 1$	$1.5N(N+1) + MN$	$2T_m + 2T_a$	0	Partially reduced
Van [16]	$2(N+1)^2 - 1$	$1.5N(N+1) + (M+1)N$	$T_m + 3T_a$	0	NO
Khoo [17] (Non-separable)	$2(N+1)^2 - 1$	$N(N+1) + 1 + [(N-1)/7] + \lfloor N/7 \rfloor N + N(M+1)$	$T_m + 3T_a$	0	NO
Separable (Type-1)	$(N+1)^2 + 2N$	$N^2 + [0.5(3N+1)] + N(M+1)$	$T_m + 2T_a$	0	NO
Separable (Type-2)	$(N+1)^2 + 2N$	$N^2 + N + 1 + \lfloor (N-1)/3 \rfloor + N(M+1)$	$T_m + 2T_a$	0	NO
2-D FIR filter					
Sid-Ahmed [11]	$(N+1)^2$	$(N+1)^2 + MN$	$T_m + T_a$	1	YES
Sunder [12]	$(N+1)^2$	$(N+1)^2 + MN$	$T_m + T_a$	1	YES
Shanbhag [13]	$(N+1)^2$	$(N+1)^2 + MN$	$T_m + 2T_a$	0	YES
Mohanty [26] (Block processing)	$L(N+1)^2$	$L(N+1)(2M - L)$	$T_m + T_a$	1	YES
Mohanty [19] (Separable)	$2N$	$N(M+1) + MN$	$T_m + T_a$	–	YES
Van [16]	$(N+1)^2$	$3(\lfloor N/3 \rfloor + 1)(N+1) + MN$	$T_m + 2T_a$	0	NO
Khoo [18]	$(N+1)^2$	$(M + N + 1)N$	$T_m + 2T_a$	0	NO
Mohanty [27] (Block processing L)	$L(N+1)^2$	$N(M + N)$	$T_m + T_a + 2T_{fa}(\log_2 N)$	0	YES
Mohanty [27] (Separable)	$2LN$	$N(M+1)$	$T_m + T_a + 2T_{fa}(\log_2 N)$	–	YES

(continued)

Table 1 (continued)

Design	Number of multipliers	Number of registers	Critical path delay	Latency	Global broadcast
Symmetry in 2-D IIR filter					
Chen [21] (Non-separable) ($p = 0$) (Diagonal)	$(N+1)^2 + N$	$N(N+1) + NM$	$2T_m + 2T_a$	0	YES
Khoo [16] (Separable) Diagonal	5 N	$(11 + 2M)N/2 + N/2 + 1$	$T_m + 2T_a$	$\frac{N}{2} - 1$	NO
Chen [22] (Non-separable) (Fourfold)	$3(N+1)^2/4 + 0.5(N+1) - 1$	$N(N+1) + MN$	$T_m + 3T_a$	0	YES

Table 2 Complexities of 2-D IIR architectures for N=3

Design		Number of multipliers	Number of registers	Critical path delay
Sid-Ahmed [11]		31	16+6M	$T_m + 2T_a$
Sunder [12]		31	16+3M	$T_m + 2T_a$
Shanbhag [13]		31	19+3M	$2T_m + 2T_a$
Van [16]		31	21+3M	$T_m + 3T_a$
Khoo [17]	Type-1	22	17+3M	$T_m + 3T_a$
	Type-2	22	17+3M	$T_m + 2T_a$
Chen [21] Diagonal		19	12+3M	$2T_m + 2T_a$
Khoo [17] Diagonal		16	18+3M	$T_m + 2T_a$
Chen [22] Fourfold		13	12+3M	$2T_m + 2T_a$
Chen [23]				
Diagonal	Type-1	16	23+3M	$T_m + 3T_a$
	Type-2	16	23+3M	$T_m + 2T_a$
Fourfold	Type-1	10	24+3M	$T_m + 3T_a$
	Type-2	10	26+3M	$T_m + 2T_a$
Quadrantal	Type-1	14	29+3M	$T_m + 3T_a$
	Type-2	14	29+3M	$T_m + 2T_a$
Octagonal	Type-1	9	22+3M	$T_m + 3T_a$
	Type-2	9	21+3M	$T_m + 2T_a$
Chen [24]				
Diagonal		16	24+3M	$T_m + 2T_a$
Fourfold		10	25+3M	$T_m + 2T_a$

of delay elements than global broadcast path containing architectures [9–13]. Chen [22] has suggested the architectures based on architectures in [17], so their critical path delay is similar to [17].

7 Conclusions

In this paper, we have presented a brief review of the design methodologies, challenges, and the suggested different efficient architectures for 2-D digital filters. Along with architectures, their mathematical rearrangements have been also discussed. Further, a discussion has been carried out on the systolic architectures designs, input scanning methodologies and due to this modification in architectures, global and local broadcast, and storage error. Since in many applications of 2-D digital filter,

the magnitude response of 2-D digital filter poses some symmetry and due to this, in architecture design, fewer multipliers are required. Therefore, a short review on symmetries in 2-D digital filters and the suggested efficient architectures for them has been also given in this paper. Furthermore, to achieve the high-throughput in 2-D digital filters, a discussion on architectures for block filtering has been also presented. In the last section of the paper, we have presented the time as well as hardware complexities of all discussed designs in tabular form in Tables 1 and 2. From Tables 1 and 2, it can be observed that the separable 2-D IIR filter architecture with octagonal symmetry has the lowest hardware complexities among all reported architectures, and architecture suggested by Kumar [28] for 2-D block FIR filtering requires lowest delay elements per input and it has also the highest throughput among all discussed architectures.

References

1. Sid-Ahmed, M.A.: Image Processing: Theory, Algorithms, and Architectures. McGraw-Hill, NY (1995)
2. Dudgeon, D.E., Mersereau, R.M.: Multidimensional Digital Signal Processing. Prentice-Hall, Englewood Cliffs, NJ (1984). Chap. 4
3. Joshi, R.M., Madanayake, A., Adikari, J., Bruton, L.T.: Synthesis and array processor realization of a 2-D IIR beam filter for wireless applications. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **20**(12), 2241–2254 (2012)
4. Swamy, M.N.S., Rajan, P.K.: Symmetry in 2-D Filters and Its Application, Multidimensional Systems: Techniques and Applications (S.G. Tzafestas, Ed.). Marcel Dekkar, New York (1986). Chap. 9
5. Wijenayake, C.K., Madanayake, A., Bruton, L.T.: FPGA prototypes of differential-form 2D-IIR systolic-array DSP architectures for multi-beam plane-wave filters In: Proceedings IEEE SIPS, pp. 58–63. San Francisco, CA (2010)
6. Tekalp, A.M.: Digital Video Processing. Prentice-Hall, Englewood Cliffs, NJ (1995). Chap. 14
7. Gnanasekaran, R.: 2-D filter implementations for real-time signal processing. IEEE Trans. Circuits Syst. **35**(5), 587–590 (1988)
8. Mitra, S.K., Sagar, A.D., Pendergrass, N.A.: Realization of two dimensional recursive digital filters. IEEE Trans. Circuits Syst. **22**, 171–184 (1975)
9. Zhang, H., Yuan, B.: A new VLSI structure for 2-D FIR digital filters. In: IEEE 9th Annual Conference on Proceedings of TENCON'94. 'Frontiers of Computer Technology', vol. 1, pp. 534–536. Singapore (1994)
10. Venetsanopoulos, A., Mertzios, B.: High-speed architectures for digital image processing. IEEE Trans. Circuits Syst. **34**(8), 887–896 (1987)
11. Sid-Ahmed, M.A.: A systolic realization for 2-D digital filters. IEEE Trans. Acoust. Speech Signal Process. **ASSP-37**(4), 560–565 (1989)
12. Sunder, S., El-Guibaly, F., Antoniou, A.: Systolic implementations of two-dimensional recursive digital filters. In Proceedings IEEE International Symposium Circuits System, pp. 1034–1037 (1990)
13. Shanbhag, N.R.: An improved systolic architecture for 2-D digital filters. IEEE Trans. Signal Process. **39**, 1195–1202 (1991)
14. Van, L.D., Tang, C.C., Tenqchen, S., Feng, W.S.: A new VLSI architecture without global broadcast for 2-D systolic digital filters. In: IEEE International Symposium Circuits System, vol. 1, pp. 547–550. Geneva, Switzerland (2000)

15. Van, L.D., Tenqchen, S., Chang, C.H., Feng, W.S.: A new 2-D digital filter using a locally broadcast scheme and its cascade form. In: IEEE Asia Pacific Conference on Circuits System, pp. 579–582. Tianjin, China (2000)
16. Van, L.D.: A new 2-D systolic digital filter architecture without global broadcast. IEEE Trans. VLSI Syst. **10**(4), 477–486 (2002)
17. Khoo, I.H., Reddy, H.C., Van, L.D., Lin, C.T.: 2-D digital filter architectures without global broadcast and some symmetry applications. In: IEEE International Symposium Circuits System ISCAS, pp. 952–95 (2009)
18. Khoo, I.H., Reddy, H.C., Van, L.D., Lin, C.T.: Generalized formulation of 2-D filter structures without global broadcast for VLSI implementation. In: IEEE MWSCAS, pp. 426–429 (2010)
19. Mohanty, B.K., Meher, P.K.: New scan method and pipeline architecture for VLSI implementation of separable 2-D FIR filters without using transposition. In: Proceedings of IEEE Region 10 TENCON 2008 Conference. Hyderabad, India (2008)
20. Reddy, H.C., Khoo, I.H., Rajan, P.K.: 2-D symmetry: theory and filter design applications. IEEE Circuits Syst. Mag. **3**, 4–33 (2003)
21. Chen, P.Y., Van, L.D., Reddy, H.C., Lin, C.T.: A new VLSI 2-D diagonal-symmetry filter architecture design. In: Proceedings of IEEE APCCAS, pp. 320–323. Macao, China (2008)
22. Chen, P.Y., Van, L.D., Reddy, H.C., Lin, C.T.: A new VLSI 2-D fourfold-rotational-symmetry filter architecture design. In: Proceeding IEEE International Symposium on Circuits System (ISCAS), pp. 93–96 (2009)
23. Chen, P.Y., Van, L.D., Khoo, I.H., Reddy, H.C., Lin, C.T.: Power efficient cost-effective 2-D symmetry filter architecture. IEEE Trans. Circuit Syst. I, Reg. Papers **58**(1), 112–125 (2011)
24. Chen, P.Y., Van, L.D., Reddy, H.C., Khoo, I.H.: Area-efficient 2-D digital filter architectures possessing diagonal and four-fold rotational symmetries. In: Proceedings of ICICS. Taiwan (2013)
25. Kawan, H.: Systolic and parallel realization of 2-D IIR digital filter. In: IEEE International Symposium On Circuit and Systems, vol. 3, pp. 2345–2348. USA (1990)
26. Mohanty, B.K., Meher, P.K.: Cost-effective novel flexible cell-level systolic architecture for high throughput implementation of 2-D FIR filters. IEEE Proc. Comput. Digital Tech. **143**(6), 436–439 (1996)
27. Mohanty, B.K., Meher, P.K., Al-Maadeed, S., Amira, A.: Memory footprint reduction for power-efficient realization of 2-D finite impulse response filters. IEEE Trans. Circuits Syst. I, Reg. Papers **61**(1), 120–133 (2014)
28. Kumar, P., Shrivastava, P., Tiwari, M., Mishra, G.: High-throughput, area-efficient architecture of 2-D block FIR filter using distributed arithmetic algorithm. Circuits Syst. Signal Process (2018). https://doi.org/10.1007/s00034-018-0897-2
29. Kumar, P., Shrivastava, P., Tiwari, M., Dhawan, A.: ASIC implementation of area-efficient high-throughput 2-D IIR filter using distributed arithmetic. Circuits System Signal Process **37**(7), 2934–2957 (2018). https://doi.org/10.1007/s00034-017-0698-z
30. Weste, N.H.E, Eshraghian, K.: Principles of COMS VLSI Design: A Systems Perspective, 2nd ed. Addison-Wesley, Reading, MA (1993). Chaps. 3–5
31. Van, L.D.: Design of efficient VLSI architectures: multiplier, 2-D digital filter and adaptive digital filter, Ph.D. dissertation, Department of the Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (2001)

New LMI Criteria to the Global Asymptotic Stability of Uncertain Discrete-Time Systems with Time Delay and Generalized Overflow Nonlinearities

Pushpendra Kumar Gupta and V. Krishna Rao Kandanvli

Abstract This paper investigates the problem of stability analysis of discrete-time systems under the effect of generalized overflow nonlinearities, parameter uncertainties, and time delay. The systems under assumption involve norm-bounded parameter uncertainties. Two stability criteria based on Linear Matrix Inequality (LMI) approach are presented. The usefulness of the presented criteria is numerically proved.

Keywords Generalized overflow nonlinearity · Global asymptotic stability · Time delay · Parameter uncertainty · Lyapunov method · Linear matrix inequality

1 Introduction

Several practical engineering systems such as Markovian jump systems [1], network control systems [2], neural networks [3], sensor networks [4], etc. can be transformed as discrete systems which can be represented as state-space model.

In these discrete-time systems, the delay can be occurred due to channel, for e.g., transportation delay or some other reason, which may tend the system to be unstable. So, stability analysis of systems having delay is important. Delay may be constant, time-varying, and random in nature. Several studies based on the concept of delay are previously reported [1, 3, 5–13].

Parameter uncertainty is also an important factor for the instability of the discrete systems. Parameter uncertainty arises due to various factors such as modeling errors, finite resolution of the measuring equipment, variation in system parameters, and some other ignored factors. [14–16].

P. K. Gupta (✉) · V. K. R. Kandanvli
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Prayagraj 211004, India
e-mail: pushpendranit30@gmail.com

V. K. R. Kandanvli
e-mail: rao@mnnit.ac.in

While implementing linear discrete-time systems in digital computer with finite wordlength processors, nonlinearities may occur in the systems. Owing to the presence of such nonlinearities, there may be a chance that systems exhibit unstable nature. So, it is very important to find the system parameters range for which the given system is stable. In the present work, the effects of overflow nonlinearities are only taken into account and the quantization effects are assumed as negligible [5, 17–23].

The stability analysis problem of a class of discrete-time uncertain systems having generalized overflow nonlinearities and time delay is significant and more realistic. Very little contribution has been done in the literature [5, 17] so far. In [5], a delay-independent stability criterion is presented and in [17], delay-dependent stability result is provided by employing free-weighting matrix method. Generally, delay-dependent stability analysis provides less conservative results than delay-independent one. A tighter bound inequality, i.e., Wirtinger-based inequality is used in [24, 25] for getting better stability results. Motivated by these concerns and inspired by the work presented in [5, 7, 17, 19, 24–27], we revisit the problem under consideration.

The main aim of the paper is to establish improved delay-dependent criteria for the stability of the systems under consideration. The key involvements of the paper are (1) The system presented covers a wider class of discrete-time systems employing norm-bounded parameter uncertainties, time delay, and generalized overflow nonlinearities. (2) Wirtinger-based inequality [24, 25] is used to deal the sum and cross terms present in the forward difference of Lyapunov function for deriving the delay-dependent stability criteria of the present system which may provide better results. (3) The presented criteria are computationally less complex as they are Linear Matrix Inequality (LMI) based.

The rest of the paper and its organization are given as follows: In Sect. 2, descriptions of the considered systems and some required lemmas are given. Delay-dependent criteria for the stability analysis of the discrete-time systems are presented in Sect. 3. Section 4 shows the effectiveness of the proposed criteria with numerical examples. Finally, concluding remarks are provided in Sect. 5.

In this paper, notations are considered as follows: $\mathbb{R}^k$ represents k-dimensional Euclidean space, $\mathbb{R}^{\alpha \times \beta}$ denotes set of $\alpha \times \beta$ matrices with real elements, $\mathbf{0}$ is a matrix or vector with all elements are zero, $\mathbf{I}$ refers identity matrix of compatible dimension, $\mathbf{P}^{\mathrm{T}}$ stands transpose of the matrix $\mathbf{P}$, $\mathbf{P} > \mathbf{0}(< \mathbf{0})$ shows $\mathbf{P}$ is positive (negative) definite real symmetric matrix, symbol $*$ stands symmetric terms in a symmetric matrix. $diag(a, b, c)$ means diagonal matrix with diagonal elements a,b,c.

2 System Description

The description of the system under consideration is given by

$$\hat{x}(r+1) = \hat{f}(\hat{y}(r))$$
$$= [\hat{f}_1(\hat{y}_1(r)) \, \hat{f}_2(\hat{y}_2(r)) \, \cdots \, \hat{f}_n(\hat{y}_n(r))]^{\mathrm{T}} \tag{1a}$$

$$\hat{y}(r) = \overline{A} \, \hat{x}(r) + \overline{A}_d \, \hat{x}(r-d)$$
$$= \left[\hat{y}_1(r) \, \hat{y}_2(r) \, \cdots \, \hat{y}_n(r) \right]^{\mathrm{T}} \tag{1b}$$

$$\hat{x}(r) = \varphi(r), \, \forall r \in [-d, \, 0] \tag{1c}$$

$$\overline{A} = A + \Delta A, \ \overline{A}_d = A_d + \Delta A_d \tag{1d}$$

where $\hat{x}(r) \in \mathbb{R}^n$ is the state variable; $A, \ A_d \in \mathbb{R}^{n \times n}$ are matrices (known constant); the matrices (unknown) $\Delta A, \ \Delta A_d \in \mathbb{R}^{n \times n}$ having uncertainties in $A, \ A_d$, respectively; at time r, the initial state value is $\varphi(r) \in \mathbb{R}^n$; d is the positive integer for time delay.

The characteristic of generalized overflow nonlinearities $\hat{f}_i(\hat{y}_i(r))$ is specified by

$$\left. \begin{array}{ll} L \leq \hat{f}_i(\hat{y}_i(r)) \leq 1, & \hat{y}_i(r) > 1 \\ \hat{f}_i(\hat{y}_i(r)) = \hat{y}_i(r), & -1 \leq \hat{y}_i(r) \leq 1 \\ -1 \leq \hat{f}_i(\hat{y}_i(r)) \leq -L, & \hat{y}_i(r) < -1 \end{array} \right\} i = 1, 2, \ldots n \tag{2a}$$

where

$$-1 \leq L \leq 1. \tag{2b}$$

With proper choice of L, (2) covers different types of overflow arithmetics, for e.g., saturation ($L = 1$), zeroing ($L = 0$), two's complement and triangular ($L = -1$), etc. In the state matrices, the parameter uncertainties are assumed as

$$\Delta A = B_0 \, F_0 \, C_0 \tag{3a}$$

$$\Delta A_d = B_1 \, F_1 \, C_1 \tag{3b}$$

where $B_i \in \mathbb{R}^{n \times p_i}$, $C_i \in \mathbb{R}^{q_i \times n}$ ($i = 0, 1$) are matrices (known constant) and $F_i \in \mathbb{R}^{p_i \times q_i}$ ($i = 0, 1$) is matrix (unknown) which satisfies

$$F_i^{\mathrm{T}} F_i \leq I, \, i = 0, 1. \tag{3c}$$

For the proof of main results, we present the following lemmas.

Lemma 1 [5, 17–19] *A positive definite matrix* $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{T}} = [\hat{h}_{ij}] \in \boldsymbol{R}^{n \times n}$ *satisfies*

$$\hat{\boldsymbol{y}}^{\mathrm{T}}(r)\boldsymbol{M}\hat{\boldsymbol{y}}(r) - \hat{\boldsymbol{f}}^{\mathrm{T}}(\hat{\boldsymbol{y}}(r))\boldsymbol{M}\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r)) \geq 0 \tag{4}$$

provided that the matrix $\boldsymbol{M}$ *is characterized by*

$$\hat{h}_{ii} = s_i + \sum_{j=1,\, j\neq i}^{n} (\hat{o}_{ij} + \hat{\rho}_{ij}), \ i = 1, 2, \ldots n \tag{5a}$$

$$\hat{h}_{ij} = \hat{h}_{ji} = \left(\frac{1+L}{2}\right)(\hat{o}_{ij} - \hat{\rho}_{ij}), \quad i, j = 1, 2, \ldots n \ (i \neq j) \tag{5b}$$

$$\hat{o}_{ij} = \hat{o}_{ji} > 0, \ \hat{\rho}_{ij} = \hat{\rho}_{ji} > 0 \ \ i, j = 1, 2, \ldots n \ (i \neq j) \tag{5c}$$

$$s_i > 0, \quad i = 1, 2, \ldots n \tag{5d}$$

$$-1 \leq L \leq 1 \tag{5e}$$

where $\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r))$ *is the nonlinearities described by (2) and it is very easy to understand that, for the case where* $n = 1$, $\boldsymbol{M}$ *corresponds to a positive scalar* 'γ'.

Lemma 2 [24, 25] *For a matrix* $\boldsymbol{R} > \boldsymbol{0}$ *and three nonnegative integers* a_1, a_2, r, *as* $a_1 \leq a_2 \leq r$, *if*

$$\boldsymbol{\xi}(r, a_1, a_2) = \frac{1}{a_2 - a_1}[(2\sum_{s=r-a_2}^{r-a_1-1} \hat{\boldsymbol{x}}(s)) + \hat{\boldsymbol{x}}(r - a_1) - \hat{\boldsymbol{x}}(r - a_2)], \ a_1 < a_2$$

$$= 2\hat{\boldsymbol{x}}(r - a_1), \ a_1 = a_2 \tag{6}$$

then

$$-(a_2 - a_1)\sum_{s=r-a_2}^{r-a_1-1} \boldsymbol{\eta}^{\mathrm{T}}(s)\boldsymbol{R}\boldsymbol{\eta}(s) \leq -\begin{bmatrix}\boldsymbol{\theta}_0\\ \boldsymbol{\theta}_1\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}\boldsymbol{R} & \boldsymbol{0}\\ \boldsymbol{0} & 3\boldsymbol{R}\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}_0\\ \boldsymbol{\theta}_1\end{bmatrix} \tag{7}$$

where

$$\boldsymbol{\theta}_0 = \hat{\boldsymbol{x}}(r - a_1) - \hat{\boldsymbol{x}}(r - a_2), \tag{8}$$

$$\boldsymbol{\theta}_1 = \hat{\boldsymbol{x}}(r - a_1) + \hat{\boldsymbol{x}}(r - a_2) - \boldsymbol{\xi}(r, a_1, a_2) \tag{9}$$

$$\boldsymbol{\eta}(s) = \hat{\boldsymbol{x}}(s + 1) - \hat{\boldsymbol{x}}(s) \tag{10}$$

Lemma 3 [14, 15] *Let $\boldsymbol{\Xi}$, Θ, $\boldsymbol{F}$ and $\boldsymbol{U}$ be matrices (real) of suitable dimensions with $\boldsymbol{U} = \boldsymbol{U}^{\mathrm{T}}$, then*

$$\boldsymbol{U} + \boldsymbol{\Xi}\,\boldsymbol{F}\,\Theta + \Theta^{\mathrm{T}}\,\boldsymbol{F}^{\mathrm{T}}\,\boldsymbol{\Xi}^{T} < \boldsymbol{0} \tag{11}$$

for all $\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} \leq \boldsymbol{I}$, iff there exists a positive scalar ε such that

$$\boldsymbol{U} + \varepsilon^{-1}\,\boldsymbol{\Xi}\,\boldsymbol{\Xi}^{\mathrm{T}} + \varepsilon\,\Theta^{\mathrm{T}}\,\Theta < \boldsymbol{0} \tag{12}$$

Lemma 4 [19–22, 27] *Consider $\boldsymbol{B} = [\hat{c}_{uv}] \in \boldsymbol{R}^{n \times n}$ is given by*

$$\hat{c}_{uu} = \sum_{v=1,\,v \neq u}^{n} (\chi_{uv} + \delta_{uv}),\ u = 1, 2 \ldots n \tag{13a}$$

$$\hat{c}_{uv} = L(\chi_{uv} - \delta_{uv}),\ u, v = 1, 2, \ldots n\,(u \neq v) \tag{13b}$$

$$\chi_{uv} > 0,\ \delta_{uv} > 0,\ u, v = 1, 2, \ldots n\,(u \neq v) \tag{13c}$$

$$0 \leq L \leq 1, \tag{13d}$$

(*for $n = 1$, $\boldsymbol{B}$ becomes a positive scalar*)*, then*

$$\sum_{u=1}^{n} 2[\hat{y}_u(r) - \hat{f}_u(\hat{y}_u(r))]\left[\sum_{v=1,\,v \neq u}^{n} (\chi_{uv} + \delta_{uv})\hat{f}_u(\hat{y}_u(r)) + L(\chi_{uv} - \delta_{uv})\hat{f}_v(\hat{y}_v(r)) \right]$$
$$= \hat{\boldsymbol{y}}^{\mathrm{T}}(r)\boldsymbol{B}\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r)) + \hat{\boldsymbol{f}}^{\mathrm{T}}(\hat{\boldsymbol{y}}(r))\boldsymbol{B}^{\mathrm{T}}\hat{\boldsymbol{y}}(r) - \hat{\boldsymbol{f}}^{\mathrm{T}}(\hat{\boldsymbol{y}}(r))(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}})\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r)) \geq 0, \tag{14}$$

where $\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r))$ is given by (2a) *and* (13d).

Lemma 5 [19, 20, 23, 27] *Consider $\boldsymbol{Z} = diag\,(d_1, d_2, \ldots, d_n) > \boldsymbol{0}$. Then, pertaining to* (2a) *and $-1 \leq L < 0$, the following relation is satisfied*

$$\sum_{k=1}^{n} 2d_k[\hat{y}_k(r) - \hat{f}_k(\hat{y}_k(r))]\,[-L\hat{y}_k(r) + \hat{f}_k(\hat{y}_k(r))]$$
$$= (1 + L)\hat{\boldsymbol{y}}^{\mathrm{T}}(r)\boldsymbol{Z}\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r)) + (1 + L)\hat{\boldsymbol{f}}^{\mathrm{T}}(\hat{\boldsymbol{y}}(r))\boldsymbol{Z}\hat{\boldsymbol{y}}(r)$$
$$- 2\hat{\boldsymbol{f}}^{\mathrm{T}}(\hat{\boldsymbol{y}}(r))\boldsymbol{Z}\hat{\boldsymbol{f}}(\hat{\boldsymbol{y}}(r)) - 2L\hat{\boldsymbol{y}}^{\mathrm{T}}(r)\boldsymbol{Z}\hat{\boldsymbol{y}}(r) \geq 0. \tag{15}$$

Remark 1 In this paper, we have considered two ranges of L, i.e., $0 \leq L \leq 1$ and $-1 \leq L < 0$ which together cover (2b).

3   Main Results

In Sect. 3, two criteria are presented for the global asymptotic stability of the system (1a), (1b), (1c), (1d)–(3a), (3b), (3c). The first criterion (Theorem 1) is applicable to the situation where $0 \leq L \leq 1$ and the second one (Theorem 2) is valid for $-1 \leq L < 0$.

Theorem 1 *The system* (1a), (1b), (1c), (1d)–(3a), (3b), (3c) *with* $0 \leq L \leq 1$ *is globally asymptotically stable if there exit suitable dimensioned matrices* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 \\ * & \boldsymbol{p}_3 \end{bmatrix} > \boldsymbol{0},\ \boldsymbol{Q} = \boldsymbol{Q}^{\mathrm{T}} > \boldsymbol{0},\ \boldsymbol{R} = \boldsymbol{R}^{\mathrm{T}} > \boldsymbol{0}$ *and positive scalars* $\varepsilon_0,\ \varepsilon_1$

$$
\left[
\begin{array}{cccc}
-\boldsymbol{p}_1 + ((\boldsymbol{p}_2 + \boldsymbol{p}_2^{\mathrm{T}})/2) + \boldsymbol{Q} + d^2\boldsymbol{R} - 4\boldsymbol{R} + \varepsilon_0\boldsymbol{C}_0^{\mathrm{T}}\boldsymbol{C}_0 & -(\boldsymbol{p}_2/2) - 2\boldsymbol{R} & -d((\boldsymbol{p}_2 - \boldsymbol{p}_3)/2) + 3\boldsymbol{R} \\
* & -\boldsymbol{Q} - 4\boldsymbol{R} + \varepsilon_1\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{C}_1 & -d(\boldsymbol{p}_3/2) + 3\boldsymbol{R} \\
* & * & -3\boldsymbol{R} \\
* & * & * \\
* & * & * \\
* & * & * \\
* & * & *
\end{array}
\right.
$$

$$
\left.
\begin{array}{cccc}
(\boldsymbol{p}_2^{\mathrm{T}}/2) - d^2\boldsymbol{R} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{M} & \boldsymbol{0} & \boldsymbol{0} \\
-(\boldsymbol{p}_2^{\mathrm{T}}/2) + \boldsymbol{A}_d^{\mathrm{T}}\boldsymbol{B} & \boldsymbol{A}_d^{\mathrm{T}}\boldsymbol{M} & \boldsymbol{0} & \boldsymbol{0} \\
d\,\boldsymbol{p}_2^{\mathrm{T}}/2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\
\boldsymbol{p}_1 + (d^2\boldsymbol{R}) - \boldsymbol{M} - (\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) & \boldsymbol{0} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}_0 & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}_1 \\
* & -\boldsymbol{M} & \boldsymbol{M}\boldsymbol{B}_0 & \boldsymbol{M}\boldsymbol{B}_1 \\
* & * & -\varepsilon_0\boldsymbol{I} & \boldsymbol{0} \\
* & * & * & -\varepsilon_1\boldsymbol{I}
\end{array}
\right] < \boldsymbol{0}
$$

$$(16)$$

where the matrices $\boldsymbol{M}$ *and* $\boldsymbol{B}$ *are given by (5) and (13), respectively.*

Proof Consider a quadratic Lyapunov function

$$
V(\hat{\boldsymbol{x}}(r)) = \boldsymbol{\varGamma}^{\mathrm{T}}(r)\boldsymbol{P}\,\boldsymbol{\varGamma}(r) + \sum_{s=r-d}^{r-1} \hat{\boldsymbol{x}}^{\mathrm{T}}(s)\boldsymbol{Q}\hat{\boldsymbol{x}}(s) + d \sum_{\theta=-d+1}^{0} \sum_{s=r+\theta-1}^{r-1} \boldsymbol{\eta}^{\mathrm{T}}(s)\boldsymbol{R}\boldsymbol{\eta}(s)
$$

$$(17)$$

where

$$
\boldsymbol{\varGamma}^{\mathrm{T}}(r) = [\hat{\boldsymbol{x}}^{\mathrm{T}}(r)\ \sum_{s=r-d}^{r-1} \hat{\boldsymbol{x}}^{\mathrm{T}}(s)\,]
$$

and

$$\eta(r) = \hat{x}(r+1) - \hat{x}(r) = \hat{f}(\hat{y}(r)) - \hat{x}(r).$$

Defining

$$\Delta V(\hat{x}(r)) = V(\hat{x}(r+1)) - V(\hat{x}(r))$$
$$= \chi^{\mathrm{T}}(r)\Psi(d)\chi(r) + \hat{x}^{\mathrm{T}}(r)Q\hat{x}(r) - \hat{x}^{\mathrm{T}}(r-d)Q\hat{x}(r-d)$$
$$+ d^2\eta^{\mathrm{T}}(r)R\,\eta(r) - d\sum_{s=r-d}^{r-1}\eta^{\mathrm{T}}(s)R\eta(s) \tag{18}$$

where

$$\chi^{\mathrm{T}}(r) = \left[\,\hat{x}^{\mathrm{T}}(r)\,\hat{x}^{\mathrm{T}}(r-d)\,\xi^{\mathrm{T}}(r,0,d)\,\hat{f}^{\mathrm{T}}(\hat{y}(r))\,\right],$$

$$\Psi(d) = \begin{bmatrix} -p_1 + (p_2 + p_2^{\mathrm{T}})/2 & -p_2/2 & -d((p_2 - p_3)/2) & p_2^{\mathrm{T}}/2 \\ * & 0 & -d\,p_3/2 & -p_2^{\mathrm{T}}/2 \\ * & * & 0 & d\,p_2^{\mathrm{T}}/2 \\ * & * & * & p_1 \end{bmatrix} \tag{19}$$

and $\xi(r,0,d)$ is given by (6).

Next, by the use of Lemma 2, the last term of $\Delta V(\hat{x}(r))$ is expressed as

$$-d\sum_{s=r-d}^{r-1}\eta^{\mathrm{T}}(s)R\eta(s)$$
$$\leq -d\begin{bmatrix} \hat{x}(r) - \hat{x}(r-d) \\ \hat{x}(r) + \hat{x}(r-d) - \xi(r,0,d) \end{bmatrix}^{\mathrm{T}}\begin{bmatrix} R & 0 \\ 0 & 3R \end{bmatrix}\begin{bmatrix} \hat{x}(r) - \hat{x}(r-d) \\ \hat{x}(r) + \hat{x}(r-d) - \xi(r,0,d) \end{bmatrix} \tag{20}$$

Now, from Lemmas 1 and 4, (18) can be rearranged as

$$\Delta V(\hat{x}(r)) = \chi^{\mathrm{T}}(r)\phi(d)\chi(r) - \beta \tag{21a}$$

where

$$\beta = \hat{y}^{\mathrm{T}}(r)M\hat{y}(r) - \hat{f}^{\mathrm{T}}(\hat{y}(r))M\hat{f}(\hat{y}(r)) + \hat{y}^{\mathrm{T}}(r)B\hat{f}(\hat{y}(r))$$
$$+ \hat{f}^{\mathrm{T}}(\hat{y}(r))B^{\mathrm{T}}\hat{y}(r) - \hat{f}^{\mathrm{T}}(\hat{y}(r))(B + B^{\mathrm{T}})\hat{f}(\hat{y}(r)) \geq 0. \tag{21b}$$

$$
\boldsymbol{\phi}(d) = \begin{bmatrix} -\boldsymbol{p}_1 + ((\boldsymbol{p}_2 + \boldsymbol{p}_2^{\mathrm{T}})/2) + \boldsymbol{Q} + d^2\boldsymbol{R} - 4\boldsymbol{R} + \bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{M}\bar{\boldsymbol{A}} & -(\boldsymbol{p}_2/2) - 2\boldsymbol{R} + \bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{M}\bar{\boldsymbol{A}}_d \\ * & -\boldsymbol{Q} - 4\boldsymbol{R} + \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{M}\bar{\boldsymbol{A}}_d \\ * & * \\ * & * \\ -d((\boldsymbol{p}_2 - \boldsymbol{p}_3)/2) + 3\boldsymbol{R} & (\boldsymbol{p}_2^{\mathrm{T}}/2) - d^2\boldsymbol{R} + \bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{B} \\ -d\boldsymbol{p}_3/2 + 3\boldsymbol{R} & -(\boldsymbol{p}_2^{\mathrm{T}}/2) + \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{B} \\ -3\boldsymbol{R} & d\boldsymbol{p}_2^{\mathrm{T}}/2 \\ * & \boldsymbol{P}_1 + (d^2\boldsymbol{R}) - \boldsymbol{M} - (\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) \end{bmatrix} \tag{21c}
$$

From (21a), (21b), (21c), it is implied that $\Delta V(\hat{\boldsymbol{x}}(r)) < 0$ if $\boldsymbol{\phi}(d) < \boldsymbol{0}$. By the aid of Schur's complement, $\boldsymbol{\phi}(d) < \boldsymbol{0}$ is written as

$$
\boldsymbol{\phi}(d) = \begin{bmatrix} -\boldsymbol{p}_1 + ((\boldsymbol{p}_2 + \boldsymbol{p}_2^{\mathrm{T}})/2) + \boldsymbol{Q} + d^2\boldsymbol{R} - 4\boldsymbol{R} & -(\boldsymbol{p}_2/2) - 2\boldsymbol{R} & -d((\boldsymbol{p}_2 - \boldsymbol{p}_3)/2) + 3\boldsymbol{R} \\ * & -\boldsymbol{Q} - 4\boldsymbol{R} & -d\boldsymbol{p}_3/2 + 3\boldsymbol{R} \\ * & * & -3\boldsymbol{R} \\ * & * & * \\ * & * & * \\ (\boldsymbol{p}_2^{\mathrm{T}}/2) - d^2\boldsymbol{R} + \bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{B} & \bar{\boldsymbol{A}}^{\mathrm{T}}\boldsymbol{M} \\ -(\boldsymbol{p}_2^{\mathrm{T}}/2) + \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{B} & \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{M} \\ d\boldsymbol{p}_2^{\mathrm{T}}/2 & \boldsymbol{0} \\ \boldsymbol{p}_1 + d^2\boldsymbol{R} - \boldsymbol{M} - (\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) & \boldsymbol{0} \\ * & -\boldsymbol{M} \end{bmatrix} < \boldsymbol{0} \tag{22}
$$

Using (3a), the inequality (22) can be expressed as

$$
\boldsymbol{\phi}_0(d) + \bar{\boldsymbol{B}}_0\boldsymbol{F}_0\bar{\boldsymbol{C}}_0 + \bar{\boldsymbol{C}}_0^{\mathrm{T}}\boldsymbol{F}_0^{\mathrm{T}}\bar{\boldsymbol{B}}_0^{\mathrm{T}} < \boldsymbol{0} \tag{23}
$$

where

$$
\bar{\boldsymbol{B}}_0^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{0} \ \boldsymbol{0} \ \boldsymbol{0} \ \boldsymbol{B}_0^{\mathrm{T}}\boldsymbol{B} \ \boldsymbol{B}_0^{\mathrm{T}}\boldsymbol{M} \end{bmatrix}, \bar{\boldsymbol{C}}_0 = [\boldsymbol{C}_0 \ \boldsymbol{0} \ \boldsymbol{0} \ \boldsymbol{0} \ \boldsymbol{0}] \tag{24}
$$

$$
\boldsymbol{\phi}_0(d) = \begin{bmatrix} -\boldsymbol{p}_1 + (\boldsymbol{p}_2 + \boldsymbol{p}_2^{\mathrm{T}})/2 + \boldsymbol{Q} + d^2\boldsymbol{R} - 4\boldsymbol{R} & -(\boldsymbol{p}_2/2) - 2\boldsymbol{R} & -d((\boldsymbol{p}_2 - \boldsymbol{p}_3)/2) + 3\boldsymbol{R} \\ * & -\boldsymbol{Q} - 4\boldsymbol{R} & -d\boldsymbol{p}_3/2 + 3\boldsymbol{R} \\ * & * & -3\boldsymbol{R} \\ * & * & * \\ * & * & * \\ (\boldsymbol{p}_2^{\mathrm{T}}/2) - d^2\boldsymbol{R} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{M} \\ -(\boldsymbol{p}_2^{\mathrm{T}}/2) + \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{B} & \bar{\boldsymbol{A}}_d^{\mathrm{T}}\boldsymbol{M} \\ d\boldsymbol{p}_2^{\mathrm{T}}/2 & \boldsymbol{0} \\ \boldsymbol{p}_1 + d^2\boldsymbol{R} - \boldsymbol{M} - (\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}}) & \boldsymbol{0} \\ * & -\boldsymbol{M} \end{bmatrix} \tag{25}
$$

Employing Lemma 3, (23) can be written as

$$
\boldsymbol{\phi}_0(d) + \varepsilon_0^{-1}\bar{\boldsymbol{B}}_0\bar{\boldsymbol{B}}_0^{\mathrm{T}} + \varepsilon_0\bar{\boldsymbol{C}}_0^{\mathrm{T}}\bar{\boldsymbol{C}}_0 < \boldsymbol{0} \tag{26}
$$

where $\varepsilon_0 > 0$. Next, with the aid of Schur's complement, (26) yields

$$
\begin{bmatrix}
-p_1 + ((p_2 + p_2^T)/2) + Q + d^2R - 4R + \varepsilon_0 C_0^T C_0 & -(p_2/2) - 2R & -d((p_2 - p_3)/2) + 3R \\
* & -Q - 4R & -d(p_3/2) + 3R \\
* & * & -3R \\
* & * & * \\
* & * & * \\
* & * & *
\end{bmatrix}
$$

$$
\begin{bmatrix}
(p_2^T/2) - d^2R + A^TB & A^TM & 0 \\
-(p_2^T/2) + \bar{A}_d^TB & \bar{A}_d^TM & 0 \\
dp_2^T/2 & 0 & 0 \\
p_1 + d^2R - M - (B + B^T) & 0 & B^TB_0 \\
* & -M & MB_0 \\
* & * & -\varepsilon_0 I
\end{bmatrix} < 0
$$

(27)

Next, following the analysis similar to (23)–(27), one can easily get (16) from (27).
    This ends the proof.
    Next, we present the following result.

Theorem 2 *The system* (1a), (1b), (1c), (1d)–(3a), (3b), (3c)*along with* $-1 \le L < 0$
is globally asymptotically stable if there exit suitable dimensioned matrices $P = \begin{bmatrix} p_1 & p_2 \\ * & p_3 \end{bmatrix} > 0$, $Q = Q^T > 0$, $R = R^T > 0$, *and positive scalars* ε_0, ε_1 *satisfying*

$$
\begin{bmatrix}
\begin{matrix} -p_1 + (p_2 + p_2^T)/2 + Q + \\ d^2R - 4R + \varepsilon_0 C_0^T C_0 \end{matrix} & -(p_2/2) - 2R & -d(p_2 - p_3)/2 + 3R & (p_2^T/2) - d^2R + (1+L)A^TZ \\
* & -Q - 4R + \varepsilon_1 C_1^T C_1 & -d(p_3/2) + 3R & -(p_2^T/2) + (1+L)A_d^TZ \\
* & * & -3R & d(p_2^T/2) \\
* & * & * & p_1 + d^2R - M - 2Z \\
* & * & * & * \\
* & * & * & * \\
* & * & * & * \\
* & * & * & *
\end{bmatrix}
$$

$$
\begin{bmatrix}
A^TM & (-2L)^{1/2}A^TZ & 0 & 0 \\
A_d^TM & (-2L)^{1/2}A_d^TZ & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & (1+L)ZB_0 & (1+L)ZB_1 \\
-M & 0 & MB_0 & MB_1 \\
0 & -Z & (-2L)^{1/2}ZB_0 & (-2L)^{1/2}ZB_1 \\
* & * & -\varepsilon_0 I & 0 \\
* & * & * & -\varepsilon_1 I
\end{bmatrix} < 0
$$

(28)

where M *is given by (5) and* Z *is a positive definite diagonal matrix of appropriate dimension.*

Proof Pertaining to the case where $-1 \le L < 0$ and in view of Lemmas 1 and 5, Eq. (18) (obtained in the proof of Theorem 1) can be mathematically re-expressed

as

$$\Delta V(\hat{x}(r)) = \chi^{T}(r)\phi_1(d)\chi(r) - \beta_1 \tag{29a}$$

where

$$
\begin{aligned}
\beta_1 &= \hat{y}^{T}(r)M\hat{y}(r) - \hat{f}^{T}(\hat{y}(r))M\hat{f}(\hat{y}(r)) + (1+L)\hat{y}^{T}(r)Z\hat{f}(\hat{y}(r)) \\
&\quad + (1+L)\hat{f}^{T}(\hat{y}(r))Z\hat{y}(r) - 2\hat{f}^{T}(\hat{y}(r))Z\hat{f}(\hat{y}(r)) - 2L\hat{y}^{T}(r)Z\hat{y}(r) \geq 0,
\end{aligned}
\tag{29b}
$$

$$
\phi_1(d) = \begin{bmatrix}
\begin{array}{c} -p_1 + ((p_2 + p_2^{T})/2) + Q + d^2 R - 4R + \\ \bar{A}^{T}M\bar{A} - 2L\bar{A}^{T}Z\bar{A} \end{array} & -(p_2/2) - 2R + \bar{A}^{T}M\bar{A}_d - 2L\bar{A}^{T}Z\bar{A}_d \\
* & -Q - 4R + \bar{A}_d^{T}M\bar{A}_d - 2L\bar{A}_d^{T}Z\bar{A}_d \\
* & * \\
* & *
\end{array} \\
\begin{array}{cc}
-d((p_2 - p_3)/2) + 3R & (p_2^{T}/2) - d^2 R + (1+L)\bar{A}^{T}Z \\
-d(p_3/2) + 3R & -(p_2^{T}/2) + (1+L)\bar{A}_d^{T}Z \\
-3R & d\,p_2^{T}/2 \\
* & p_1 + d^2 R - M - 2Z
\end{array}
\end{bmatrix}
\tag{30}
$$

From (29a), (29b), it follows that $\Delta V(\hat{x}(r)) < 0$ if $\phi_1(d) < 0$.

Next, following the similar steps as shown in the proof of Theorem 1, the condition $\phi_1(d) < 0$ leads to (28). This ends the proof.

Remark 2 The given criteria are in LMI setting and can be mathematically tractable by MATLAB with YALMIP 3.0 parser [28].

Remark 3 Note that it is worth to compare the present results with Corollary 9 of [17]. To handle the sum and cross terms of the forward difference of the Lyapunov function, Wirtinger-based inequality (see Lemma 2) is considered in this paper while in [17], free-weighting matrix approach is used for the same. Consequently, the present approach requires less number of decision variables as compared to [17]. Hence, the present approach gains improvement in terms of computational complexity relative to [17].

Remark 4 As future work, the idea of the present paper can be extended to the stability of discrete-time uncertain systems with external disturbances and/or time-varying delay using generalized overflow nonlinearities, H_{∞} stability of discrete-time uncertain systems with generalized overflow nonlinearities and time delay, etc. Further, one can also apply the present idea for two dimensional (2-D) and multidimensional ($n > 2$) systems.

Remark 5 The time delay considered in the existing literature may be constant and/or time-varying. In the modeling of the practical engineering/industrial systems, i.e., microgrid systems with constant communication delay [29] and networked control systems with constant feedback delay [30] where constant time delay has also been taken into account, the main objective of all these concerns including the present paper is to find the allowable time delay that assures the system stability.

4 Numerical Examples

The usefulness of the presented criteria is given by the following numerical examples.

Example 1 Consider the system (1a), (1b), (1c), (1d)–(3a), (3b), (3c), where

$$A = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.91 \end{bmatrix}, \; A_d = \begin{bmatrix} -0.1 & 0 \\ -0.1 & -0.1 \end{bmatrix}$$

$$B_0 = B_1 = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, \; C_0 = [0.01 \quad 0], \; C_1 = [0 \quad 0.01] \tag{31}$$

Pertaining to $L = 1$, it is found that (16) is feasible for maximum delay $d = 18$. Hence, according to Theorem 1, global asymptotically stability of the system under consideration is achieved whereas Corollary 9 of [17] fails to obtain the same.

Example 2 Consider the discrete-time system described by (1a), (1b), (1c), (1d)–(3a), (3b), (3c) with (31) and $L = -1/3$. It is verified easily that Theorem 2 assures the global asymptotic stability of the system over a maximum delay $d = 13$. By contrast, Corollary 9 of [17] fails to this end.

Examples 1 and 2 support the fact that the presented approach may yield stability results which are not covered by [17]. In summary, the present results may establish improved stability conditions (less conservative and computationally efficient) which are uncovered by existing results.

5 Conclusions

In this paper, two delay-dependent criteria for stability analysis of discrete-time uncertain systems with time delay and generalized overflow nonlinearities have been established. The presented criteria are in the form of LMIs and, hence, computationally tractable. The effectiveness of the proposed criteria is proved with the numerical examples.

References

1. Li, X., Zhang, X., Wang, X.: Stability analysis for discrete-time markovian jump systems with time-varying delay: a homogeneous polynomial approach. IEEE Access **5**, 27573–27581 (2017). https://doi.org/10.1109/ACCESS.2017.2775606
2. Wu, L., Lam, J., Yao, X., Xiong, J.: Robust guaranteed cost control of discrete-time networked control systems. Optim. Control. Appl. Methods **32**(1), 95–112 (2011). https://doi.org/10.1002/oca.932
3. Zhang, C.K., He, Y., Jiang, L., Wang, Q.G., Wu, M.: Stability analysis of discrete-time neural networks with time-varying delay via an extended reciprocally convex matrix inequality. IEEE Trans. Cybern. **47**(10), 3040–3049 (2017). https://doi.org/10.1109/TCYB.2017.2665683
4. Zhang, D., Shi, P., Zhang, W.A., Yu, L.: Energy-efficient distributed filtering in sensor networks: a unified switched system approach. IEEE Trans. Cybern. **47**(7), 1618–1629 (2017). https://doi.org/10.1109/TCYB.2016.2553043
5. Kandanvli, V.K.R., Kar, H.: Robust stability of discrete-time state-delayed systems employing generalized overflow nonlinearities. Nonlinear Anal.: Theory, Methods Appl. **69**(9), 2780–2787 (2008). https://doi.org/10.1016/j.na.2007.08.050
6. Kandanvli, V.K.R., Kar, H.: Robust stability of discrete-time state-delayed systems with saturation nonlinearities: Linear matrix inequality approach. Signal Process **89**(2), 161–173 (2009). https://doi.org/10.1016/j.sigpro.2008.07.020
7. Kandanvli, V.K.R., Kar, H.: An LMI condition for robust stability of discrete-time state-delayed systems using quantization/overflow nonlinearities. Signal Process **89**(11), 2092–2102 (2009). https://doi.org/10.1016/j.sigpro.2009.04.024
8. Guan, X., Lin, Z., Duan, G.: Robust guaranteed cost control for discrete-time uncertain systems with delay. IEE Proc. Control Theory Appl. **146**(6), 598–602 (1999). https://doi.org/10.1049/ip-cta:19990714
9. Bakule, L., Rodellar, J., Rossell, J.M.: Robust overlapping guaranteed cost control of uncertain state-delay discrete-time systems. IEEE Trans. Automat. Control **51**(12), 1943–1950 (2006). https://doi.org/10.1109/TAC.2006.886536
10. Chen, W.H., Guan, Z.H., Lu, X.: Delay-dependent guaranteed cost control for uncertain discrete-time systems with delay. IEE Proc. Control Theory Appl. **150**(4), 412–416 (2003). https://doi.org/10.1049/ip-cta:20030572
11. Mahmoud, M.S.: Robust Control and Filtering for Time-Delay Systems. CRC Press, Marcel-Dekker, New York (2000)
12. Mahmoud, M.S., Boukas, E.K., Ismail, A.: Robust adaptive control of uncertain discrete-time state-delay systems. Comput. Math. Appl. **55**(12), 2887–2902 (2008). https://doi.org/10.1016/j.camwa.2007.11.021
13. Xu, S.: Robust H_∞ filtering for a class of discrete-time uncertain nonlinear systems with state delay. IEEE Trans. Circuits Syst. I **49**(12), 1853–1859 (2002). https://doi.org/10.1109/tcsi.2002.805736
14. Boyd, S., Ghaoui, L.E., Feron, E., Balakrishnan, V.: Linear Matrix Inequalities in System and Control Theory. SIAM, Philadelphia, PA (1994)
15. Xu, S., Lam, J., Lin, Z., Galkowski, K.: Positive real control for uncertain two-dimensional systems. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **49**(11), 1659–1666 (2002). https://doi.org/10.1109/TCSI.2002.804531

16. Bakule, L., Rodellar, J., Rossell, J.M.: Robust overlapping guaranteed cost control of uncertain state-delay discrete-time systems. IEEE Trans. Autom. Control **51**(12), 1943–1950 (2006). https://doi.org/10.1109/TAC.2006.886536

17. Kandanvli, V.K.R., Kar, H.: A delay-dependent approach to stability of uncertain discrete-time state-delayed systems with generalized overflow nonlinearities. ISRN Comput. Math. (2012). https://doi.org/10.5402/2012/171606

18. Liu, D., Michel, A.N.: Asymptotic stability of discrete-time systems with saturation nonlinearities with applications to digital filters. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **39**(10), 798–807 (1992). https://doi.org/10.1109/81.199861

19. Rani, P., Kokil, P., Kar, H.: l_2–l_∞ suppression of limit cycles in interfered digital filters with generalized overflow nonlinearities. Circuits, Syst. Signal Process. **36**(7), 2727–2741 (2017). https://doi.org/10.1007/s00034-016-0433-1

20. Dey, A., Kar, H.: LMI-based criterion for the robust stability of 2D discrete state-delayed systems using generalized overflow nonlinearities. J. Control Sci. Eng. **23** (2011). https://doi.org/10.1155/2011/271515

21. Kar, H.: An LMI based criterion for the nonexistence of overflow oscillations in fixed-point state-space digital filters using saturation arithmetic. Digit. Signal Proc. **17**(3), 685–689 (2007). https://doi.org/10.1016/j.dsp.2006.11.003

22. Kokil, P., Kandanvli, V.K.R., Kar, H.: A note on the criterion for the elimination of overflow oscillations in fixed-point digital filters with saturation arithmetic and external disturbance. AEU-Int. J. Electron. Commun. **66**(9), 780–783 (2012). https://doi.org/10.1016/j.aeue.2012.01.004

23. Kar, H., Singh, V.: A new criterion for the overflow stability of second-order state-space digital filters using saturation arithmetic. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **45**(3), 311–313 (1998). https://doi.org/10.1109/81.662720

24. Nam, P.T., Pathirana, P.N., Trinh, H.: Discrete wirtinger-based inequality and its application. J. Franklin Inst. **352**(5), 1893–1905 (2015). https://doi.org/10.1016/j.jfranklin.2015.02.004

25. Tadepalli, S.K., Kandanvli, V.K.R., Vishwakarma, A.: Criteria for stability of uncertain discrete-time systems with time-varying delays and finite wordlength nonlinearities. Trans. Inst. Meas. Control **40**(9), 2868–2880 (2017). https://doi.org/10.1177/0142331217709067

26. Kokil, P., Parthipan, C.G., Jogi, S., Kar, H.: Criterion for realizing state-delayed digital filters subjected to external interference employing saturation arithmetic. Cluster Comput. 1–8 (2018)

27. Kokil, P., Arockiaraj, S.X., Kar, H.: Criterion for limit cycle-free state-space digital filters with external disturbances and generalized overflow nonlinearities. Trans. Inst. Meas. Control **40**(4), 1158–1166 (2018). https://doi.org/10.1177/0142331216680287

28. Lofberg, J.: YALMIP: a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE International Symposium on Computer Aided Control Systems Design, pp. 284–289. IEEE, New Orleans, LA, USA (2001). https://doi.org/10.1109/cacsd.2004.1393890

29. Mary, T.J., Rangarajan, P.: Delay-dependent stability analysis of microgrid with constant and time-varying communication delays. Electr. Power Compon. Syst. **44**(13), 1441–1452 (2016). https://doi.org/10.1080/15325008.2016.1170078

30. Razeghi-Jahromi, M., Seyedi, A.: Stabilization of distributed networked control systems with constant feedback delay. In: IEEE 52nd Annual Conference on Decision and Control (CDC), pp. 4619–4624. IEEE, Florence, Italy (2013). https://doi.org/10.1109/cdc.2013.6760612

Improved DWT-SVD-Based Medical Image Watermarking Through Hamming Code and Chaotic Encryption

S. Thakur, A. K. Singh, Basant Kumar and S. P. Ghrera

Abstract This paper presents an efficient watermarking technique to enhance the performance of DWT-SVD-based approach. The method uses well-known error-correcting code (ECC) and chaotic encryption to reduce the channel noise distortion and improve the security of the technique, respectively. In the proposed method, the cover image is transformed by DWT and the sub-bands are selected for embedding the watermarks. Subsequently, the selected sub-bands are further transformed by SVD. The more robust watermark "patient report" and less robust watermark "patient medical image" is embedded into singular values of the selected DWT sub-bands. The use of transform domain techniques along with hamming code ensures that the approach offers more robustness and reliability. Inclusion of chaos-based encryption carries twofold benefits such as obscuring the watermarked content and strengthening the overall security of the projected scheme. The proposed technique offers a significant value of peak signal-to-noise ratio (PSNR), normalized correlation (NC), bit error rate (BER), number of pixel change rate (NPCR), and unified average changing intensity (UCAI) in presence of major attacks. The performance comparisons of our technique with a similar approach show that the proposed technique is superior for nine considered attacks.

Keywords DWT · SVD · Image watermarking · Error-correcting code · Chaotic encryption · PSNR · NC · BER · NPCR · UACI

S. Thakur · S. P. Ghrera
Department of CSE, JUIT, Solan, HP, India
e-mail: sritithakur19@gmail.com

S. P. Ghrera
e-mail: sp.ghrera@juit.ac.in

A. K. Singh (✉)
Department of CSE, NIT Patna, Patna, Bihar, India
e-mail: amit_245singh@yahoo.com

B. Kumar
Department of ECE, MNNIT Allahabad, Allahabad, UP, India
e-mail: singhbasant@yahoo.com

897

1   Introduction

Nowadays, dissemination of modern health facilities is majorly depending upon information and communication technology (ICT) based networked transmission mediums [1]. Though it comes with an advantage of the ease of access and easy and swift exchange of crucial information, yet it bears a major disadvantage too. Maintaining proper security of such sensitive transmission channels becomes the utmost priority of health personnel and concerned management [1–3]. This draws our attention for an all-encompassing approach which fulfills the confidentiality, integrity, and authenticity requirements and at the same time, maintains the privacy of the exchanged data. Digital watermarking is proving to be an effective approach which can act as a countermeasure against such issues [4, 5]. Recently, cryptographic schemes are also used in combination with watermarking algorithms to provide a strong barrier against security threats [5].

In our study, we develop a dual watermarking scheme for medical images to increase the performance of DWT-SVD-based approach. The method uses the concept of DWT to transform the cover image into different frequency components for optimum embedding. However, SVD is harnessed for obtaining the matrix with maximum energy and minimum coefficient values which facilitates smooth and accurate embedding [6, 7]. Prior to embedding, the "patient medical image" is separated into sub-images and the "patient report" is encoded with Hamming code [8]. After embedding, butterfly concept which governs the chaos-based encryption [9, 10] is used to encrypt the watermarked image.

The remaining section of the paper is ordered as follows: current literature related to our work is given in the "related work" section. In section "design and outline of the proposed scheme", the concept behind our scheme is discussed followed by results in "results" section. The paper is summarized in the "conclusion" section.

2   Related Research

In this section, the related researches are discussed.

A pixel-based watermarking approach is presented in [11] for medical images. The method uses bit replacement scheme to embed ROI information into background part (NROI) of the host image. Further, the method identified the tampered area at the receiver end. Experimental results proved it's superiority to existing techniques [12–15].

A dual watermarking scheme is implemented using medical images in [16]. The authors estimated the performance of four different ECCs. Experimentally evaluating the proposed technique yielded improved results in terms of robustness and it

was observed that Reed–Solomon performed better than the rest of the ECC's. Comparison with similar techniques, [17, 18] showed better performance in terms of NC. In [19], a reversible watermarking method for e-health application is presented. The ROI section of original image is transformed by DWT then SVD is performed on the chosen sub-bands of image. The watermark is embedded into the particular SVD component ("U" matrix). Further, ECC is used to encode EPR watermark which is hidden into the RONI section of original image. The method performed better than similar state-of-the-art methods [20, 21]. In [22], the authors presented a watermarking method using lossless compression addressing the bandwidth/storage issue of the medical data. The method offered good quality images at higher compression ratios. In addition to this, it is robust when subjected to attacks. A secure and robust DWT-based watermarking is developed in [21]. The embedding of watermark is performed using spread-spectrum method. Further, BCH code is used to enhance the robustness.

3   Design and Outline of the Proposed Scheme

Our approach comprises of two parts, i.e., insertion and significant recovery of both watermarks. Initially, the host image is transformed by second level DWT and three sub-bands are chosen for embedding the watermarks. Subsequently, the selected sub-bands are further transformed by SVD. The singular component of high–low (HL) and low–high (LH) frequency components of first level DWT image is altered with partitioned watermark medical images. Further, singular vector of the second level DWT coefficients (high–high (HH2)) is selected for embedding of "patient report" watermark. After embedding, chaos-based encryption [9, 10] is applied on watermarked image to increase the authenticity of our scheme. DWT decomposition (up to second level) is depicted in Fig. 1 [4]. The entire process of watermark insertion and recovery is depicted in Fig. 2.
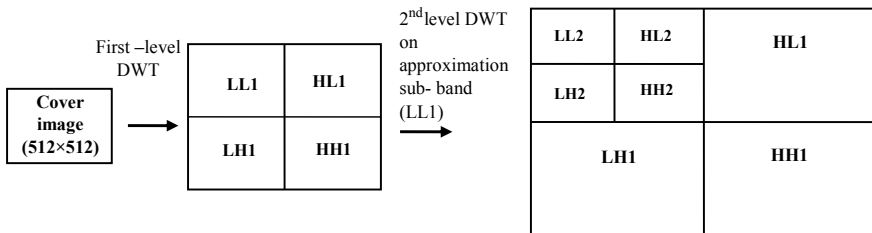
Fig. 1 DWT decomposition of an image
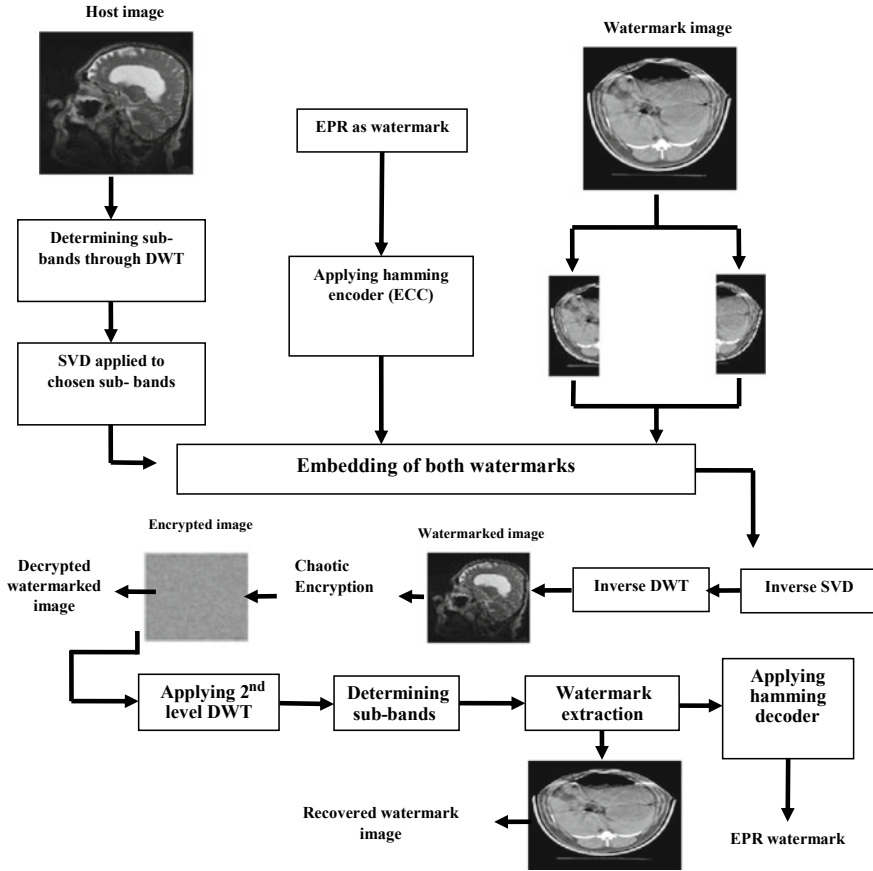
Fig. 2 Proposed watermark insertion and recovery approach

4 Results

The proposed work is experimentally evaluated using MATLAB software-version R2013a. To estimate the performance of our scheme, six medical images and Barbara image are selected as cover [23]. Further, to improve the authenticity, two watermarks "patient report" (12 characters) and "patient medical image" (256 × 256) are used for testing and evaluating our method. The results are carried out in the form of PSNR, NC, BER [1]. Further, NPCR and UACI estimate the efficacy of the encryption scheme [7, 9, 24]. The images used in our experimental analysis are—(a) Cover (b) watermark (c) watermarked and (d) "patient report", as shown in Fig. 3. Tabular representation of the performance of our scheme is shown in Tables 1, 2, 3 and 4.

Our work is evaluated on particular gain factor values as shown in Table 1. Table 1 shows the highest PSNR = 38.3571 dB at a gain = 0.005. The BER, NPCR, and UACI values are approximately reaching the acceptable values for all gain factors,

(a)

Original image

(b)

Watermark image

(c)

Watermarked image

(d)

Sriti Thakur
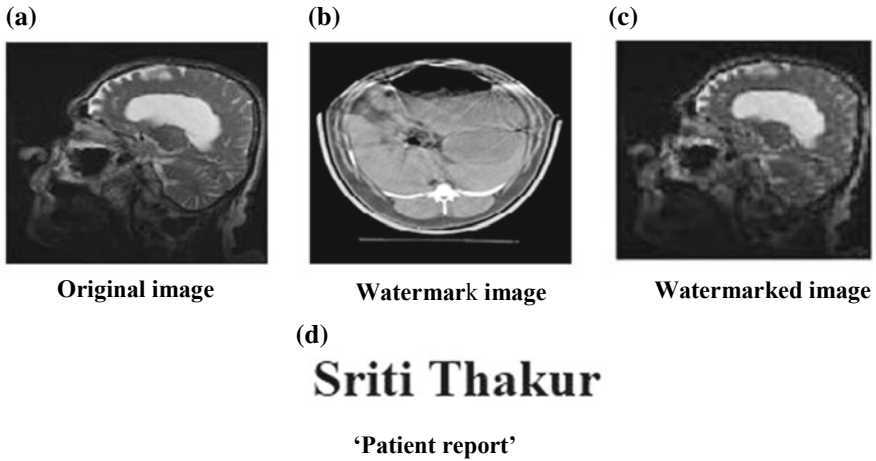
'Patient report'

Fig. 3 Considered images as **a** Cover, **b** watermark, **c** watermarked, and **d** "patient report"

Table 1 Performance analysis at a varying gain

Gain factor	PSNR (in dB)	NC (image)	BER (text)	NPCR	UACI
0.005	38.3571	0.9154	0	0.9961	0.4010
0.01	38.3538	0.9676	0	0.9960	0.4012
0.03	38.0954	0.9916	0	0.9960	0.4010
0.05	36.9234	0.9154	0	0.9961	0.4010
0.07	35.2507	0.9977	0	0.9961	0.4011
0.09	33.5399	0.9985	0	0.9960	0.4014
0.1	32.7280	0.9986	0	0.9958	0.4012

Table 2 Performance analysis for various cover images

Cover media	PSNR (in dB)	NC (image)	BER	NPCR	UACI
CT-scan	21.4548	0.8865	0	0.9961	0.4546
PET-CT	38.4215	0.9648	0	0.9960	0.4811
Ultrasound	24.5103	0.9327	0	0.9961	0.4132
PET-scan	28.0304	0.9488	0	0.99612	0.44804
Brain MRI	36.9234	0.9154	0	0.9961	0.4010
Hand X-ray	38.0842	0.9869	0	0.9961	0.4906
Barbara	36.4312	0.9793	0	0.9960	0.4206

Table 3 Robustness analysis for image and text watermark(s) under attacks

Attack	Noise density	NC (image)	BER (text)
Salt and pepper noise	0.0001	0.9975	0
	0.0005	0.9630	0
	0.001	0.8761	0
Gaussian noise	0.0001	0.9785	0
	0.0005	0.8311	0
Rotation	1°	0.9308	55.95
	5°	0.8908	55.95
	10°	0.8913	0.55
JPEG compression	QF = 10 and 50	0.8994 and 0.9626	0
Sharpening mask	0.1	0.8042	0
	0.5	0.8320	0
	0.9	0.8445	0
Speckle noise	0.001	0.9947	0
	0.01	0.8277	0
Image scaling	×1	0.9973	0
	×2	0.8242	0
Cropping	[5 5, 10 10]	0.9800	7.1428
	[10 10, 200 200]	0.9009	7.1428
Gaussian low-pass filter at Var = 0.6	Mean = 1 and 3	0.9973 and 0.9124	0
Median filter	[1 1]	0.9973	0
	[2 2]	0.9099	0
	[3 3]	0.9290	0
Poisson noise		0.8026	0
Histogram equalization		0.6624	2.38

respectively. The highest NC values are 0.9986 obtained for a gain of 0.1. Performance measures for different host images are presented in Table 2. In this table, we notice that the proposed work achieves significant value of PSNR = 38.0842 dB and NC = 0.9869 for "X-ray" image. Further, Table 3 shows the analysis of the work when subjected to several attacks. The highest NC = 0.9975 for salt and pepper noise (density = 0.0001). The proposed work has achieved acceptable range of NC value, i.e., above 0.7 except for histogram equalization. Similarly, BER values are equal to "0" except for rotation attack, cropping, and histogram equalization.

The performance comparison of our technique with a similar approach [16] is shown in Table 4 and it is found that our technique is superior for nine considered attacks.

Table 4 Comparative analysis of our method with Singh et al.

Attack	Noise density	NC and BER by [16]		NC and BER [proposed scheme]	
		NC	BER	NC	BER
JPEG compression	QF = 100	0.9950	0	0.9974	0
Sharpening mask	0.1, 0.5 and 0.9	0.5986, 0.6293 and 0.6457	0	0.80422, 0.8320 and 0.8445	0
Gaussian noise	0.05	0.3150	8.5714	0.4157	1.190
Image scaling	2 and 2.5				
Salt and pepper noise	0.001	0.7553	0	0.8765	0
Median filter	[3 3]	0.8885	0	0.9290	0
Gaussian low-pass filter	0.6	0.8780	0	0.9124	0
Cropping		0.7451	4.5714	0.9800	7.1428
Histogram equalization		0.5880	1.4286	0.6624	2.3809

5 Conclusion

In this research, DWT SVD hamming chaotic encryption-based dual watermarking scheme was developed. To improve the authenticity of our method, dual watermarks were embedded in the different wavelet coefficients/frequency sub-bands of the host image. On obtaining the watermarked image, it is subjected to an encryption mechanism governed by chaos which ensures the security of the entire process. Tabular representation of the simulation results has shown the suitability of our method for medical applications. In comparison to previous similar approach, our technique achieved better "NC" and "BER" values. Although the evaluated performance of our technique is acceptable, in future work, we plan to test our method for other applications such as audio watermarking and video watermarking.

References

1. Singh, A.K., Kumar, B., Singh, S.K., Ghrera, S.P., Mohan, A.: Multiple watermarking technique for securing online social network contents using back propagation neural network. Future Gener. Comput. Syst. **86**, 926–939 (2018). https://doi.org/10.1016/j.future.2016.11.023
2. Priya, S., Varatharajan, R., Manogaran, G., Sundarasekar, R., Kumar, P.M.: Paillier homomorphic cryptosystem with poker shuffling transformation based water marking method for the secured transmission of digital medical images. Pers. Ubiquitous Comput., 1–11 (2018). https://doi.org/10.1007/s00779-018-1131-8
3. Lakshmi, C., Thenmozhi, K., Rayappan, J.B., Amirtharajan, R.: Encryption and watermark-treated medical image against hacking disease—an immune convention in spatial and frequency

domains. Comput. Methods Programs Biomed. **159**, 11–21 (2018). https://doi.org/10.1016/j.cmpb.2018.02.021

4. Singh, A.K., Kumar, B., Singh, G., Mohan, A. (eds.) Medical Image Watermarking: Techniques and Applications. Book series on Multimedia Systems and Applications. Springer (2017)

5. Singh, A.K., Dave, M., Mohan, A.: Wavelet based image watermarking: futuristic concepts in information security. Proc. Natl. Acad. Sci., India, Sect. A **84**(3), 345–359 (2014). https://doi.org/10.1007/s40010-014-0140-x

6. Chao, H.M., Hsu, C.M., Miaou, S.G.: A data-hiding technique with authentication, integration, and confidentiality for electronic patient records. IEEE Trans. Inf. Technol. Biomed. **6**(1), 46–53 (2002). https://doi.org/10.1109/4233.992161

7. Thakur, S., Singh, A.K., Ghrera, S.P., Elhoseny, M.: Multi-layer security of medical data through watermarking and chaotic encryption for tele-health applications. Multimed. Tools Appl., 1–14 (2018). https://doi.org/10.1007/s11042-018-6263-3

8. Singh, A.K., Kumar, B., Dave, M., Ghrera, S.P., Mohan, A.: Digital image watermarking: techniques and emerging applications. In: Handbook of Research on Modern Cryptographic Solutions for Computer and Cyber Security, USA, IGI Global, pp. 246–272 (2016). https://doi.org/10.4018/978-1-5225-0105-3.ch011

9. Wu, Y., Noonan, J.P., Yang, G., Jin, H.: Image encryption using the two-dimensional logistic chaotic map. J. Electron. Imaging **21**(1) (2012). https://doi.org/10.1117/1.jei.21.1.013014

10. Abd El-Samie, F.E., Ahmed, H.E.H., Elashry, I.F., Shahieen, M.H., Faragallah, O.S., El-Rabaie, E.-S.M., Alshebeili, S.A.: Image Encryption: A Communication Perspective. CRC Press (2013)

11. Bakthula, R., Shivani, S., Agarwal, S.: Self authenticating medical X-ray images for telemedicine applications. Multimed. Tools Appl., 1–18 (2017). https://doi.org/10.1007/s11042-017-4738-2

12. Tan, C.K., Ng, J.C., Xu, X., Poh, C.L., Guan, Y.L., Sheah, K.: Security protection of DICOM medical images using dual-layer reversible watermarking with tamper detection capability. J. Digit. Imaging **24**(3), 528–540 (2011). https://doi.org/10.1007/s10278-010-9295-4

13. Eswaraiah, R., Reddy, E.S.: Medical image watermarking technique for accurate tamper detection in ROI and exact recovery of ROI. Int. J. Telemed. Appl., 13 (2014). https://doi.org/10.1155/2014/984646

14. Memon, N.A., Chaudhry, A., Ahmad, M., Keerio, Z.A.: Hybrid watermarking of medical images for ROI authentication and recovery. Int. J. Comput. Math. **88**(10), 2057–2071 (2011). https://doi.org/10.1080/00207160.2010.543677

15. Wu, J.H., Chang, R.F., Chen, C.J., Wang, C.L., Kuo, T.H., Moon, W.K., Chen, D.R.: Tamper detection and recovery for medical images using near-lossless information hiding technique. J. Digit. Imaging **21**(1), 59–76 (2008). https://doi.org/10.1007/s10278-007-9011-1

16. Singh, A.K., Kumar, B., Dave, M., Mohan, A.: Robust and imperceptible dual watermarking for telemedicine applications. Wirel. Pers. Commun. **80**(4), 1415–1433 (2015). https://doi.org/10.1007/s11277-014-2091-6

17. Tripathi, S., Ramesh, N., Bernito, A., Neeraj, K.J.: A DWT based dual image watermarking technique for authenticity and watermark protection. An Int. J. (SIPIJ) **1**(2), 33–45 (2010). https://doi.org/10.5121/sipij.2010.1204

18. Mahajan, L.H., Patil, S.A.: Image watermarking scheme using SVD. Int. J. Adv. Res. Sci. Eng. **2**(6), 69–77 (2013)

19. Thakkar, F.N., Srivastava, V.K.: A blind medical image watermarking: DWT-SVD based robust and secure approach for telemedicine applications. Multimed. Tools Appl. **3**(76), 3669–3697 (2016). https://doi.org/10.1007/s11042-016-3928-7

20. Parah, S.A., Sheikh, J.A., Ahad, F., Loan, N.A., Bhat, G.M.: Information hiding in medical images: a robust medical image watermarking system for e-healthcare. Multimed. Tools Appl. **8**(76), 10599–10633 (2015). https://doi.org/10.1007/s11042-015-3127-y

21. Singh, A.K., Kumar, B., Dave, M., Mohan, A.: Multiple watermarking on medical images using selective discrete wavelet transform coefficients. J. Med. Imaging Health Inform. **5**(3), 607–614 (2015). https://doi.org/10.1166/jmihi.2015.1432

22. Amri, H., Khalfallah, A., Gargouri, M., Nebhani, N., Lapayre, J.C., Bouhlel, M.S.: Medical image compression approach based on image resizing, digital watermarking and lossless compression. J. Signal Process. Syst. **87**(2), 203–214 (2017). https://doi.org/10.1007/s11265-016-1150-5
23. MedPix online medical image database. https://medpix.nlm.nih.gov/
24. Wu, Y., Noonan, J.P., Agaian, S.: NPCR and UACI randomness tests for image encryption. Cyber J. Multidiscip. J. Sci. Technol. JSAT, 31–38 (2011)

Dynamically Tuned PIDD^2 Controller for Single-Link Flexible Manipulator

Komal Agrawal, Richa Negi and Nitin Singh

Abstract This paper proposes a modified optimal PIDD^2 controller for flexible-link manipulator. The single flexible link is modeled mathematically in which the flexible link and base rotation are modeled as stiff systems using Lagrange's method. The system obtained as a result will have one degree of freedom. In the proposed work, the comparison of two types of controller, i.e., PID and PIDD^2, is done for controlling the position and trajectory of the single-link manipulator. The main objective is to control the trajectory with minimum tip oscillation. The tuning of the controllers is done using the Ziegler–Nichols (Z-N) method and Dynamic Particle Swarm Optimization (DPSO) algorithm. The dynamic particle swarm optimization algorithm is an improved version of the particle swarm optimization algorithm which identifies and eliminates the dilemma of stagnation and local optima. The findings show that the PIDD^2 controller with dynamically tuned parameters is better in controlling the position and trajectory of the single-link manipulator. All the simulations were performed on MATLAB–SIMULINK.

Keywords Flexible-link manipulator · PID controller · PID plus second-order derivative controller · Ziegler–Nichols (ZN) · Particle Swarm Optimization (PSO) · Dynamic Particle Swarm Optimization (DPSO)

1 Introduction

The use of manipulators has massively increased in the space, defense, and medical industry in the twentieth century. The advantages offered by the single-link manipulators like high volume of work, lightweight, low energy consumption, small

---

K. Agrawal · R. Negi · N. Singh (✉)
Electrical Engineering Department, MNNIT, Allahabad, India
e-mail: nitins@mnnit.ac.in

K. Agrawal
e-mail: ag.komal80@gmail.com

R. Negi
e-mail: richa@mnnit.ac.in

actuators, better maneuverability, and transportability have made them widely used manipulators [1]. The main problem is that with the use of lightweight robots in the industries, structural flexibility has become a real challenge [2]. Sometimes the arm flexibility may lead to the tip oscillation of the link for the duration of the movement, which is highly undesirable. These oscillations have made control of such links more challenging and attracted the interest of various researchers worldwide to model and control the oscillations of the flexible-link manipulators.

A single-link manipulator is a link which is controlled by an actuator for fulfilling the specified objective, e.g., placing a load from one point to another. The nonlinearity is introduced in the dynamics of the single-link manipulator due to the flexibility and it increases the vibration and overshoot of the single-link manipulator. The rotary flexible single-link manipulator is shown in Fig. 1, and it comprises of the stainless steel flexible link. The complete assembly is mounted on the Quanser SRV02 plant. The motion of the link in horizontal plane is achieved by a DC motor connected to it. In order to sense the angular deflection of the tip, a strain gauge is installed at the motor end and an optical encoder is installed with the motor.

The research on the modeling and control of the flexible-link manipulators was started in 1970s [3], thereafter, various works have been published on this issue. There are various existing techniques to model the single-link manipulator, e.g., assumed mode method [4], the finite element method [5], or the Ritz expansion method [6], using the Lagrangian formulation or Newton-Euler formulation and Hamilton Principle [7]. Each of the modeling methods has got their own pros and cons related to complexity, adaptability, and accuracy. In the presented work, Lagrangian's method is used for deriving the motion equation of the flexible-link manipulator.

In order to control the flexible-link manipulators with dynamic deformation, the smart materials for actuators were proposed to be used by some researchers [8–10]. The torque wheel was mounted at the manipulator end point for suppressing the active vibration and correcting the position [11]. The elastic oscillations of the manipulator were eliminated by the application of fast and high-accuracy micromanipulator. There are some approaches that controlled the active vibration using the additional actuator [3, 12, 13]. Some researchers also suggested the joint-actuator method [14, 15] and joint manipulator method [28, 29] for suppressing the vibrations.

Fig. 1 Quanser flexible link

There are various researches who have proposed various strategies and establish the effectiveness of the state feedback controller and LQR controller for controlling the single-link manipulator [16, 17]. The effectiveness of the vibration controller with respect to the deflection feedback for controlling the flexible link is shown by Tso et al. [18]. Whereas, the use of linear optimal controller can achieve the improved control of single-link manipulator which is shown by Piedboeuf et al. [19].

In the proposed work, two different controllers are used for controlling the tip vibration of the single-link manipulator. The dynamic analysis of the single point flexible manipulator is presented using the state-space model. The parameters of the two controllers are tuned using two different techniques, i.e., Z-N method and dynamic particle swarm optimization algorithm.

This paper is arranged in various sections, Sect. 2 describes the mathematical modeling of the single flexible-link manipulator. Section 3 defines the problem, which describes the different control strategies designed for maintaining the accurate position of manipulator with minimum tip vibration. Section 4 discusses the methodology for finding the optimized values of controller's gain using Z-N and DPSO method. Comparison and discussion of simulation results are discussed in Sect. 5. In the last section, i.e., Sect. 6 the findings of the article are summarized.

2 Modeling the Flexible Link

2.1 Single-Link Manipulator System Model

Experimental setup of the single flexible-link manipulator consists of combination of electrical and mechanical components. The mechanical subsystem is modeled by stating the deflection of the tip deflection and rotation dynamics of the base. DC servomotor modeling is performed using electrical subsystem relating torque and voltage [20, 21].

2.2 Servomotor Model

Dynamics of servo system is developed using the block diagram as shown in Fig. 2. Block diagram provides the transfer function of servo system in the form of expression given in Eq. (1).

$$\frac{\theta(s)}{V_{in}(s)} = \frac{A_m}{s\left(s + \frac{1}{\tau_t}\right)} \tag{1}$$
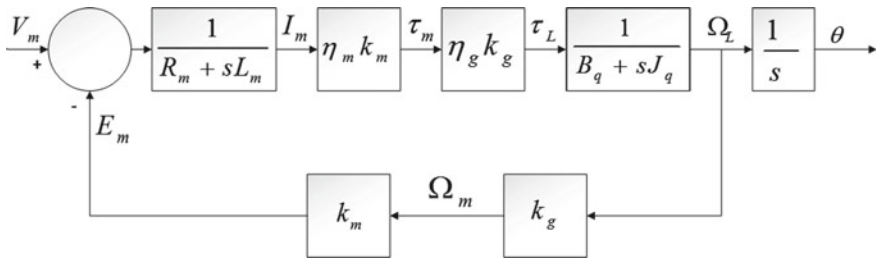
Fig. 2 Block diagram of SRV02 Servomotor

where, $\theta(s)$ = servomotor angle, $V_{in}(s)$ = Input voltage, A_m = motor gain and τ_t = time constant. The motor gain A_m can be further expanded as Eq. (2).

$$A_m = \frac{\eta_{mo}\eta_{ge}K_bK_{gt}}{R_{am}J_q} \tag{2}$$

where $\eta_m = \eta_{mo}$ = motor efficiency, $\eta_g = \eta_{ge}$ = gearbox efficiency, $R_m = R_{am}$ = motor armature resistance, J_q = high-gear moment of inertia, L_m = armature inductance, K_{gt} = high-gear to total gearbox ratio, $k_m = K_b$ = back-emf constant, and B_q = high-gear damping coefficient. The time constant can be written as Eq. (3).

$$\frac{1}{\tau_t} = \frac{B_q}{J_q} + \frac{\eta_{mo}\eta_{ge}K_b^2K_{gt}^2}{R_{am}J_q} \tag{3}$$

All the parameters in Eq. (3) are considered without any additional load [21]. The final torque can be expressed as Eq. (4).

$$\tau = \frac{\eta_{mo}\eta_{ge}K_bK_{gt}}{R_{am}}\left(V_{in} - K_bK_{gt}\Omega_L\right) \tag{4}$$

Ω_L = generated angular velocity by the servomotor at load end.

2.3 Mathematical Model

The schematic diagram of the flexible link is depicted in Fig. 3, where, $\theta(t)$ is the servo angle and $\alpha(t)$ is the deflection angle of the flexible link relative to under formed link of generalized variables [20]. The output angle and velocity of the DC servomotor be expressed as Eqs. (5) and (6), respectively.

$$y(t) = \theta(t) + \alpha(t) \tag{5}$$

Fig. 3 Flexible link
schematic

Fig. 3 Flexible link schematic
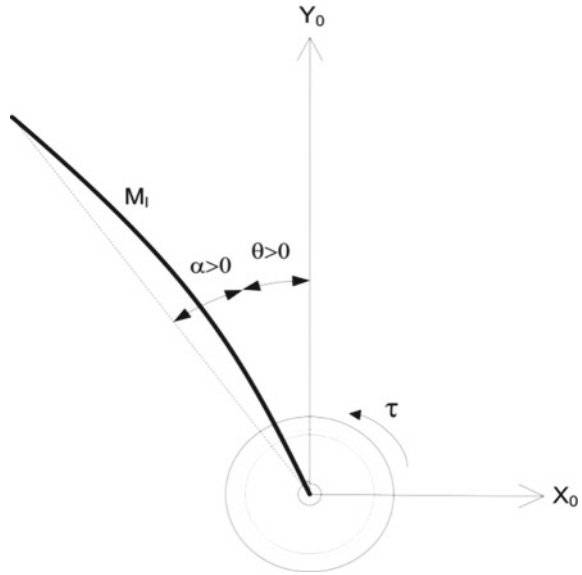
$$\dot{\theta}(t) = \omega \tag{6}$$

Now, the total potential and kinetic energy needs to be calculated in order to determine the system's Lagrangian. The total potential energy generated by gravitational force is neglected. So, the remaining equivalent potential energy is simply because of the elasticity of the flexible arm given as Eq. (7).

$$V = \frac{1}{2}\left[K\alpha^2(t)\right] \tag{7}$$

where K = torsion spring stiffness. The base and link rotation produce the kinetic energy which can be expressed as Eq. (8).

$$T = \frac{1}{2}J_q\left(\frac{d\theta}{dt}\right) + \frac{1}{2}J_{link}\left(\frac{d\theta}{dt} + \frac{d\alpha}{dt}\right)^2 \tag{8}$$

where J_{link} = moment of inertia of the flexible link. Generalized coordinates and viscous damping forces generate the nonconservative forces given as Eqs. (9) and (10).

$$Q_\theta = \tau - B_q\left(\frac{d\theta}{dt}\right) \tag{9}$$

$$Q_\alpha = B_{link}\left(\frac{d\alpha}{dt}\right) \tag{10}$$

where B_{link} = link viscous damping force, but, it is ignored, i.e., $B_{link} = 0$ and "τ" as shown in Eq. (4). Euler–Lagrange's equation can be expressed as Eq. (11).

$$\frac{\partial^2 L}{\partial t \partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = Q_i \tag{11}$$

The generalized coordinates are shown as $L = T - V$ and nonconservative forces. In order to obtain the state-space system, Eqs. (7), (8), (9), and (10) are put into Eq. (11) and it is solved. After solving it, obtained state-space system is given in Eq. (12).

$$\dot{x} = Ax + Bu \tag{12}$$

where

$$x = [\theta \; \alpha \; \omega \; \dot{\alpha}]^T$$

$$A =$$
$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & K / J_q & -\left(\eta_{ge}\eta_{mo}K_{gt}^2 K_b + B_q R_{am}\right) / R_{am} J_q & 0 \\ 0 & -K\left(J_q + J_{link}\right) / J_q J_{link} & \left(\eta_{ge}\eta_{mo}K_{gt}^2 K_b + B_q R_{am}\right) / R_{am} J_q & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ 0 \\ \left(\eta_{ge}\eta_{mo}K_{gt}K_t\right)/R_{am} J_q \\ -\left(\eta_{ge}\eta_{mo}K_{gt}K_t\right)/R_{am} J_q \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The numerical values of different variables of the system are given in Table 1.

Now, from the state-space form of the system, i.e., Eq. (12) and from the parameter values of the system shown in Table 1, the closed-loop transfer function of the single-link flexible manipulator system with respect to servo angle and deflection angle of the flexible link is given as Eqs. (13) and (14), respectively.

$$G_\theta = \frac{61.73s^2 + 2.741E - 014s + 1.927E004}{s^4 + 35.11s^3 + 974.1s^2 + 1.096E004 + 1.927E004} \tag{13}$$

$$G_\alpha = \frac{-61.73s + 1.569E - 013}{s^3 + 35.11s^2 + 850.6s + 1.096E004} \tag{14}$$

Table 1 Numerical values of the parameter [22]

Symbol	Description	Value
B_q	Viscous damping coefficient	4.00E-03 (Nm/(rad/s))
J_q	Moment of inertia	2.08E-03 (Kg)
K_t	Motor torque constant	7.680E-03 (Nm/A)
K	Torsion spring constant	1.2485
η_{mo}	Motor efficiency	0.69
η_{ge}	Gearbox efficiency	0.90
R_{am}	Motor armature resistance	2.6Ω
J_{link}	Moment of inertia	0.004 (Kg)
K_{gt}	Total gearbox rotation	70
K_b	Back-emf constant	7.68E-03 (V/(rad/s))

Fig. 4 State space model
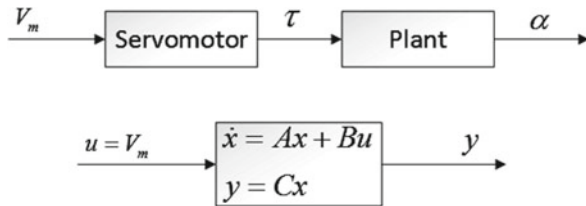
The final state-space model of the manipulator will be a bigger subsystem comprising of two subsystems, i.e., servomotor and flexible link. The final state-space model is shown in Fig. 4. Here, $V_m = V_{in}$ is the input voltage.

3 Problem Definition

Two control strategies are utilized for maintaining the accurate position of the manipulator and eliminate tip's vibration (during fast response), for a defined set of servo angle. The two strategies are as follows.

3.1 PID Controller

It is the most commonly used structure of feedback controllers. PID controllers are used by most of the control loops in the process control. It is normally utilized at the low level and the set point to these lower level controllers is given by the multivariable controller. The PID controller is also called as "bread and butter" of
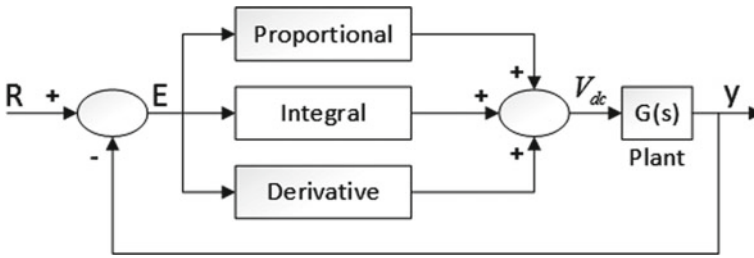
Fig. 5 PID controller block diagram

the control engineers. In the PID controller three variables are presents, i.e., $K_p =$ Proportional gain, $K_i =$ Integral gain, and $K_d =$ Derivative gain. The above-defined variables are used for providing the controlled output of the PID controller. The controlled output variable known as "u" of PID can be written as Eq. (15). The block diagram of the PID controller is shown in Fig. 5.

$$u(t) = \left[K_p e(t) + K_i \int e(t)dt + K_d de(t)/dt \right] \tag{15}$$

Error between the desired set point and measured variable is corrected using a PID controller. Steady-state error is decreased by the use of proportional controller. But by no means it can manage to eliminate the steady-state error. Now for removing the steady-state error, proportional plus integral (PI) controller is used. But it also has a negative impact on speed of response and overall stability of the system. It does not have the ability to predict the future errors of the system. Proportional Plus Derivative (PD) controller has an ability to predict the future error of the system response. PID controller has zero steady-state error, fast response (short rise time), no oscillations, and higher stability.

3.2 PIDD^2 Controller

Its function is similar to the PID controller, except that it has got an extra derivative square term that plays an important role in reducing the vibration error caused due to oscillation and reducing the response settling time of the rotary flexible link. The gain equation of the PID plus second-order derivative controller can be written as Eq. (16). The block diagram of the PIDD^2 controller is shown in Fig. 6.

$$u(t) = \left[K_p + K_i/s + K_{d1}s + K_{d2}s^2 \right] \tag{16}$$

where K_p, K_i, K_{d1}, and K_{d2} are the proportional, integral, derivative, and second-order derivative gains, respectively. In Eq. (16), it is clearly seen that a term is added. It is well known that the derivative control improves the transient response. So, the
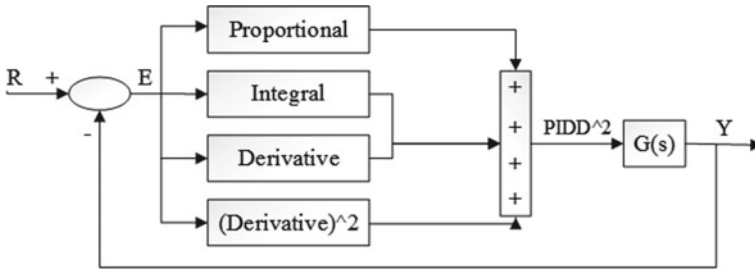
Fig. 6 PIDD^2 controller block diagram

peak overshoot and transient present in the vibration of the rotary flexible link are reduced more effectively.

4 Tuning of Controllers

The system response becomes unstable for incorrect parameter values of the PID controller. So, for finding the optimum values of controller, the parameter tuning is done. In the proposed work, the parameters of the PID and PIDD^2 controllers are tuned using the Ziegler–Nichols (ZN) method and Dynamic Particle Swarm Optimization (DPSO) algorithm.

4.1 Ziegler–Nichols Method

It is a prying method of tuning. Initially, it takes integral (I) and derivative (D) gains as zero. Now, increase the proportional (P) gain K_p, at which the output of the system has consistent and stable oscillations. The settling time of the P controller is taken as extreme period (T_e). Both K_e and T_e are used to find the optimal values of the PID controller gain.

4.2 Dynamic Particle Swarm Optimization (DPSO)

The new variant of particle swarm optimization was proposed in [23] which overcomes the two major goals of the population-dependent algorithm, i.e., the rate of the convergence and the technique for eliminating the problem of local minima, which is complicated to accomplish as both the objectives are contradictory to each other. The fundamental PSO [24] makes use of the global best (gBest) for updating the velocity of the particles and simultaneously it also affects the position of the particles resulting

in a rapid convergence, because of which it becomes susceptible to the dilemma of local optima especially in the case of multimodal systems [25, 26]. Various variants have already been proposed by the researchers to improve the performance of the PSO and provide solution to this problem by balancing the two parameters [23, 27].

The dynamic PSO (DPSO) variant reduces the difficulty of stagnation and local optima and simultaneously tracks and delivers the rapid convergence rate. DPSO achieves the objective of fast convergence rate without getting trapped into local optima problem by tracking and recording the modification in the individual best positions attained by the particles, i.e., pBest. The DPSO algorithm records and tracks the duration for which the position pBest has not changed and simultaneously it records and tracks the number of iterations during which the gBest had not been updated. If there is no change in the pBest and gBest for a specified number of iterations, then the current values of both the parameters are changed with the best values that they have attained previously at some point of time and iteration.

The pBest and gBest are again monitored for any changes in their values that might have taken place due to the changed values and it was adjudged that with the replaced values whether they are able to get any improvements in the target values or not. If any positive improvement is observed in the target values the replaced values are made permanent otherwise the change is reversed and the old values are restored in the algorithm. This process is repeated until the optimal value is achieved or the termination condition is reached. The objective function of the system is given as Eq. (17) and is used for optimization.

$$J = \int \left[(R - \theta)^2 + (\alpha)^2 \right] dt \qquad (17)$$

Here, the Integral of Square Error (ISE) performance parameter is used to calculate the optimal values of the controller gains for minimizing the error performance index.

4.3 Dynamic Particle Swarm Optimization Procedure

The various steps involved in tuning the parameters of PID and PIDD^2 controllers are given below [23]. The Dynamic PSO algorithm (DPSO) procedure involves mainly three steps, i.e., Step-(i), the identification and preservation of the potential region which are possibly unexplored, Step-(ii), restructuring pBest and gBest of the stagnated particles for whom the value of pBest and gBest has not changed for a long time and Step-(iii), monitoring the acceptability of the changes made in the values of pBest and gBest. The DPSO procedure is explained through flowchart shown in Fig. 7. The terminology used in DPSO algorithm is shown in Table 2. The parameters of the Dynamic PSO are initialized as follows:

- Constant inertia weight (w) is taken as 0.783.
- The acceleration coefficients c_1 and c_2 are taken as $c_1 = c_2 = 1.367$.
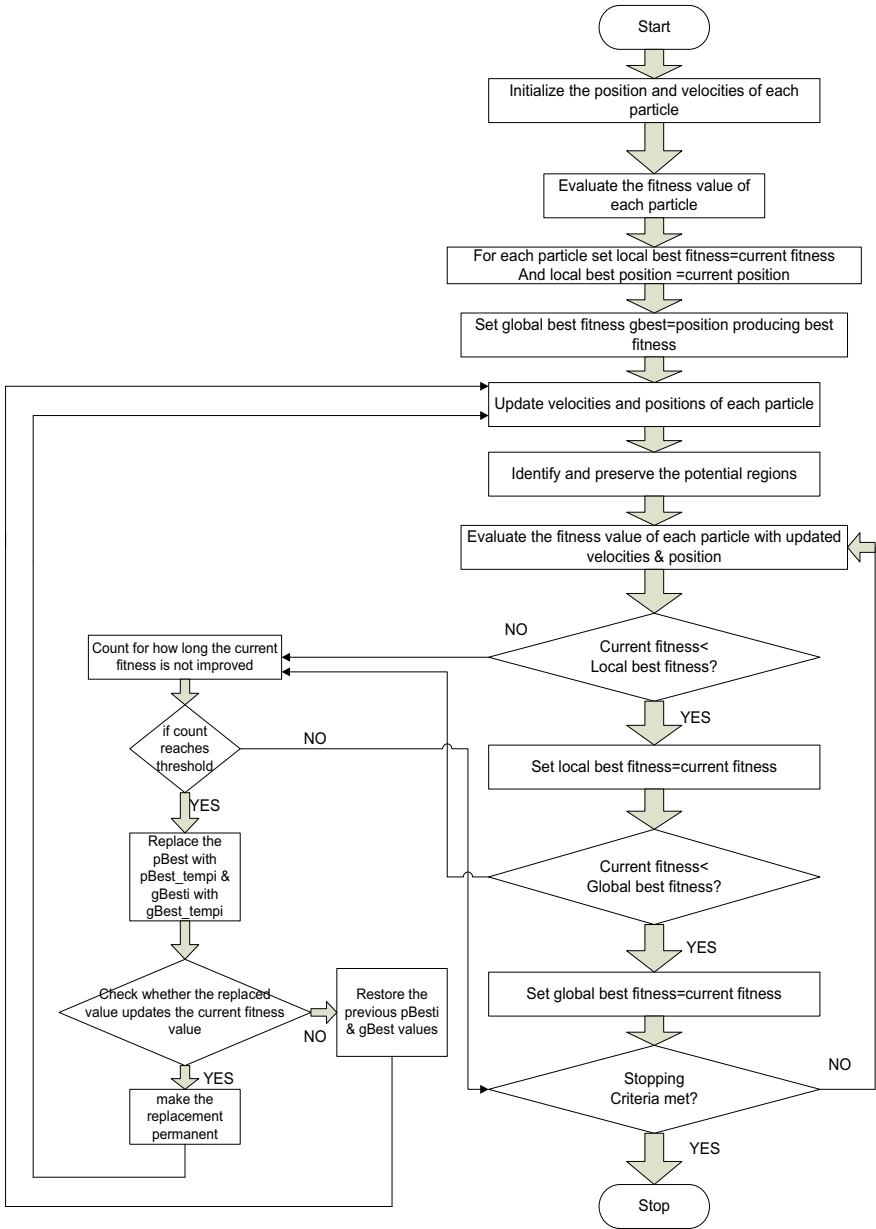
Fig. 7 Flowchart showing the Dynamic Particle Swarm Optimization (DPSO) procedure

Table 2 Dynamic particle swarm optimization terminology

Symbol	Description
M	No. of particles in Swarm (population size)
X_i	Position vector of the ith particle, i = 1 to m
V_i	Velocity vector of the ith particle, i = 1 to m
pBest$_i$	Best position obtained by ith particle
gBest	Global best position of the swarm
pBest_count$_i$	Counter when pBest$_i$ do not get better in a successive generation for particle i
pBest_iter_thresh	Utmost number of iteration ith particle wait for an improvement in pBest$_i$
pBest_temp$_i$	Provisionally restructured vector of pBest$_i$ for particle I, i = 1 to m
pBest_chance_count$_i$	Counter when pBest_temp$_i$ do not improve in a successive generation for particle i
pBest_chance_max	Maximu number of iterations pBest_temp$_i$ will get chance to improve pBest$_i$ (constant)
gBest_h	Vector for historical values of gBest
gBest_h_maxcount	Maximum permissible number of elements in gbest_h (constant)
gBest_count	Counter when gBest do not improve in a successive generation for particle i
gBest_iter_thresh	Maximum number of iterations group wait for an improvement in gBest (constant)
gBest_temp	Provisionally restructured vector of pBest$_i$ for particle i = 1 to m
gBest_chance_count	Counter when gBest_temp do not improve in gBest_chance_max successive iterations
gBest_chance_max	Maximum number of iterations gBest_temp will get chance to improve gBest (constant)

- Max number of gBest stored as history in gBest_h = 10*m, pBest_iter_thresh = gBest_iter_thresh = 5*m (dimension).

Initialization

The numbers of particles containing the PID and PIDD^2 controller parameters are initialized randomly within search space, here, number of particles is taken as 50. Counter for pBest_count$_i$ and pBest_chance_count$_i$ (for i = 1 to m) are initialized to 0 and –1, respectively, these counts keep a track of the succeeding generations for which pBest$_i$ and pBest_temp$_i$ have not improved. Similarly, gBest_count and gBest_chance_count are set to 0 and –1, respectively, these counts keep a track of the number of iterations for which gBest and gBest_temp have remain unchanged.

PBest, GBest and Fitness Calculation

The Integral Square Error (ISE) shown in Eq. (17) is taken as the objective function for calculating the optimal values of the gains, i.e., K_p, K_i, K_{d1}, and K_{d2} of the PIDD^2 controller, where R is the reference of the deflection, θ is the actual value of the servo

angle, and α is the deflection angle. The fitness of each particle is determined; the value of initial position vector is assigned as pBest$_i$ (i = 1 to m) and best position produced as gBest.

Updating Velocity and Position
The velocity and position vector, i.e., v$_i$ and x$_i$ (i = 1 to m), respectively, of the individual particle are updated using Eqs. (18) and (19), respectively.

$$v_i^{k+1} \Leftarrow w \times v_i^k + c_1\tau_1(pxbest_i - x_i^k) + c_2\tau_2(gxbest - x_i^k) \qquad (18)$$

$$x_j^{k+1} = x_j^k + v_j^{k+1} \qquad (19)$$

Identify and Preserve Potential Regions
The potential regions for the most part obtained in the form of gBest up to count gBest_h_maxcount are stored as historical global best position in gBest_h.

Updating PBest and GBest
Fitness of the individual particle is computed again with the updated position, if the new position of particle x$_i$ is able to make better the personal best position of the particle pBest$_i$ (for i = 1 to m), it is updated and similarly if the current global best position gBest of the particle is better than previous it gets updated.

Controlling Stagnation Problem
The count pBest_count$_i$ and gBest_count$_i$ (for i = 1 to m) are increased by 1 in each generation and if the pBest$_i$ and gBest have not improved from the previous values they are reset to 0. If pBest_count$_i$ and group gBest_count attain the threshold value for every particle i (for i = 1 to m), i.e., pBest_iter_thresh and gBest_iter_thresh, respectively, then pBest and gBest are restructured; otherwise, the termination condition of the algorithm is tested.

Restructuring PBest and GBest
The particles whose pBest_counti and gBest_count has crossed the threshold limit, their pBesti and gBest are restructured with the new values from the history as pBest_tempi and gBest_temp, respectively.

Checking Acceptability of PBest and GBest
The particles whose personal best, i.e., pBesti, is replaced with pBest_tempi are checked for the maximum number of iterations given as pBest_chance_max, if the value of pBesti further improves from the previous value, then the replaced value is accepted as fixed otherwise old values are reinstated and replaced with a new pBest_tempi until an improved value of the pBesti is obtained. Similarly, gBets replaced with gBest_temp is checked for an improved value then the previous gBest, if it is obtained then the change is taken as permanent otherwise original value is restored.

Termination Condition

If the evaluation of the function has exceeded the maximum allowed evaluation number, then the algorithm terminates. Otherwise, the value of velocity and position vector is again updated and all the steps are repeated until the termination condition is reached.

5   Simulation Results and Discussion

In this section, after using Table 1 parameters, respectively, simulation has been performed and the obtained results are mentioned in Figs. 8, 9, 10, and 11. Optimal values of the PID and PIDD^2 controllers were obtained for minimizing the cost function (objective function) as mentioned in Table 3.

Here, the proposed PIDD^{2}controller is tested for controlling the single-link flexible manipulator system defined in Eq. (12). The performances of conventional PID controller and PIDD^{2}controller, tuned with the Ziegler–Nichols (Z-N) and Dynamic Particle Swarm Optimization (DPSO) techniques, are compared. The realization of PID and PIDD^2 controller is done in the MATLAB. For tuning the PID controller and PIDD^2 controller gain parameter, DPSO algorithm uses ISE performance criterion.

Figure 8 shows servo angle (θ) which is generated by the servomotor torque at the joint of motor and the link. Figure 9 represents the deflection of the flexible link angle (α) at the tip of the flexible link (vibration). Figures 10 and 11 show the response
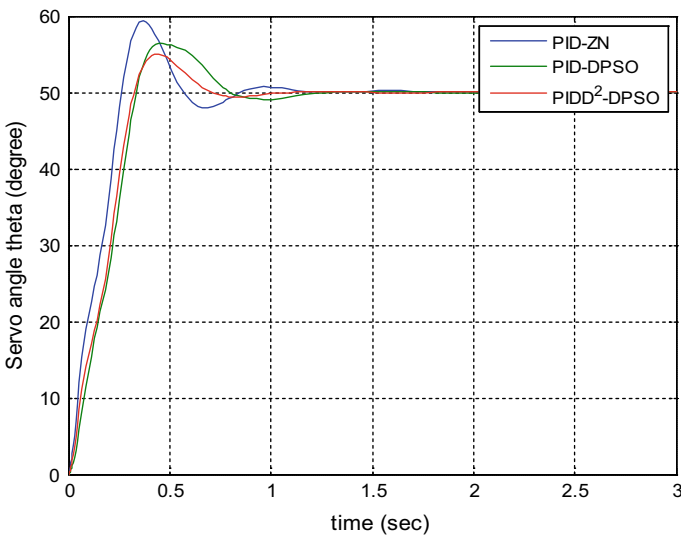
Fig. 8 Deflection of servo angle (θ)
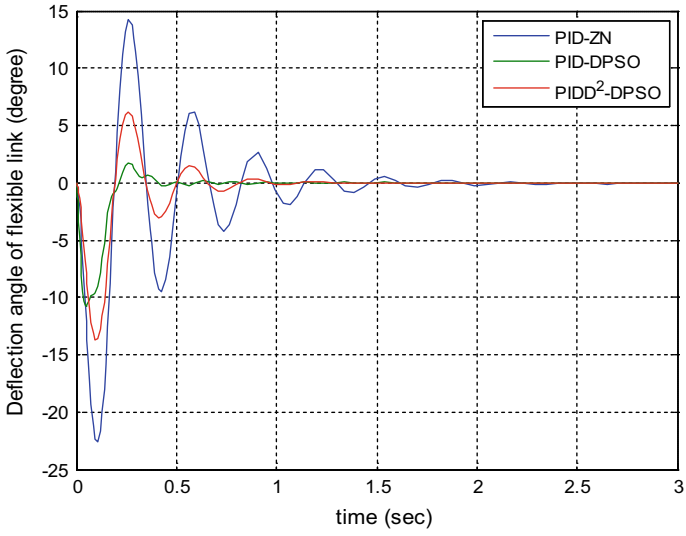
Fig. 9 Tip deflection angle (α)
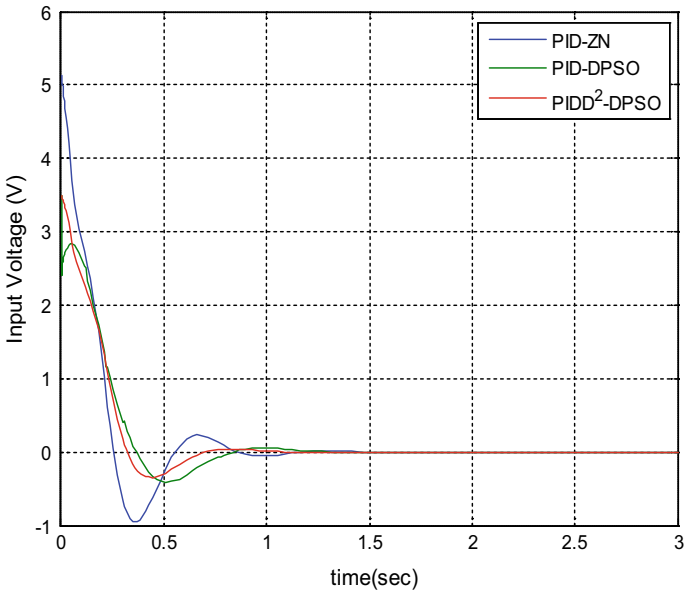
Fig. 10 Input voltage (V) for servo angle (θ)
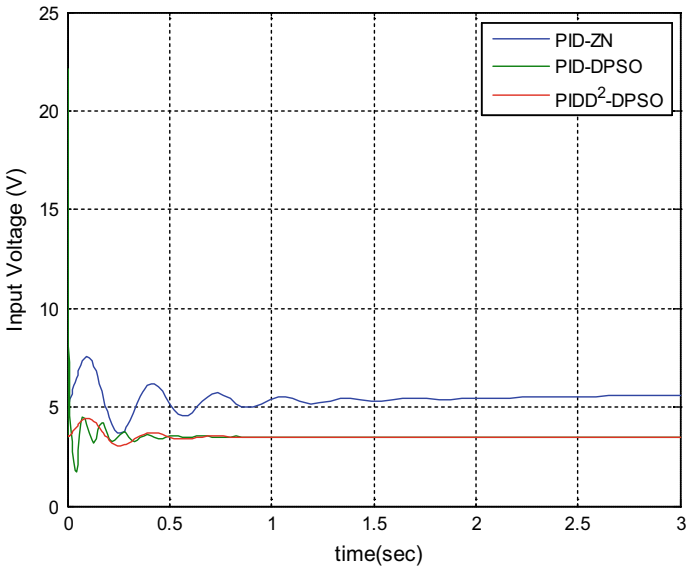
Fig. 11 Input voltage (V) for tip deflection angle (α)

Table 3 Optimal values of the controller gain

Controller gain	K_p	K_i	K_d
PID-ZN	5.88	0.1779	0.0445
PID-DPSO	4.19532	0.0017689	0.015248
PIDD^2-DPSO	3.98925	0.00147624	$\mathbf{K_{d1}} = 0.010548$ $\mathbf{K_{d2}} = 0.029158$

of the controlled input (voltage) for servo angle and deflection angle, respectively, which is given to plant.

After close observation of the results, it becomes evident that the oscillations and settling time are reduced significantly for all the responses except the tip deflection angle response. Figures 10 and 11 show the variation of input voltage given to the servo motor for controlling the servo angle (θ) and tip deflection angle (α), it signifies that operation of the motor will be stable and smooth for the PIDD^2 controller which is tuned using DPSO algorithm as compared to the other tuning methods.

6   Conclusion

In this paper, a comparative study of PID and PIDD^2 controller is done, in order to control the dynamics of the single-link flexible manipulator. Proposed PIDD^2 controller

has four control terms, proportional, integral, derivative, and second derivative. For tuning the gains of the PID and PIDD^2 controllers, Z-N method and DPSO algorithm are used which use integral of square error (ISE) as the performance criterion. The dynamic performance of single-link flexible manipulator with DPSO-PIDD^2 is compared with the Ziegler–Nichols (ZN)-tuned PID- and DPSO-tuned PID controller. Simulation results confirm the superior performance of the PIDD^2 controller.

References

1. Stieber, M.E., McKay, M., Vukovich, G., Petriu, E.: Vision-based sensing and control for space robotics applications. IEEE Trans. Instrum. Meas. **48**(4), 807–812 (1999)
2. Ahmad, M.A., Mohamed, Z.: Modelling and simulation of vibration and input tracking control of a single-link flexible manipulator. Pertanika J. Sci. Technol. **18**(1), 61–76 (2010)
3. Book, W.J.: Modeling, Design and Control of Flexible Manipulator Arms. Massachusetts Institute of Technology (1974)
4. Dalong, W., Youfang, L., Yan, L., Xiaoguang, L.: Dynamic model and tip trajectory tracking control for a two-link flexible robotic manipulator, vol. 2, pp. 1020–1024 (1996)
5. Ge, S.S., Lee, T.H., Zhu, G.: A nonlinear feedback controller for a single-link flexible manipulator based on a finite element model. J. Robot. Syst. **14**(3), 165–178 (1997)
6. Nicosia, S., Valigi, P., Zaccarian, L.: Dynamic modelling of a two link flexible robot and experimental validation, vol. 3, pp. 1953–1958 (1996)
7. Low, K.H., Vidyasagar, M.: A lagrangian formulation of the dynamic model for flexible manipulator systems. J. Dyn. Syst. Meas. Contr. **110**(2), 175 (1988)
8. Kakatcioglu, S., Giray, M., Asmer, H.: Vibration control of flexible manipulators using smart structures, pp. 415–420 (1997)
9. Tzou, H.S.: Integrated distributed sensing and active vibration suppression of flexible manipulators using distributed piezoelectrics. J. Robot. Syst. **6**(6), 745–767 (1989)
10. Sun, D., Mills, J.K.: Study on piezoelectric actuators in control of a single-link flexible manipulator, vol. 2, pp. 849–854 (1999)
11. Khorrami, F., Gomez, A.A., Hills, M.D.: Utilization of torque wheels for active damping of flexible manipulators, pp. 1816–1821 (1994)
12. Sharon, A., Hogan, N., Hardt, D.E.: The macro/micro manipulator: an improved architecture for robot control. Robot. Comput.-Integ. Manuf. **10**(3), 209–222 (1993)
13. Ballhaus, W.L., Rock, S.M.: End-point control of a two-link flexible robotic manipulator with a mini-manipulator: Initial experiments. In: American Control Conference, pp. 2510–2514 (1992)
14. Rossi, M., Zuo, K., Miang, D.: Issues in the design of passive controllers for flexible link robots, pp. 321–326 (1994)
15. Trautman, C., Wang, D.: Non-collocated passive control of a flexible link manipulator, vol. 2, pp. 1107–1114 (1996)
16. Ge, S.S., Lee, T.H., Zhu, G.: Improving regulation of a single-link flexible manipulator with strain feedback. IEEE Trans. Robot. Autom. **14**(1), 179–185 (1998)
17. Baroudi, M., Saad, M., Ghie, W.: State-feedback and linear quadratic regulator applied to a single-link flexible manipulator, pp. 1381–1386 (2009)
18. Tso, S.K., Yang, T.W., Xu, W.L., Sun, Z.Q.: Vibration control for a flexible-link robot arm with deflection feedback. Int. J. Non-Linear Mech. **38**(1), 51–62 (2003)
19. Piedboueuf, J.-C., Dochain, D., Hurteau, R., Benameur, K.: Optimal control of the tip of a flexible arm. In: Engineering Institute of Canada, Canadian Conference on Electrical and Computer Engineering, vol. 1, p. 24 (SEE N 94–22101 05-33) (1991)

20. Gurses, K.: Dynamic modeling and vibration control of a single-link flexible manipulator using a combined linear and angular velocity feedback controller. University of Victoria (2007)
21. Zain, B.A.M., Tokhi, M.O., Toha, S.F.: PID-based control of a single-link flexible manipulator in vertical motion with genetic optimisation, pp. 355–360 (2009)
22. Baroudi, M., Saad, M., Ghie, W., Kaddouri, A., Ziade, H.: Vibration controllability and observability of a single-link flexible manipulator, pp. 1–6 (2010)
23. Saxena, N., Tripathi, A., Mishra, K.K., Misra, A.K.: Dynamic-PSO: an improved particle swarm optimizer, pp. 212–219 (2015)
24. Eberhart, R.C., Kennedy, J.: A new optimizer using particle swarm theory. Proc. Sixth Int. Symp. Micro Mach. Human Sci. **1**, 39–43 (1995)
25. Seo, J.H., Im, C.-H., Heo, C.-G., Kim, J.-K., Jung, H.-K., Lee, C.-G.: Multimodal function optimization based on particle swarm optimization. IEEE Trans. Magn. **42**(4), 1095–1098 (2006)
26. Liang, J.J., Qin, A.K., Suganthan, P.N., Baskar, S.: Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. IEEE Trans. Evol. Comput. **10**(3), 281–295 (2006)
27. Shi, Y., Eberhart, R.: A modified particle swarm optimizer, pp. 69–73 (1998)
28. Millian, A.S., Feliu, V., Garcia, A.: A two stage control scheme of single-link flexible manipulators. In: 23rd Mediterranean Conference on Control and Automation (MED), June 16–19, 2015, Torremolinos, Spain (2015)
29. Yan, Z., Lai, X., Meng, Q.: Stable control of single-link flexible-joint manipulator. In: Proceedings of the 37th Chinese Control Conference, July 25–27, 2018. Wuhan, China (2018)

Design of an Optimal Microstrip Butterworth Low-Pass Filter Using Colliding Bodies Optimization

Mridul Gupta, Mayank Kansal, Shriram Thyagarajan, Prajjwal Singh Chauhan and D. K. Upadhyay

Abstract A new methodology to design a microwave filter is discussed. The filter designed is a Butterworth low-pass filter of order five which is obtained in digital domain using chain scattering parameters for serial transmission lines and a shunt-connected open-circuited stub. Colliding Bodies Optimization (CBO) algorithm is used for approximating the magnitude response of desired transfer function with the ideal one. Superiority of the results obtained from CBO in comparison to some widely used optimization algorithms, viz., real-coded genetic algorithm, gravitation search algorithm, and particle swarm optimization is shown. It is seen that CBO outperforms the other algorithms in terms of magnitude error parameters with passband and stopband error values of -43.62 dB and -48.205 dB, respectively. Afterward, the obtained configuration is then simulated on ADS using the substrate Rogers RO3003 having permittivity value of 3 and 0.75 mm thickness. The microstrip Butterworth low-pass filter is found to have 3.4 GHz cutoff frequency which is capable of covering various microwave applications.

Keywords Low-pass filter · Line elements · Microwave filter · Optimization

M. Gupta (✉) · M. Kansal · S. Thyagarajan · P. S. Chauhan · D. K. Upadhyay
Division of Electronics and Communication Engineering, Netaji Subhas Institute of Technology, Sector 3, Dwarka, New Delhi, India
e-mail: mri.gupta@gmail.com

M. Kansal
e-mail: mayank.kansal23@gmail.com

S. Thyagarajan
e-mail: shriram365@hotmail.com

P. S. Chauhan
e-mail: mail2prajj@gmail.com

1 Introduction

A Microwave Low-Pass Filter (LPF) is used to only pass specific lower frequency components up to the cutoff frequency and attenuate the higher frequency harmonics. Microwave LPFs find wide range of applications in communication systems such as, Bluetooth, Wi-Fi, WiMAX, GSM, navigation systems like GPS, GNSS, and GLONASS in the bands ranging from 1.2 to 1.9 GHz, bio-medical applications, many S-band applications like Direct-to-Home satellite television, powering the particle accelerators, airport surveillance radars, imaging systems, in receivers and transmitters of many wireless systems, and to improve signal-to-noise ratio in a communication link [1–4].

Commonly used microwave LPFs are designed using Butterworth approximation. The magnitude response of these type of filters is observed to be maximally flat in the passband region for a given filter order and complexity [4]. Along with the maximally flat passband response, a Butterworth Low-Pass Filter (BLPF) gives an attenuation equal to $-120 \times n$ dB/decade. The magnitude response of an nth order ideal BLPF is given as

$$|H_i(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}}} \tag{1}$$

where ω_c is the cutoff frequency with both ω and ω_c having units of radians/second, and n denotes the order of filter.

Various techniques are identified for designing microwave LPF circuits. Hsue et al. [5] have designed a microstrip LPF in Z-domain using chain scattering parameters of multi-section serial transmission lines and shunt-open-circuited stubs. Most commonly used method to design microstrip BLPFs is the stepped impedance method, wherein LC ladder network is approximated in the form of alternate high impedance (serial inductor) and low impedance (shunt capacitor) lines as has been shown in [6–9]. Transformed radial stubs were used in [10] to propose LPF microstrip circuits. Tri-section stepped impedance resonators were used to design a 1 GHz cutoff frequency LPF [11]. Liu et al. [12] have used stepped impedance hairpin units to propose a microstrip LPF with a 3 dB cutoff frequency of 1.6 GHz. Hayati and Shama [13] proposed a microstrip LPF using stepped impedance resonator having ultrawide stopband. Authors in [14], used cylindrical- and radial-shaped resonator to design a very compact microwave LPF. Oraizi et al. [15] successfully designed microwave LPF using Particle Swarm Optimization (PSO) technique.

In this work, an efficient metaheuristic optimization technique known as Colliding Bodies Optimization (CBO) [16] has been employed to find out the optimal impedance values of transmission line sections to design a 5th order BLPF in Z-domain. Additionally, the filter design has also been carried out using three more widely used algorithms, namely, Gravitation Search Algorithm (GSA) [17], Real-Coded Genetic Algorithm (RCGA) [18], and PSO [19] to prove the superiority of

CBO over others. The configuration of the most optimal design obtained from CBO has been simulated on Advanced Design System (ADS) software and the 3 dB cutoff frequency of the microstrip LPF is found to be 3.4 GHz.

This work has been organized into the following sections. In Sect. 2, design of the Butterworth LPF is discussed. Also, the detailed analysis of CBO is presented and the obtained results are compared with other metaheuristic algorithms. The obtained configuration for microstrip BLPF is simulated on ADS in Sect. 3. Finally, the work is concluded in Sect. 4.

2 Design of Butterworth Low-Pass Filter

2.1 Problem Formulation

Implementation of microwave filters using chain scattering matrices (T-matrices) of transmission line sections in Z-domain was shown by Chang and Hsue [20]. Based on this approach an LPF of order $n = 5$ was designed using multi-section serial and shunt-open transmission lines [5]. It is noticed that an open-circuited shunt stub provides a zero at $z = -1(\omega = л)$ [21] in Z-domain, which in turn is responsible for giving the behavior of an LPF. In this paper, we have achieved a much simpler design of a BLPF of order $n = 5$, by making use of a single shunt-connected open stub along with cascaded serial transmission line sections. For serial transmission line and shunt-connected open stub, T-matrices are, respectively, given by Eqs. (2) and (3) [20].

$$T_S = \frac{1}{z^{-1/2}(1 - \Gamma^2)} \begin{bmatrix} 1 - \Gamma^2 z^{-1} & -(\Gamma - \Gamma z^{-1}) \\ \Gamma - \Gamma z^{-1} & -\Gamma^2 + z^{-1} \end{bmatrix} \quad (2)$$

$$T_{OC} = \frac{1}{(1 + z^{-1})} \begin{bmatrix} (1 + a) - (1 - a)z^{-1} & (a + az^{-1}) \\ -a - az^{-1} & (1 - a) + (1 + a)z^{-1} \end{bmatrix} \quad (3)$$

where $z = e^{jw}$, $\Gamma = (Z_1 - Z_0)/(Z_1 + Z_0)$ and $a = Z_0/2Z_2$. Here, Z_0 is the reference characteristic impedance, Z_1 denotes the characteristic impedance of serial transmission line and Z_2 is characteristic impedance for shunt-open stub. The product of the T-matrices of the line sections gives us the overall transfer function. After this, T_{11} of the overall cascaded T-matrix for the network is computed and transmission coefficient is given by

$$H_p(z) = \frac{1}{T_{11overall}(z)} \quad (4)$$

To obtain the impedances of line elements forming the network, an error function based on Least Mean Square (LMS) is formulated on the above defined transmission coefficient and the ideal magnitude response of a BLPF which is given as

$$E(\omega) = \sum_{\omega} \left[|H_p(\omega)| - |H_i(\omega)| \right]^2 \tag{5}$$

The different designs are compared on the basis of their passband error (E_{pass}) and stopband error (E_{stop}) which are, respectively, given by (6) and (7) as

$$E_{pass} = 20 \log_{10} \sqrt{\frac{\sum_{\omega} [|H_p(\omega)| - |H_i(\omega)|]2}{P}} \tag{6}$$

where P = 500 and $0.01 \le \omega \le 1$.

$$E_{stop} = 20 \log_{10} \sqrt{\frac{\sum_{\omega} \left[|H_p(\omega)| - |H_i(\omega)| \right]2}{P}} \tag{7}$$

where P = 500 and $1 < \omega \le \pi$.

In this work, CBO has been used for minimizing the error function defined in Eq. (5). CBO is found to give the best magnitude response and is computationally superior in achieving fastest convergence of iteration cycles as compared to many existing benchmark metaheuristic algorithms [22].

2.2 Colliding Bodies Optimization Algorithm

CBO is a new age nature-inspired metaheuristic optimization technique proposed by Kaveh and Mahdavai in 2014 [16]. A collision happens between a physical pair of objects in this optimization technique to find the best solutions. Using a random search and simplified formulation, CBO gains a decisive edge over other optimization algorithms in terms of accuracy of results. The procedure of CBO is explained below.

A population of "n" individual solutions are considered which are called CBs where the ith CB is denoted by $X_i = [x_{1i}, x_{2i}, \ldots, x_{Di}] = [a_{1i}, a_{2i}, \ldots, a_{N+1i}, b_{1i}, b_{2i}, \ldots, b_{N+1i}]$ where i $= 1, 2, \ldots,$ n and D $= 2 \times (N+1)$ represent the problem's dimension.

Step 1: The CBs are generated and initialised in an arbitrary fashion within the problem search space given by the equation

$$x_i^0 = x_{\min} + \text{rand} \times (x_{\max} - x_{\min}), \quad i = 1, 2, \ldots, n \tag{8}$$

where $x_{\min}$ is the minimum impedance which is chosen as $10\,\Omega$ and $x_{\max}$ is the maximum impedance chosen as $150\,\Omega$, and rand is a number randomly chosen in the range [0, 1].

The fitness values are then evaluated and the minimum value is termed $gbest$.

Step 2: The kth CB's magnitude is found out to be

$$m_k = \frac{1/fit(k)}{\sum_{i=1}^{n} 1/fit(i)}, \quad i = 1, 2, \ldots, n \tag{9}$$

where $fit(i)$ denotes the value of i in the objective function and n shows the size of population.

Step 3: CBs are sorted in ascending order based on their value of fitness function. Two equal groups of these CBs are then formed. Group 1 consists of higher body mass CBs and are considered stationary and group 2 consists of lighter body mass and are termed as moving CBs. Moving CBs are then made to undergo collision in 1-d with respective stationary CBs.

Step 4: Before colliding, the velocities associated with the stationary and moving CBs are given as follows:

- Stationary CBs

$$v_i^s = 0, \quad i = 1, 2, \ldots, n/2 \tag{10}$$

- Moving CBs

$$v_i^m = x_i - x_{i-n/2}, \quad i = n/2 + 1, \ldots, n \tag{11}$$

where v_i^s is the velocity of the ith stationary CB, v_i^m is the velocity, and x_i is the co-ordinate of the ith moving CB.

Step 5: The after collision velocities and positions are given as

- Stationary CBs:

$$v_i^{snew} = \frac{m_{i+\left(\frac{n}{2}\right)} + \epsilon \times m_{i+\left(\frac{n}{2}\right)}}{m_i + m_{i+\left(\frac{n}{2}\right)}} \times v_{i+\left(\frac{n}{2}\right)}^m, \quad i = 1, 2, \ldots, n/2 \tag{12}$$

$$x_i^{snew} = x_i + \text{rand} \times v_i^{snew}, \quad i = 1, 2, \ldots, n/2 \tag{13}$$

- Moving CBs:

$$v_i^{mnew} = \frac{m_i - \epsilon \times m_{i-\left(\frac{n}{2}\right)}}{m_i + m_{i-\left(\frac{n}{2}\right)}} \times v_{i+\left(\frac{n}{2}\right)}^m, \quad i = n/2 + 1, \ldots, n \tag{14}$$

$$x_i^{mnew} = x_{i-\frac{n}{2}} + \text{rand} \times v_i^{mnew}, \quad i = n/2 + 1, \ldots, n \tag{15}$$

where v_i^{mnew}, x_i^{mnew} and v_i^{snew}, x_i^{snew} are the velocities and positions of the moving and stationary CB, m_i post collision, and the COR, $\epsilon = 1 - \frac{presiter}{maxiter}$. Here, presiter is the current iteration number and maxiter is the set number of maximum iterations.

Step 6: The fitness values with the updated positions of all the CBs are calculated and *gbest* is revised if necessary.

Step 7: Before declaring *gbest* as the most optimal design, repeat the process from Step 2 until the iterations complete.

2.3 Proposed Design and Comparisons

An optimal design for the microstrip low-pass filter is obtained using CBO. Order five of the BLPF is achieved by choosing a configuration consisting of five line elements, viz., a shunt-connected open stub placed at second place from the left and four serial transmission line sections in cascade. Here, superiority of the design is verified by comparing with the results obtained from some widely used evolutionary optimization algorithms in the literature, namely, GSA [17], RCGA [18], and PSO [19]. For the cutoff frequency of $\omega_c = 1$ rad/s, simulations are carried out. For all four algorithms used, control parameters are taken as mentioned in Table 1. The impedance range during the optimization process is chosen from $10\,\Omega$ to $150\,\Omega$ for ensuring practically realizable widths and lengths, and each algorithm is run for 500 iterations. Values of the impedances for designs obtained through different optimization algorithms and their corresponding E_{pass} and E_{stop} values are listed in Table 2.

From the results shown in Table 2, it is seen that the design D1, obtained through CBO has improved values for E_{pass} and E_{stop} which proves its superiority over GSA, RCGA, and PSO. Furthermore, the magnitude response comparison of BLPF designs

Table 1 Control parameters for different optimization algorithms

Algo.	Control parameter	Value	Algo.	Control parameter	Value
CBO	Population size	20	GSA	Gravitation constant	100
				Coefficient of decrease	20
				Number of agents	50
RCGA	Population size	100	PSO	Inertia weight	1
				Inertia damping ratio	0.99
	Crossover percentage	0.7			
	Mutation percentage	0.3			
	Mutation rate	0.1			
	Selection type: RW			Personal learning coefficient	1.5
				Global learning coefficient	2
				Swarm size	100

Table 2 Values of the impedances for designs obtained through different optimization algorithms and their corresponding E_{pass} and E_{stop} values

Design	Algo.	Impedance values	E_{pass} (dB)	E_{stop} (dB)
D1	CBO	121.46, 13.79, 150, 22.05, 56.06	−43.6232	−48.2027
D2	GSA	52.09, 23.68, 150, 12.02, 100.52	−38.4885	−39.7280
D3	RCGA	142.73, 17.33, 70.02, 102.84, 23.38	−39.1353	−33.1658
D4	PSO	115.11, 21.52, 96.04, 104.81, 18.76	−32.6255	−30.4433
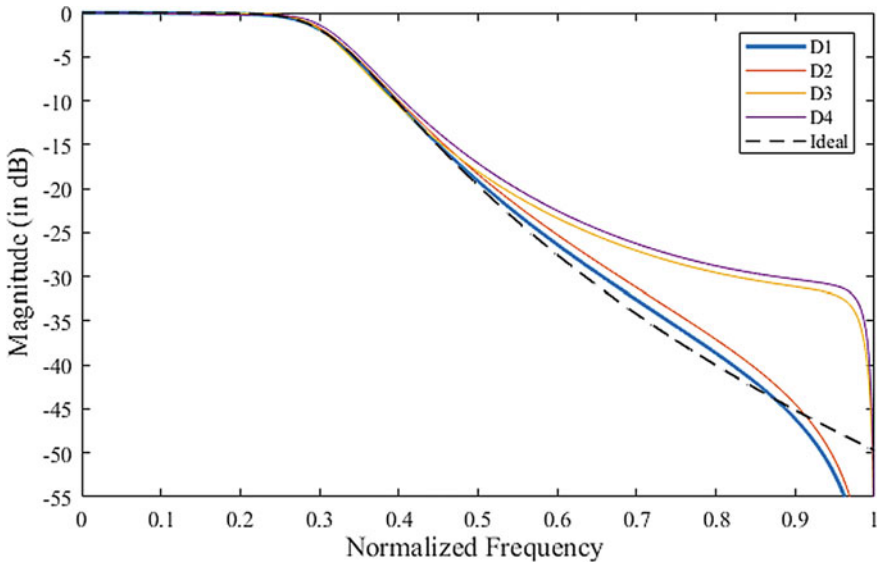
Fig. 1 Magnitude comparison of BLPF designs through different algorithms

through different algorithms shown in Fig. 1 clearly shows that results obtained from CBO give the magnitude response which is closest to the ideal one.

3 Simulation Results for the Proposed Microstrip BLPF

Configuration for the proposed design using CBO is shown in Fig. 2, which has been realized on microstrip format. The design is simulated using Agilent ADS software on a Rogers RO3003 substrate with relative dielectric constant ϵ_r: 3, substrate thickness of 0.75 mm and loss tangent value 0.001. Physical length for each line element is taken as $l = \lambda_o/4$ where λ_o is the wavelength at normalizing frequency of 10 GHz. The corresponding lengths and widths for the impedance values of design D1 are computed using a line calculator and are listed in Table 3.
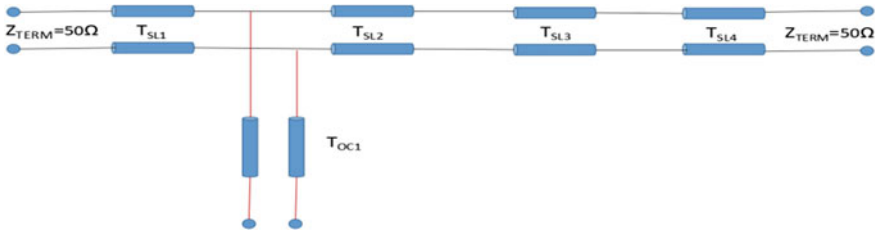
Fig. 2 Configuration of the proposed microstrip BLPF

Table 3 Width and length of the proposed microstrip BLPF

Line element	Impedance (Ω)	Width (mm)	Length (mm)
T_{SL1}	121.46	0.195	5.154
T_{OC1}	13.79	12.838	4.334
T_{SL2}	150	0.179	5.164
T_{SL3}	22.05	0.417	5.056
T_{SL4}	56.06	9.158	4.476
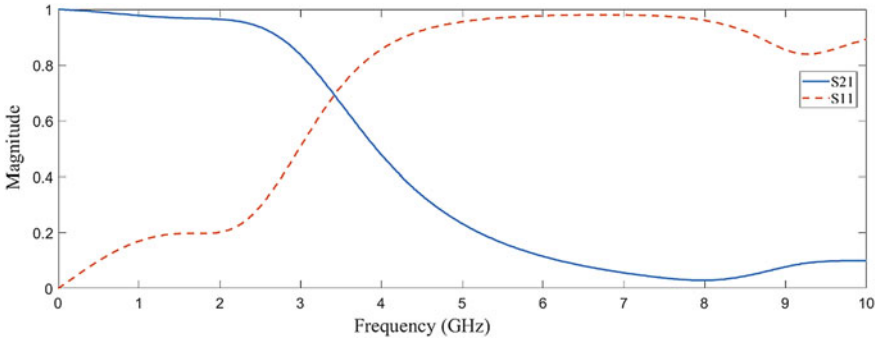
Fig. 3 Magnitude responses of scattering parameters for proposed microstrip BLPF

Figure 3 gives the magnitude responses of scattering parameters, viz., $S_{21}(f)$ and $S_{11}(f)$ for the proposed microstrip BLPF which is having 3 dB cutoff frequency of 3.4 GHz.

4   Conclusion

A Butterworth LPF of order five has been designed in the digital domain using T-matrices of equal electrical-length line elements. CBO is used for minimizing an LMS-based error function to obtain the desired impedance values of the proposed

configuration. Proposed BLPF supersedes designs obtained from other well-known metaheuristic algorithms with improved E_{pass} and E_{stop} values. Configuration for the microwave BLPF consists of a shunt-connected open stub and four serial transmission line sections. Simulations are carried out in ADS environment using Rogers RO3003 substrate and results obtained for the transmission coefficient $S_{21}(f)$ indicate the 3 dB cutoff frequency of 3.4 GHz.

References

1. Pozar, D.M.: Microwave Engineering, 3rd edn. Wiley Inc, New York (2004)
2. Chen, W.-Y., Su, Y.-H., Kuan, H., et al.: Simple method to design a tri-band bandpass filter using asymmetric SIRs for GSM, WiMAX, and WLAN application. Microw. Opt. Technol. Lett., **53**, 1573–1576 (2011)
3. Laouej, D., Daoud, H., Loulou, M.: Design of sixth order butterworth Gm-C filter using particle swarm optimization program for biomedical application. In: 29th International Conference on Microelectronics (ICM), pp. 1–5. IEEE, Beirut (2017)
4. Hong, J.-S., Lancaster, M.J.: Microstrip Filters for RF/Microwave Applications. Wiley Inc, New York (2001)
5. Hsue, C., Tsai, Y., Hsu, C., Wu, C.: Sharp rejection low-pass filter using three section stub and Z-transform technique. IET Microw. Antennas Propag. **4**(9), 1240–1246 (2010)
6. Maity, B.: Stepped impedance low pass filter using microstrip line for C-band wireless communication. In: International Conference on Computer Communication and Informatics (ICCCI), pp. 1–4. IEEE, Coimbatore (2016)
7. Jadhav, S.A., Misal, S.B., Mishra, A., et al.: Designing of stepped impedance butterworth and chebyshev filters for wireless Communication. In: Applied Electromagnetics Conference (AEMC), pp. 1–2. IEEE, Aurangabad (2017)
8. Chen, F.-C., Li, R.-S., Chu, Q.-X.: Ultra-wide stopband low-pass filter using multiple transmission zeros. IEEE Access **5**, 6437–6443 (2017)
9. Hayati, M., Abbasi, H., Shama, F.: Microstrip low pass filter with ultrawide stop band and sharp roll-off. Arabian J. Sci. Eng. **39**, 6249–6253 (2014). https://doi.org/10.1007/s13369-014-1237-x
10. Ma, K., Yeo, K.S.: New ultra-wide stopband low-pass filter using transformed radial stubs. IEEE Trans. Microw. Theory Tech. **59**(3), 604–611 (2011)
11. Packiaraj, A.D., Vinoy, K.J., Ramesh, M., Kalghatgi, A.T.: Design of compact low pass filter with wide stop band using tri-section stepped impedance resonator. AEU Int. J. Electron. Commun. **65**(12), 1012–1014 (2011)
12. Liu, S., Xu, J., Xu, Z.T.: Compact lowpass filter with wide stopband using stepped impedance hairpin units. Electron. Lett. **51**(1), 67–69 (2015)
13. Hayati, M., Shama, F.: A compact lowpass filter with ultra wide stopband using stepped impedance resonator. Radio Eng. **26**(1), 269–274 (2017)
14. Imani, M.A., Shama, F., Alirezapoori, M., et al.: Miniaturized microstrip lowpass filter using cylindrical-shaped resonators for integrated applications. Analog Integer Circ. Sig. Process **95**(2), 223–229 (2018)
15. Oraizi, H., Esfahlan, M.S., et al.: Design of stepped-impedance low pass filters with impedance matching by the particle swarm optimization and conjugate gradient method. In: European Conference on Circuit Theory and Design, pp. 639–642. IEEE, Turkey (2009)
16. Kaveh, A., Mahdavi, V.R.: Colliding Bodies Optimization. Springer International Publishing, New York City (2015)
17. Rashedi, E., Saryazdi, S., et al.: GSA: a gravitational search algorithm. Elsevier Inf. Sci. **179**, 2232–2248 (2009)

18. Chuang, Y.-C., et al.: A simple and efficient real-coded genetic algorithm for constrained optimization. Appl. Soft Comput. J. (2015)
19. Marini, F., Walczak, B.: Particle swarm optimization (PSO) a tutorial. Elsevier Chemom. Intell. Lab. Syst. **149**, 153–165 (2015)
20. Chang, D.C., Hsue, C-W.: Design and implementation of filters using transfer functions in the Z domain. IEEE Trans. Microw. Theory Tech. **49**(5), (2001)
21. Hsue, C-W., Tsai, L-C., et al.: Implementation of a trapezoidal-rule microwave integrator. Microw. Opt. Technol. Lett. **48**(4), 822–825 (2005)
22. Mahata, S., Kar, R., Saha, S.K., et al.: Optimal design of wideband infinite impulse response fractional order digital integrators using colliding bodies optimisation algorithm. IET Signal Proc. **10**(9), 1135–1156 (2016)

Short-Term Load Forecasting Using Hybrid ARIMA and Artificial Neural Network Model

Check for updates

Rahul Singhal, Niraj Kumar Choudhary and Nitin Singh

Abstract Load forecasting is basic for building up a power supply strategy to enhance the reliability of the power line and gives optimal load scheduling to numerous developing nations where the demand can be expanded with high development rate. Short-Term Electric Load Forecast (STLF) is very important because it can be used to preserve optimum behaviour in daily operations of electrical system. For this purpose, Autoregressive Integrated Moving Average Model (ARIMA) is utilised which is a linear prediction procedure. Neural networks have capability to model complex and nonlinear relationship. The aim of this paper is to explain how neural network is able to change linear ARIMA model to create short-term load forecasts. The hybrid methodology, combining ARIMA and ANN model, will purposely take advantages of the unique power of ARIMA and ANN models in linear and nonlinear domains, respectively.

Keywords Load forecasting · ARIMA · Artificial neural network · Hybrid technique

1 Introduction

Load forecasting has got the main role in operation and planning of electric utilities. Accurate load forecasting is important if it is accurate, there will be a better chance to save in the decision-making and control operations, such as unit commitment, dispatch, fuel allocation, offline analysis and power system security assessment. If an error comes in forecasting the electrical load demand will rise then this will tend to increase in operating cost. It is discovered [1] that 1% rise in electric load demand forecasting's error tends to 10 million pound rise in the operating values. If the forecasted electrical load is more than real load demand then operating cost can also increase and it decays scarce facilities and if the forecasted electric load is below than the real electrical load demand, it raises the condition of black-outs, which could be

R. Singhal · N. K. Choudhary · N. Singh (✉)
Department of Electrical Engineering, MNNIT Allahabad, Prayagraj, India
e-mail: nitins@mnnit.ac.in

more expensive. In addition to this reliable load forecasting is able to reduce energy consumption and decrease in environmental pollution.

The prediction of a short-term load is complicated by the fact that at the indicated times the load depends not only on the last hour's load but also depends on the same weight of the previous days, and the same load would be loaded with the same value on the same day in the past week. In addition, prediction methods need to be modelled using load and other variables such as weather, leisure activities, etc.

Load forecasting is an essential part of energy management in power system. Accurate load forecasting encourages the electricity suppliers to make decisions in unit commitment, schedule device maintenance plan properly and reduce spinning reserve capacity [2]. This is not only to reduce the cost of production but also for the infallibility of the power networks. System managers use load forecasting results as a base for offline network analysis to govern whether system can remain stable or not. If so, corrective measures, e.g. energy procurement and load shedding, must be prepared.

2 Short-Term Load Forecasting

The Short-Term Load Forecast (STLF) depicts a significant role in the creation of reliable, economic and safe operating techniques for the electrical system. For any model, predicting demand is the main aspect of development, particularly in the improvement of the structure of today's energy system for energy planning.

In operation of the electrical power network, the ability to give load to all consumers is challenging task, because it says that it should always have to meet the requirements of load immediately and always. Additional load is always reserve from generator so it can be used whenever there is any requirement. Due to fluctuations in loads every day, operator of system needs to guess electric load demand for coming few hours or for coming years so that a plan will be ready. Take an example [3] like fossil fuel generators requires a lot of time to synchronise, to provide adequate generation of resources for power generation. Therefore, prior information of electric load requirements aware the operator to better allocates resources for the system. For prior information of the electric load requirements, load forecasting will be needed.

2.1 Factors Influencing the Load Behaviour

The load at the system is the summation of all loads of the consumers at the same time. Aim of 'STLF' is to predict future demand for the load. A better knowledge of the attributes of the system leads to make rational forecast models and choose suitable models in contrasting situations. The load behaviour of the system is affected by several factors [4] such as

- Weather
- Time
- Economy
- Random Disturbance

2.2 Need for Short-Term Load Forecasting [5]

1. For unit commitment, scheduling of energy translations, hydro thermal coordination, load flow studies, power system security analysis.
2. To evaluate the safety of the power system network at any time, particularly knowing that in which case the power system can be stable, therefore the corrective actions can be taken to safely operate the system such as sender switching operations and power purchasing.
3. In order to achieve the objectives of the load dispatch centre in a financially and reliable way, some key elements are required. The main elements are STLF models, data origins and man-machine interfaces.

3 Methodology Used

3.1 ARIMA Model

Before prediction, there is a need to create a model ARIMA on the basis of measured data. This is a three-stage process of repetition. These stages are shown in Fig. 1.

3.1.1 Model Identification

Modelling starts with initial recognition. The aim of the initial phase is to set correct order by Auto Correlation Function (ACF) and Partial Autocorrelation Function (PACF). For any stationary time series, the autocorrelation function usually decays rapidly up to 0, but for the non-stationary time series the autocorrelation function decays slowly because time series which is observed has the trend [6]. If a process is unstable, then data transmission (i.e. power transformations and differencing) is required to stationaries the time series.

$$\begin{aligned} y_t &= \theta_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} \dots \dots + \emptyset_p y_{t-p} \\ &= \varepsilon_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \theta_q \varepsilon_{t-q} \end{aligned} \tag{1}$$

where θ and $\emptyset$ are the model parameters and y_t are the actual values.

Model Identification

Data Preparation

- Transform time series for stationary

- Difference time series for stationary

Model Selection

- Examine ACF & PACF.

- Identify potential Models(p,q)

Model Estimation & Testing

Model Estimation

- Estimate parameters in potential models .

- Select best model using suitable criterion .

Model Diagnostics / Testing

- Check ACF / PACF of residuals white noise.

Re-identify

Model Application

Model Application

- Use selected model to forecast.

Fig. 1 Box–Jenkins methodology for ARIMA modelling [16]

Before fitting the ARIMA model, there is a need to eliminate the trend and make stable time series. Once stability can be considered, the autocorrelation function and partial autocorrelation function of the stable time series is analysed to find out the sequence of the time series. Clearly, the AR and MA procedures follow some different patterns of ACF and PACF. Table 1 is used to distinguish whether it is AR or MA model, depending on the graph of ACF and PACF [7].
The utmost importance of studying the ACF and PACF method in the study of an ARMA (p,q) with both p > 0 and q > 0 is that there will be unreliability in the choice of concerning model when checking only for the ACF and PACF.

Table 1 Procedure to identify p and q from ACF and PACF graph

Model	ACF	PACF
MA (q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR (p)	Exponential decay and/or damped	Cuts off after lag p
ARMA (p, q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

$$\nabla^d x_t = \underbrace{\sum_{i=0}^{p} \emptyset_i \nabla^d x_{t-i}}_{AR\ Term} + w_t + \underbrace{\sum_{j=1}^{q} \theta_j w_{t-j}}_{MA\ Term} \tag{2}$$

where $\emptyset_i$ are the AR coefficients, θ_j are the MA coefficients and w is a Gaussian white noise having zero variance and mean. The operator ∇^d is used in differencing the time series data.

3.1.2 Parameter Estimation

After 'order' of p and q is determined, the next stage is to find out the parameters. The main fundamental method to estimate the parameters is maximum likelihood.

3.1.3 Diagnostic Testing

Diagnostic testing is the final stage in building the ARIMA model. This stage uses to check the reasonableness of the model and to examine whether the potential developments needed or not [8, 9].

3.2 Artificial Neural Network Model

The problem occurring from the time series model is that assumption of a linear relation between the present and future observations of the electric load is taken and shows a linear relation in the weather variables and the load. To eliminate this demerit, neural networks provide some possibility of predicting nonlinear time series. As has been said, a good nonlinear model must be sufficient to catch all nonlinear events in model data [10]. The electric load predictions by neural network are mapping for future loads from a group of past load, present values of load and future climate science variables.
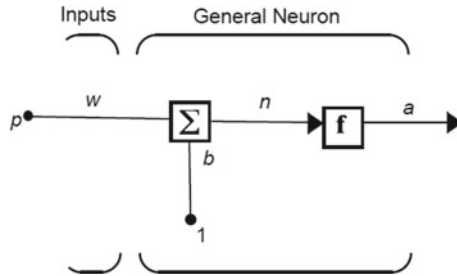
Fig. 2 General neuron
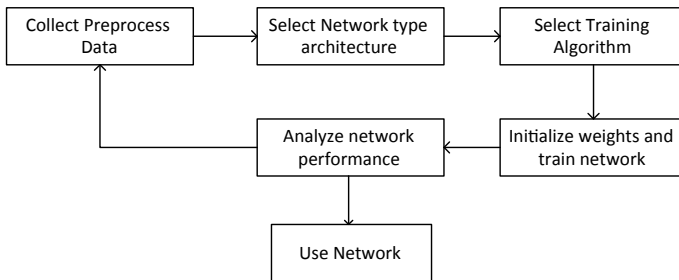
Fig. 3 Block diagram for ANN

One layer of neurons can not able to estimate arbitrary functions. There is a need to expand the network by simulating several layers simultaneously by cascading. In Fig. 2, a structure of general neuron [11] is shown how the weight (w) and biasing (b) will be given. For cascading outcome from the first layer is input to the second layer and the output from the second layer is input to the third layer.

In Fig. 3, some basic steps for modelling of artificial neural network are shown through the block diagram [12, 13].

3.3   Hybrid Model of ARIMA and Artificial Neural Network

Nonlinear autoregressive prototype with external input is a recurring dynamic system, which has a feedback link surrounding many layers of network. It is a useful model and essential for different nonlinear networks. The equation defined for the NARX model is

$$\hat{z}_t = f\left(u_{t-1},\ u_{t-2},\ \ldots,\ u_{t-p},\ z_{t-1},\ z_{t-2},\ \ldots,\ z_{t-r}\right) \tag{3}$$

where p and r are input and output order, network u_t and z_t are input and output at time t, and function f is called a nonlinear function. The succeeding value of z_t regression
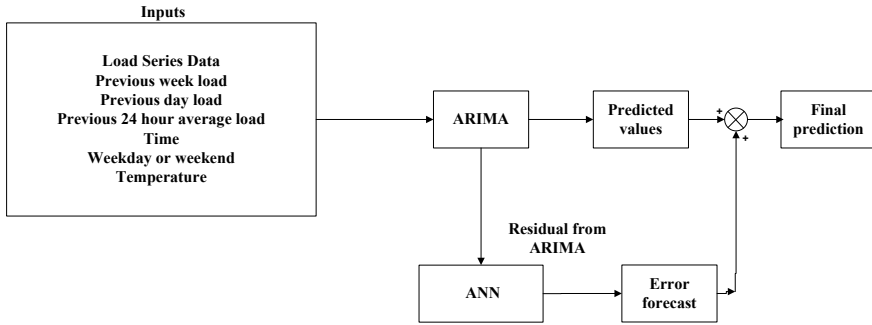
Fig. 4 Block diagram for hybrid model

depends on the output signal using past output observations and past observations of (exogenous) independent input signal. If the function f is obtained by a Multilayer Perceptron (MLP), the emerging system is known as a NARX neural network [14]. Hybrid model is applied for the better result for electric load forecasting whose block diagram is shown in Fig. 4 [15].

The inputs taken for load demand forecasting are previous week load at the same time, previous day load at same time, previous week average load and forecasted temperature at that time, what are the time and lastly the day, whether it is weekday or weekend.

4 Results and Discussion

The historical data of Denmark electrical load demand from 1 January 2012 to 31 January 2013 is taken in the intervals of 1 h.

Figure 5 describes the variation of load demand throughout the day how the load demand changes like in night hours demand is somehow less because all the commercial load are off at night hours but in day hours all the commercial and domestic loads are switched on and the demand increases but in evening time all the domestic, commercial loads and lights are switched on so it shows peak here.

Whereas, Fig. 6 shows the variation in load demand throughout the month how the load demand changes in weekdays and weekends. In weekends demand goes low due to holiday as all the offices remain closed and main commercial load is switched off.

If y_t are the actual observation for a time period t and F_t is the forecast for the same period, then the error is defined as

$$E_t = y_t - F_t \qquad (4)$$

Then the mean error can be calculated as
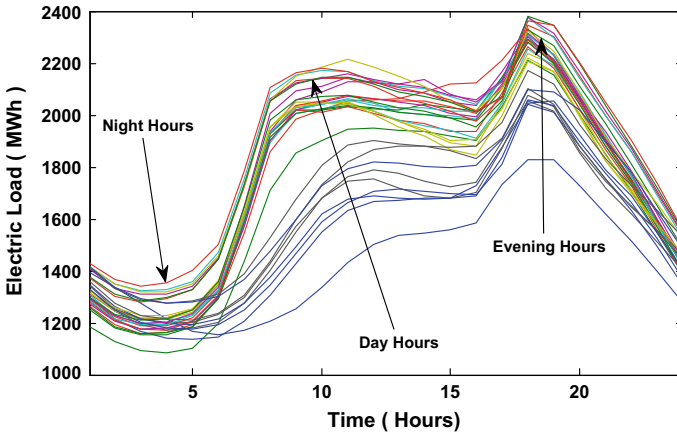
Fig. 5 Variation of electric load demand with respect to time of day
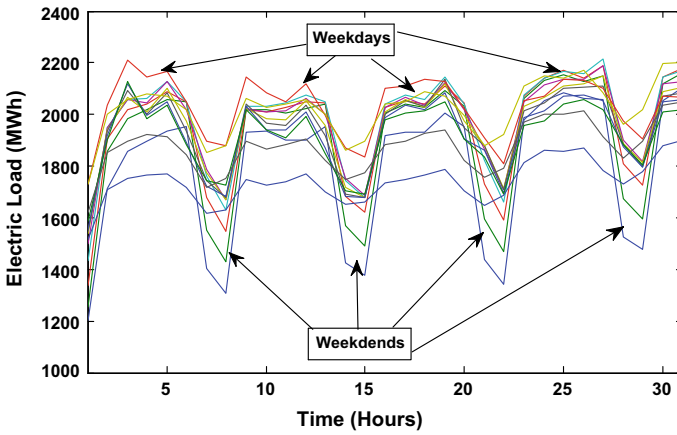
Fig. 6 Variation of electric load demand with respect to day of a month

$$ME = \frac{1}{n} \sum_{t=1}^{n} e_t \tag{5}$$

The mean square error is

$$MSE = \frac{1}{n} \sum_{t=1}^{n} e_t^2 \tag{6}$$

The mean absolute error can be estimated as

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t| \tag{7}$$

And the mean absolute percentage error is

$$MAPE = \frac{1}{n * y_t} \sum_{t=1}^{n} |e_t| * 100 \tag{8}$$

Figures 7 and 8 show the output of electric load demand for ARIMA model and ANN model individually, respectively.
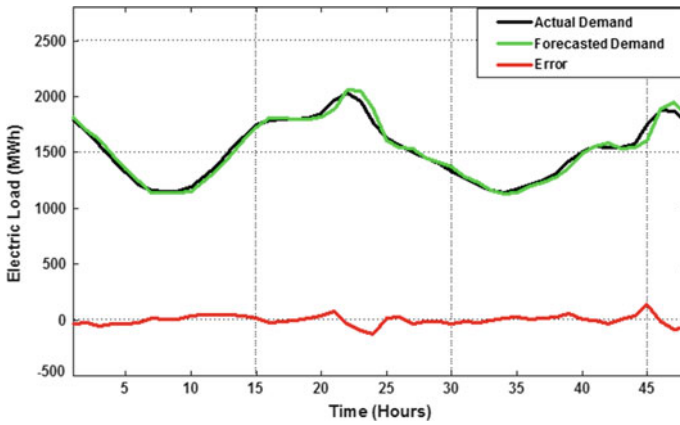
Fig. 7 ARIMA output for electric demand forecasting
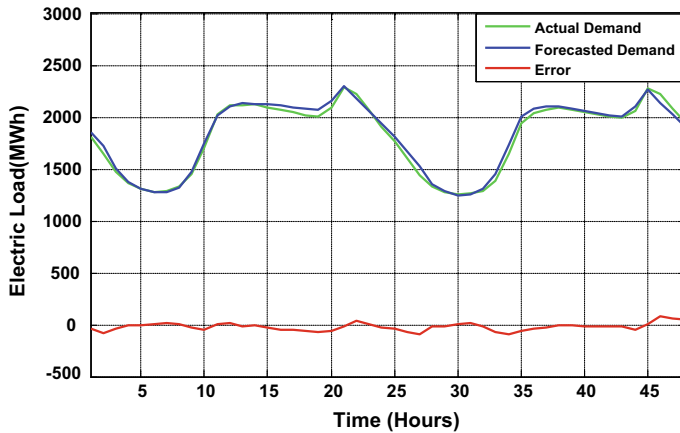
Fig. 8 Artificial neural network output for electric demand forecasting

Figures 9 and 10 show the output of electric load demand for hybrid ARIMA and ANN model for 48 h and a week, respectively.

In Table 2, the comparison between actual load demand and forecasted load demand output from different models (ARIMA, ANN and Hybrid ANN-ARIMA) is shown for 24 h and in Table 3, it compares the errors from these models which show the hybrid model of ARIMA-ANN has the least error among ARIMA, ANN and Hybrid of ARIMA-ANN model.
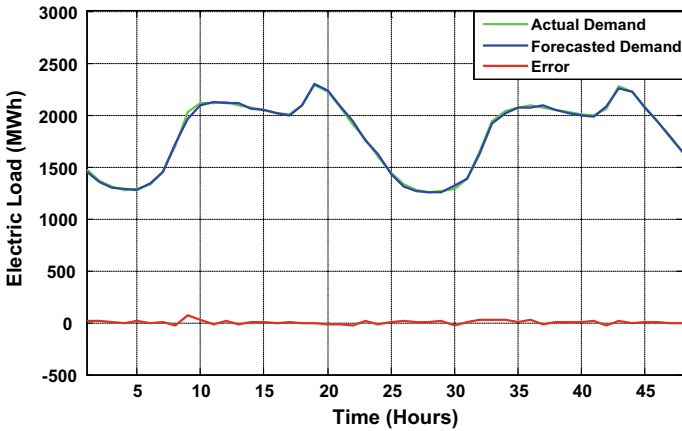
Fig. 9 Hybrid ARIMA and artificial neural network model output for electric demand forecasting
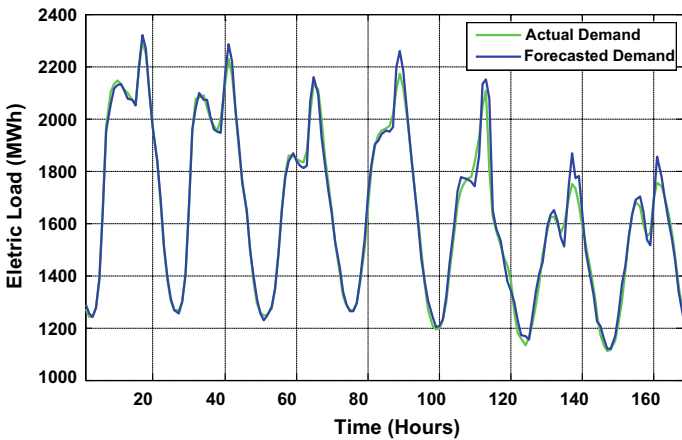
Fig. 10 Comparison for 7 days of actual demand and forecasted demand using hybrid ARIMA and ANN model

Table 2 Comparison of actual and forecasted values

Actual values (MWh)	Forecasted values (MWh)		
	ARIMA	ANN	HYBRID of ARIMA and ANN
1438	1474.789	1444.39	1400.742
1318.9	1354.505	1280.556	1303.575
1214.4	1239.413	1217.011	1237.587
1163.4	1145.014	1177.375	1167.398
1147.8	1141.671	1163.495	1143.591
1149.4	1142.191	1162.777	1150.973
1194.4	1153.666	1184.618	1180.174
1279.7	1237.313	1299.266	1332.047
1387.5	1344.775	1509.423	1488.781
1514.8	1464.947	1748.209	1578.636
1641.1	1607.212	1866.212	1665.226
1744	1726.254	1876.118	1745.806
1788.3	1809.118	1864.573	1802.657
1796.2	1805.92	1845.41	1810.796
1795.5	1795.811	1836.692	1793.394
1807.1	1793.912	1830.994	1791.332
1853.7	1816.463	1836.981	1826.915
1967.2	1893.715	1921.958	2031.691
2026.8	2059.48	2007.113	2100.521
1959.6	2050.125	1949.531	1946.036
1764.3	1888.078	1854.457	1823.099
1622.8	1608.629	1752.008	1599.881
1565.8	1542.7	1651.321	1532.154
1507	1539.469	1536.812	1489.298

Table 3 Statistical error measures

MODEL	MAE	MSE	RMSE	MAPE
ARIMA	36.6335	2.6374e + 03	51.3561	2.2698
ANN	34.0392	1.9402e + 03	44.0478	2.0312
HYBRID (ARIMA + ANN)	24.1341	1.0367e + 03	32.1978	1.6014

5 Conclusion

Any particular model cannot capture the different patterns but the association of various models can increase the ability to capture a variety of patterns in the data and improve the forecasting performance. From the error measure, it is observed that ARIMA and ANN individually are not capable of giving better result as compared to hybridization of these models. The data used in the manuscript includes linear as well as nonlinear parts of the historical load data and randomness of many uncertain factors. ARIMA and ANN have different roles to play in the hybrid model based on the inherent characteristics of the load data. ARIMA model has a very good performance when it comes to forecasting the linear part, and ANN model is good for working with nonlinear part. Their combination has resulted in improving the overall accuracy of the forecast. Furthermore, the hybrid model is highly reliable in regard to capturing the seasonal change of the data structure. The MAPE obtained in ARIMA model is 2.26 and in ANN it decreases to 2.03 while in Hybrid it is 1.60.

References

1. Alkhathami, M.: Introduction to Electric Load Forecasting Methods Received 1 April 2015; Accepted 9 May 2015; Published online 26 May 2015
2. Singh, A.K., Ibraheem, Khatoon, S., Muazzam, M., Chaturvedi, D.K.: Load Forecasting Techniques and Methodologies: A Review. In: 2012 2nd International Conference on Power, Control and Embedded Systems
3. Gross, G., Galiana, F.D.: Short-term load forecasting. Proc. IEEE **75**, 1558–1573 (1987)
4. Kyriakides, E., Polycarpou, M.: Short Term Electric Load Forecasting: A Tutorial, vol. 35 of Studies in Computational Intelligence. Springer, Berlin, Heidelberg (2007)
5. Srivastava, A.K., Pandey, A.S., Singh, D.: Short-Term Load Forecasting Methods: A Review. IEEE. 978-1-5090-2118-5/16 ©2016
6. Montgomery, D.C., Jennings, C.L., Kulahci, Murat: Introduction to Time Series Analysis and Forecasting. Wiley-Interscience, Hoboken, NJ (2008)
7. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. Wiley, Hoboken, NJ (2016)
8. de Andrade, L.C.M., da Silva, I.N.: Very short-term load forecasting based on ARIMA model and intelligent systems. IEEE. 978-1-4244-5098-5/09 ©2009
9. Singh, S., Singh, R.: ARIMA based short term load forecasting for Punjab region. Int. J. Sci. Res. (IJSR) **4**(6) (2015). ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14|Impact Factor (2013): 4.438
10. Peng, T.M., Hubele, N.F., Karady, G.G: Advancement in the application of neural networks for short term load forecasting. IEEE Trans. Power Syst. **7**, 250–257 (1992)
11. Senjyu, T., Takara, H., Uezato, K., Funabashi, T.: One-hour-ahead forecasting using neural network. IEEE Trans. Power Syst. **17**, 113–118 (2002)
12. Fidalgo, J.N., Matos, M.A.: Forecasting portugal global load with artificial neural networks. In: Proceedings of the 17th International Conference on Artificial Neural Networks. Proceeding ICANN 2007, pp. 728–737 (2007)

13. Tsekouras, G.J., Hatziargyriou, N.D., Dialynas, E.N.: An optimized adaptive neural network for annual midterm energy forecasting. IEEE Trans. Power Syst. **21**, 385–391 (2006)
14. Peter Zhang, G.: Time series forecasting using a hybrid ARIMA and neural network model Received 16 July 1999; accepted 23 November 2001
15. EI Desouky, A.A., EI Kateb, M.M.: Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA. IET Proc.—Gener. Transm. Distrib. (2000)
16. Nguyen, H., Hansen, C.K.: Short-Term Electricity Load Forecasting with Time Series Analysis. IEEE. 978-1-5090-0382-2/16 ©2017

Enhanced XOR-Based Progressive Visual Secret Sharing Using Multiple Decryptions

Vishal Singh Sachan, Mainejar Yadav and Ranvijay

Abstract This paper proposes XOR-based visual secret sharing rule using random grids with the abilities of OR as well as XOR-based visual cryptographic schemes. This scheme is different from previously proposed schemes as this paper proposes a lossless (k, n) threshold-based vcs with progressive recovery having advantages of multiple decryptions. Although previously proposed schemes were progressive with OR operator but in case of XOR their progressive nature was lost when the number of shares "t" is between "k" and "n", i.e., k < t < n. Our scheme overcomes this problem by proposing a truly progressive scheme, completely independent of the values of "k" and "n." In absence of a processing device the secret information can be simply reconstructed by OR-ing a threshold number of shares and in case, if a device having computational ability of XOR is available, then the secret information encoded is reconstructed lossless for (k, n) threshold, when all the "n" shares are collected. There is no pixel expansion, no requirement of codebook design and hence all the shares were generated of the same size of the original secret image.

Keywords Visual cryptography · Visual secret sharing · Random grids · Multiple decryptions

1 Introduction

Secret sharing is one of the most important cryptanalytic application. It is mainly employed in cases wherever the collaborating parties do not trust each other or they are unreliable whereas there is equal importance of the participants and thus the secret information is divided into parts and distributed among them in such a way that

V. S. Sachan (✉) · M. Yadav · Ranvijay
Computer Science and Engineering Department, MNNIT Allahabad, Allahabad, India
e-mail: er.singhvishal77@gmail.com

M. Yadav
e-mail: ruhulit1210@gmail.com

Ranvijay
e-mail: ranvijay@mnnit.ac.in

individual shares of the participants reveal no original secret information, however, once a predefined number of participants group together their secret shares reveal the originally intended information. Shamir et al. [1] first introduced the idea of secret sharing, this scheme was mainly based on Lagrange Interpolation polynomial. In the present world, image security and secret sharing are one of the most important fields of work because the modification of digital image and using it for unlawful purposes has become very simple. Visual secret sharing scheme is a variant of cryptography that is based on the human visual system which performs the decoding phase and no further computation is needed. Hence it is simple and can be employed by everyone. Progressive scheme has a property that with the increase in number of shares the visual quality of the recovered image increases gradually, this is helpful when you have to assign some privilege based on the level of secret information they have.

Let us consider an example where "n" thieves are present, no one trusts each other therefore in order to gain low authorization access to the account a threshold number of shares are required whereas to perform high privileged operations an additional number of shares are needed. As additional shares are collected more authorization is granted.

This paper proposes an enhanced random grid-based progressive secret sharing algorithms with the abilities of both XOR and OR operations. Our scheme is truly progressive with both XOR and OR operations which were not there in previous schemes. Moreover, in our scheme when all the "n" shares are present we can reconstruct the secret information in a lossless manner provided that we are having a device with XOR operation and in absence of such device we can simply reconstruct by stacking the shares.

2 Related Work

There are several approaches that are proposed by different researchers. Naor and Shamir [2] first proposed the threshold-based VCS scheme. In their scheme "n" shares are generated corresponding to a secret image and each share does not reveal any information regarding the original secret key. These shares were then printed on "n" number of transparencies and given to the designated participants. This method was secure as each share do not reveal any information unless a "threshold number" of shares that is defined at the time of decoding are present. Kafri and Keren [3] in 1987 proposed the random grid model in which each of the shares that were generated is meaningless random grids and the size of the shares was identical to the original secret image, thus there was no pixel expansion. The decoding operation was same as that of traditional visual cryptography, by stacking both the encoded and the master grid original information can be reconstructed. The application of random grid in image encryption removes disadvantages such as pixel expansion, shape distortion, and codebook design. But due to the lack of encryption techniques to implement the (2, 2) technique to a more generalized (k, n) or (n, n) scheme, its application was mainly limited. Later in 2008 Chen and Tsao proposed (n, n) and

(2, n) [4] random grid-based visual secret sharing scheme, they extended the (2, 2) scheme so that more number of participants can participate and the algorithm can be applied to much wider areas. This scheme is then again extended by Chen and Tsao [5] to develop a more generalized and applicable (k, n) threshold-based scheme. In their scheme, a threshold number of shares was required at the decoding phase to reconstruct the original secret image and if that threshold "k" is not met then the secret image is not revealed.

The limitations of the above work done were that the recovered image was lossy, i.e., the key information was not fully recovered because of the monotonous property of OR operator, which degrades the visual appearance of the recovered secret information. To overcome this problem XOR-based VCS was introduced owing to its better quality image reconstruction ability, however, the XOR-based scheme would be useless if no suitable computational device having XOR operation is present. Thus in order to incorporate the advantages of both OR-based VCS scheme as well as XOR-based VCS theme, schemes having the property of both XOR as well as OR-based scheme was proposed. The multiple decryption ability is helpful in cases whenever a computational device with XOR functionality is not present then we can easily reconstruct a low visual quality secret image by performing OR operation(stacking) on the shares, thus revealing the secret information. Wu and Sun [6] proposed a (k, n) threshold-based scheme with abilities of multiple decryptions but the recovered image was lossy even when all the "n" shares were stacked. Yan et al. [7] later proposed a random grid-based vcs scheme having multiple decryptions with progressive recovery of secret image. By progressive recovery, we mean that as the number of shares increases the visual quality of reconstructed image also increases. When a device with XOR operation in not available, the secret information can be revealed by stacking "k" or more shares to generate the lossy secret image whereas when computational device having XOR operation is available then the secret image is lossily reconstructed if "k" shares are collected and lossless reconstructed when all "n" shares are collected.

3 Proposed Work

In this section, we will propose an enhanced XOR-based progressive visual secret sharing algorithm with the abilities of multiple decryptions. It is different from the previous schemes [7] as the previously proposed multiple decryption schemes were not truly progressive. The generation of shares is described in algorithm 1 whereas algorithm 2 describes the secret image recovery steps. Definitions required for proper understanding and implementation of the proposed algorithms is also given. This paper is organized into three sections, first the proposed algorithms then its performance analysis and finally experimental results and its comparison with previously proposed algorithms.

Algorithm 1. **The proposed scheme**

Input: A binary secret image SB and a colored cover image C both of same size MxN and the threshold parameters (k, n).

Output: n meaningful shares MS_1, MS_2, MS_3, MS_4, MS_n of size MxN

Step 1: for each position (x, y) of the secret image such that ($1 \leq x \leq M$, $1 \leq y \leq N$)

        repeat steps 2-7

Step 2: create b_1, $\overline{b_2}$ as follows

        $b_1 \| \overline{b_2}$ =Random_pixel (SB (i, j))

Step 3: create $b_2 \sim b_k$ as cipher pixels recursively using the following algorithm

      if (n >2)

        {

              For p=2 to k-1

              $b_p \| \overline{b_{p+1}}$ =Random_pixel_generator (SB (i, j))

        }

Step 4: create b_k as the last cipher grid pixel

        $b_k = \overline{b_k}$

Step 5: if n>k , set b_{k+1} , b_{k+2}, b_{k+3} , b_{k+4} ..... $b_n = 0$

Step 6 : randomly rearrange b_1, b_2,…, b_n to c_1, c_2, c_3…, c_n

Step 7: for each share MS_i , $1 \leq i \leq n$

        do{

          $MS_i = C$;

      Randomly choose d ,d $\in$ {2,3}

      Then , $MS_i(i,j,d) = c_i$

          }

Step 8: Output n meaningful shares MS_1, MS_2, MS_3, MS_4 , … MS_n

Definition 1. **frandom ( )**

$$\text{frandom()} = \begin{cases} 0 \text{ with probability } 0.5 \\ 1 \text{ with probability } 0.5 \end{cases}$$

Definition 2. Random_pixel_generator (b_s)

    Step 1.1 Create b_k as a cipher-pixel

        b_k = frandom( ),

    Step 1.2 Create $\overline{b_{k+1}}$ as follows

$$\overline{b_{k+1}} = \begin{cases} b_k & \text{if}(b_s=0) \\ \overline{b_k} & \text{if}(b_s=1) \end{cases}$$

Algorithm 2. **The secret image recovery**

Input: t meaningful shares MS_1, MS_2, MS_3, MS_4, … MS_t of size M X N, and the

threshold parameter (k,n)

Output: A M x N binary secret image S'

Step 1: repeat steps 2 for each meaningful share MS_t

Step 2: for each position (x,y) of the secret image such that $(1 \le x \le M, 1 \le y \le N)$

if($MSi(x,y,3)$<2)

$Ci(x,y)=MS_i(x,y,3)$

else

$Ci(x,y)=MS_i(x,y,2)$

Step 3: if a device having XOR operation is not available then stack all the shares

C_1, C_2, C_3 ,…C_n to recover the secret image S', go to step 5 else goto step 4

Step 4: if computational device is present then simply XOR all the t shares, i.e.

$S'=C_1 \otimes C_2 \otimes C_3 \otimes C_4 …. C_t$

Step 5: Output the binary secret image S'.

4   Performance Analysis of Proposed Algorithm

This section describes the performance analysis of the proposed algorithm by analyzing the security and performance metrics. We will analyze our algorithm based on various parameters and compare them with the previously proposed algorithms.

4.1   Definition (Average Light Transmission)

For a certain pixel "r" in a binary image SB whose size is M × N, the light transmission of a white pixel is defined as T(r) = 1; whereas, T(r) = 0 for "r" is a black pixel. Totally, the average light transmission of R is defined as follows:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} T(SB(i, j)) \tag{1}$$

4.2 Definition (Area Representation)

Let A (0) (resp. A(1)) represents the area of white (resp. black) pixels in binary secret image SB where A = A (0) ∪ A(1) and A (0) ∩ A(1) = ø. Hence the total area of all the corresponding white and black pixels will be SB (A (1)) and SB (A (0)), respectively, in image SB.

4.3 Definition (Contrast)

The contrast of the recovered secret image S' = $S_{S1} \otimes_{S2} \otimes Ss_3 \ldots \otimes S_{Sn}$ [SB(0)] with respect to the original binary secret image SB is defined as follows:

Let

$$t_1 = T(S_{S1} \otimes_{S2} \otimes S_{s3} \ldots \ldots \otimes S_{Sn}[SB(0)]$$
$$t_2 = T(S_{S1} \otimes_{S2} \otimes S_{s3} \ldots \ldots \otimes S_{Sn}[SB(1)]$$

Then,

$$\alpha = (t_1 - t_2)/(1 + t_2) \tag{2}$$

The contrast of the reconstructed secret image after OR or XOR operation must be greater than 0 so that the reconstructed secret information or image can easily be identified by the human visual system. Average light transmission is directly proportional to the contrast (α), so the value of contrast must be as high as possible, higher value means higher average light transmission which results in good quality reconstructed secret image.

4.4 Definition (Mean Squared Error (MSE))

Let us consider a binary secret image SB and S' be the recovered image both of dimensions MxN, then the mean squared error is the cumulative squared error between the original and the reconstructed image, lower value of MSE means lesser error. The mathematical formulae for calculating is as follows:

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[SB(i, j) - S'(i, j) \right]^2 \tag{3}$$

4.5 Theorem 1. (Security)

There is no dependency among the bits $b_1, b_2, b_3 \ldots b_{k-1}$ generated using the random pixel generator function and are also completely independent of the original secret binary image "SB." The bit b_k is dependent on the previous $(k-1)$ bits generated and the original bit in the secret binary image due to the XOR operation being used in Step 4 and the remaining bits set $b_{k+1}, b_{k+2}, b_{k+3} \ldots b_n$ are generated using Step 5 of the algorithm 1. To make the scheme more secure cover image is used to hide the share bits that were generated $b_1, b_2 \ldots b_k, b_{k+1} \ldots b_{n2}, b_{n-1}, b_n$. Hence for $t < k$, the probability of secret image getting revealed is negligible, and XOR of shares less than "k" the probability of appearance of white as well as the black pixel is ½, i.e.,

$$T(S_{S1} \otimes_{S2} \otimes S3 \ldots \ldots \otimes S_{k-1}[SB(0)] = 0)$$
$$= T(S_{S1} \otimes_{S2} \otimes S3 \ldots \ldots \otimes S_{k-1}[SB(0)] = 1/2$$

Thus each pixel which appears is not fixed as pixel placement is completely independent of each other due to the usage of flipping coin procedure and hence the proposed scheme is secure.

5 Experimental Results and Analysis

In this section we will perform experiments on different test secret images for different values of "k" and "n", analyze the performance metrics of the proposed algorithm and compare them with the previously proposed scheme [7] as well. In these experiments, all the images that were used are of size 512×512 pixels and all the experiments were carried out on Intel(R) Core(TM) i7-4790 CPU @3.60 GHz octa-core processor with 4 GB of RAM and Windows 10 Pro 64-bit operating system.

Figure 1(a) represents the secret image 1 used for simulation of the proposed scheme for (2, 4), Fig. 1(b–e) represents the shares generated MS_1, MS_2, MS_3, MS_4, and Fig. 1(f–h) represents the results of XOR operation. The average light transmission and MSE values for t = 2, t = 3, and t = 4 are 0.00033, 0.40125, 1 and 0.48274, 0.24909, 0, respectively. Figure 1(i–k) represents the recovered images of OR operation. Figure 2(a) represents the secret image 2, Fig. 2(b–f) represents the meaningful shares that were generated, MS_1, MS_2, MS_3, MS_4, MS_5, and Fig. 2(g–j) represents the result of XOR operation when k < = t < = n. The average light transmission and MSE values for t = 2, t = 3, t = 4, and t = 5 are 0.06344, 0.07018, 0.30920, 1 and 0.48513, 0.44905, 0.29954, 0, respectively. In a (3, 5) scheme when less than "k" shares are stacked no information is revealed about the original secret image and hence a threshold number of shares are always mandatory to reveal the original secret image. Figure 3(a) represents the secret image 3, Fig. 3(b–e) represents
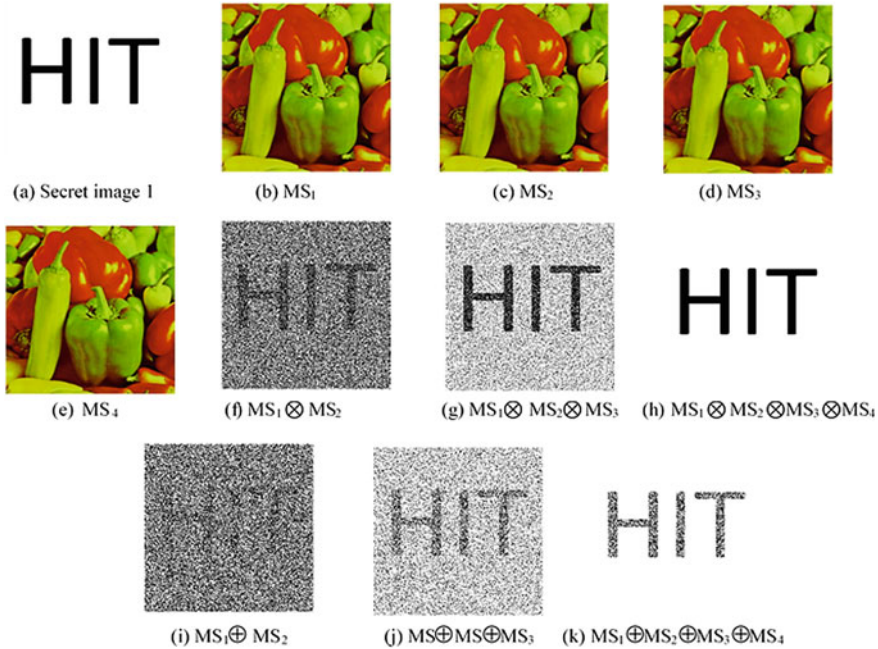
Fig. 1 Experimental results of proposed scheme for (2, 4) on secret image 1

the meaningful shares MS_1, MS_2, MS_3, MS_4, and Fig. 3(f–h) represents the results of XOR operation when a different number of shares are present.

With more number of shares better reconstructed image is obtained and the image reconstructed for values of "t", where "t" is k < t < n, the progressive nature is not lost and the proposed scheme behaves as true progressive in nature. The average light transmission and MSE values for t = 2, t = 3, and t = 4 are 0.00077, 0.39862, 1 and 0.432125, 0.25085, 0, respectively. As in a (2, 4) scheme, the average light transmission increases for t = 3 and does not decrease with increase in number of shares, i.e., for t = 4 & t = 5. Figure 3(i–k) represents the results of OR operation, the results obtained with XOR operation are truly progressive when the value of "t" lies between "k" and "n", moreover, the visual quality obtained with XOR is more than the OR operator.

Figure 4(a) represents the secret image 4 and Fig. 4(b–f) represents the five meaningful shares MS_1, MS_2, MS_3, MS_4, MS_5, and Fig. 4(g–j) represents the results of XOR operation when different number of shares are stacked. Figure 4(g) represents recovered image when t = 2, i.e., when two shares are present, Fig. 4(h–j) represents the recovered image with t = 3, t = 4 and t = 5, respectively. The average light transmission and MSE values for t = 2, t = 3, t = 4, and t = 5 are 0.12538, 0.14357, 0.50052,1 and 0.45384, 0.35032, 0.19983, 0, respectively.

From Table 1 it can be seen that in (2,4) case of previous scheme the average light transmission is increasing gradually with the increase in the number of shares "t",

(a) Secret image 2   (b) MS_1 (c) MS_2 (d) MS_3

(e) MS_4 (f) MS_5 (g) $MS_1 \otimes MS_2$ (h) $MS_1 \otimes MS_2 \otimes MS_3$

(i) $MS_1 \otimes MS_2 \otimes MS_3 \otimes MS_4$ (j) $MS \otimes MS_2 \otimes MS_3 \otimes MS_4 \otimes MS_5$ (k) $MS_1 \otimes MS_2$ (l) $MS_1 \otimes MS_2 \otimes MS_3$

(m) $MS \oplus MS \oplus MS_3 \oplus MS_4$ (k) $MS_1 \oplus MS_2 \oplus MS_3 \oplus MS_4 \oplus MS_5$
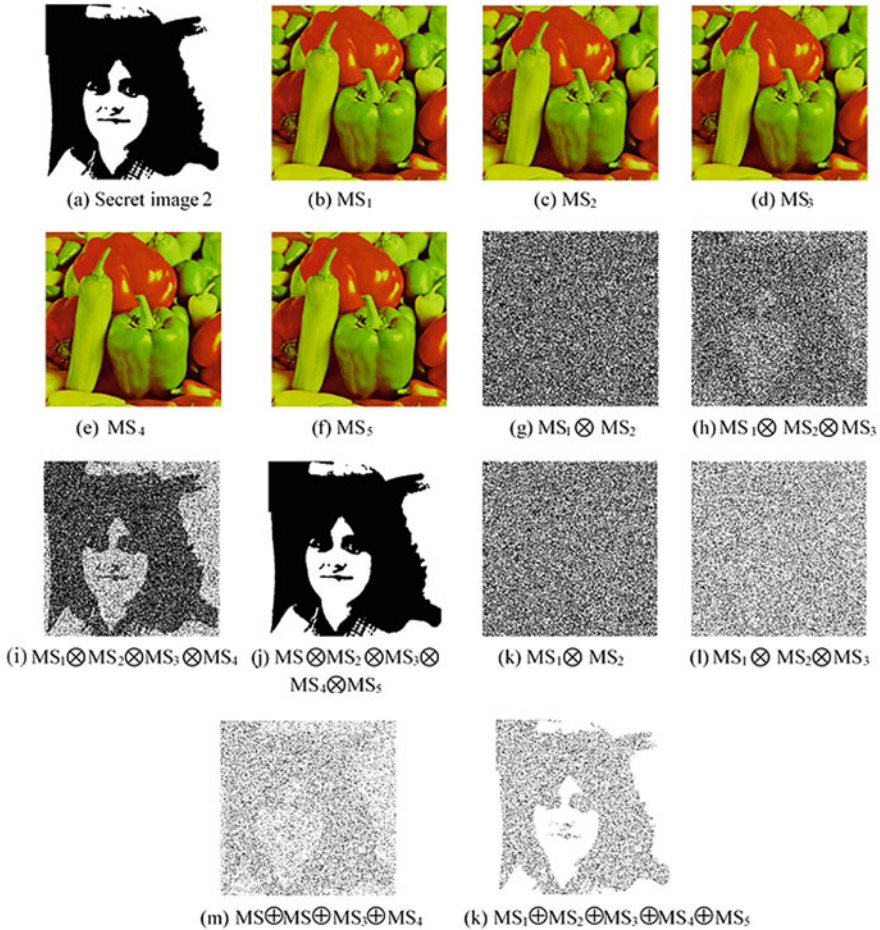
Fig. 2 Experimental results of proposed scheme for (3, 5) on secret image 2

for $t = 2$ the value is 0.000607 which gradually increases to 0.002092 and then for $t = 4$ we get complete recovery, i.e., value becomes 1, in case of proposed scheme the value of average light transmission for $t = 2, 3, 4$ are 0.001293, 0.40079, and 1, respectively whereas comparing it with (2,5) case of previous scheme for $t = 3$ and $t = 4$ there is a gradual decrease in the average light transmission, i.e., for $t = 3$ the light transmission decreases to 0.06657 as compared to 0.06825 for $t = 2$ which further decreases to 0.001943 when $t = 4$ in spite of increase in the number of shares thereby violating the progressive nature. In contrast, the secret image in the proposed scheme is reconstructed progressively and the original information is not lost with the increase in the number of shares. There is progressive recovery and the average light transmission value increases with increase in the number of shares, i.e., for $t = 2, 3, 4$, and 5 the values are 0.12538, 0.143573, 0.50052, and 1, respectively.
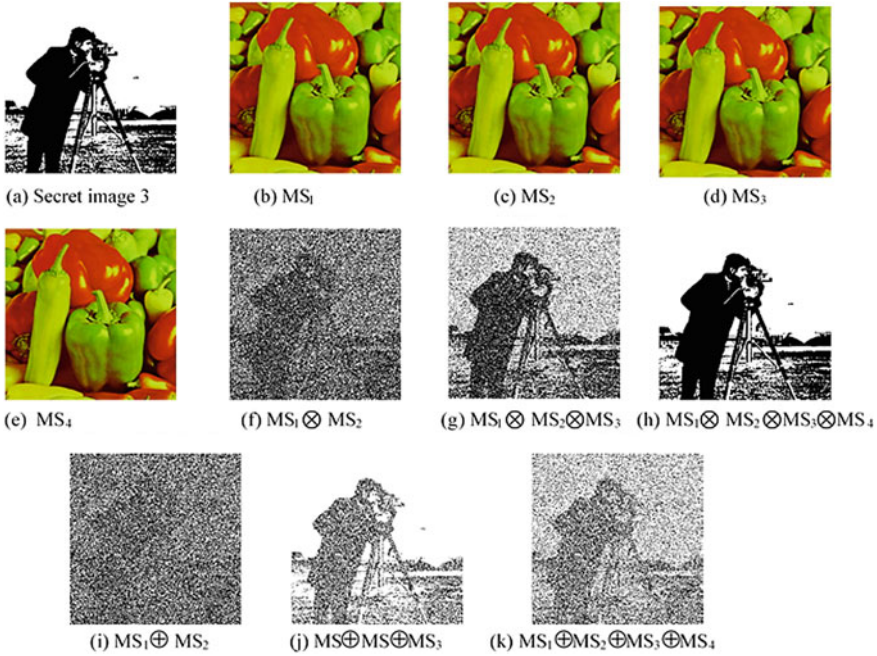
Fig. 3 Experimental results of proposed scheme for (2, 4) on secret image 3

When we compare both the cases of (2, 4) and (2, 5) we can easily infer that in case of the previous algorithm the progressive recovery is not always possible due to the self-complementing nature of the XOR operator. There might be cases where with XOR operator in spite of increasing the shares the quality of image degrades as demonstrated in (2,5) case, whereas in case of our scheme the nature of progressive recovery is always retained even for large values of n, though with the increase in the total number of participants the image recovery in more linear.

Table 2 lists the value of mean squared error of the proposed and the previous scheme for different values of k and n, lower the value of MSE higher is the quality of the image formed. Table 2 also supports the results drawn from Table 1, i.e., in some cases, previous algorithm loses the progressive nature with the increase in number of shares and the proposed algorithm overcome this problem. There might be also the case that in spite of increase in average light transmission the overall visual quality of the recovered image decreases which is also depicted by the MSE values. Table 3 lists the MSE values for recovered image with OR operation and it is evident that in the case of OR operation even when all the shares are present, secret image is not recovered completely.

In the case of OR operation even when all the shares are present, secret image is not recovered completely. The image reconstruction with OR operator is lossy and a low quality of image is formed.

(a) Secret image 4   (b) MS₁   (c) MS₂   (d) MS₃

(e) MS₄   (f) MS₅   (g) MS₁ ⊗ MS₂   (h) MS₁ ⊗ MS₂ ⊗ MS₃

(i) MS₁⊗MS₂⊗MS₃⊗MS₄   (j) MS⊗MS₂⊗MS₃⊗ MS₄⊗MS₅   (k) MS₁⊗ MS₂   (l) MS₁⊗ MS₂⊗MS₃

(m) MS⊕MS⊕MS₃   (n) MS₁⊕MS₂⊕MS₃⊕MS₄
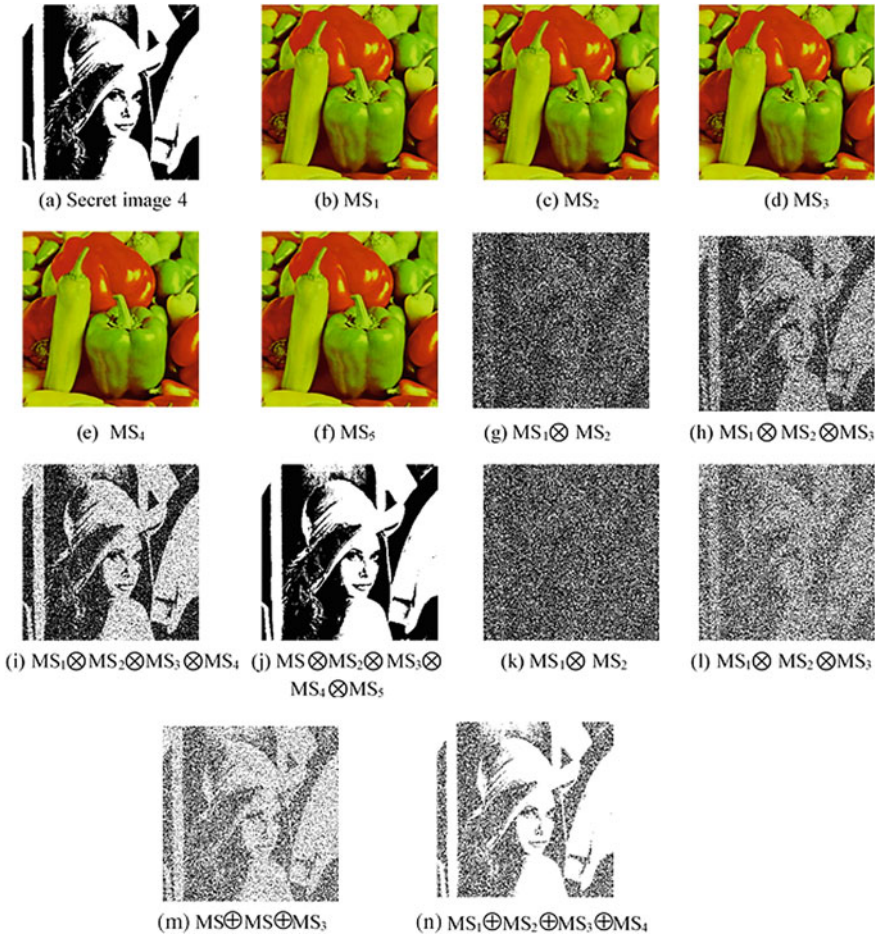
Fig. 4 Experimental results of proposed scheme for (2, 5) on secret image 4

Thus, for different values of "k" and "n" the secret image is completely recovered when t = n with XOR operation. Stacking also recovers the secret image but the recovery is lossy and complete information is not obtained. OR operator can be treated as a secondary option when a suitable device having the capability of performing XOR is not present so that a lower quality but visually recognizable version of the secret image can be easily constructed which is recognizable by the human visual system.

Figure 5(a) shows the performance of the proposed algorithm and previous algorithm [7] on secret image 1 for (2, 4). The graph is plotted between the average light transmission of each reconstructed image for different values of shares "t." It can be seen that in each case the recovery in the proposed algorithm is strictly progressive and in case of a large number of shares, i.e., Fig. 5(b–c). The Fig. 5(b) represents the

Table 1 Average light transmission, of the proposed scheme and previous scheme [7] using XOR operation

(k, n)	Proposed scheme				Previous scheme [7]			
	t = 2	t = 3	t = 4	t = 5	t = 2	t = 3	t = 4	t = 5
(2,2)	1				1			
(2,3)	0.2533	1			0.2509	1		
(3,3)		1				1		
(2,4)	0.0014	0.4001	1		0.0006	0.0021	1	
(3,4)		0.1827	1			0.1823	1	
(4,4)			1				1	
(2,5)	0.1254	0.1436	0.5005	1	0.0683	0.0666	0.0019	1
(3,5)		0.0719	0.3086	1	–	0.068	0.0023	1
(4,5)			0.1433	1			0.1431	1
(5,5)				1				1

Table 2 Mean squared error of the proposed scheme and previous scheme [7] using XOR operation

MSE	Proposed scheme				Previous scheme [7]			
(k,n)	t = 2	t = 3	t = 4	t = 5	t = 2	t = 3	t = 4	t = 5
(2,2)	0				0			
(2,3)	0.3324	0			0.3326	0		
(3,3)		0				0		
(2,4)	0.4201	0.2500	0		0.4178	0.4987	0	
(3,4)		0.3754	0			0.3746	0	
(4,4)			0				0	
(2,5)	0.4539	0.3503	0.1998	0	0.4503	0.5028	0.4998	0
(3,5)		0.4476	0.2992	0		0.4508	0.4977	0
(4,5)			0.4004	0			0.399	0
(5,5)				0				0

plot for secret image 2 for (3, 5), Fig. 5(c–d) represents the plot for secret image 3 (2, 4), and secret image 4 (2, 5), respectively. From the graph, it is also evident that the performance of the proposed scheme is better than the previous scheme [7]. In case of the previous scheme [7] when number of shares increases more than "k" then there is a sharp decrease in the overall visual quality of the recovered image with XOR operation and the scheme does not behave as truly progressive in nature.

Table 3 MSE of proposed scheme and previous scheme [7] using OR operation

MSE	Proposed scheme				Previous scheme			
(k,n)	t = 2	t = 3	t = 4	t = 5	t = 2	t = 3	t = 4	t = 5
(2,2)	0.2439				0.2462			
(2,3)	0.4158	0.2447			0.4165	0.2452		
(3,3)		0.3682				0.3660		
(2,4)	0.4599	0.3718	0.2438		0.4550	0.4304	0.3669	
(3,4)		0.4617	0.3667			0.4624	0.3664	
(4,4)			0.4280				0.4276	
(2,5)	0.4772	0.4241	0.3452	0.2439	0.4709	0.4555	0.4545	0.4286
(3,5)		0.4825	0.4430	0.3665		0.4807	0.4658	0.4278
(4,5)			0.4792	0.4282			0.4791	0.4281
(5,5)				0.4584				0.4581
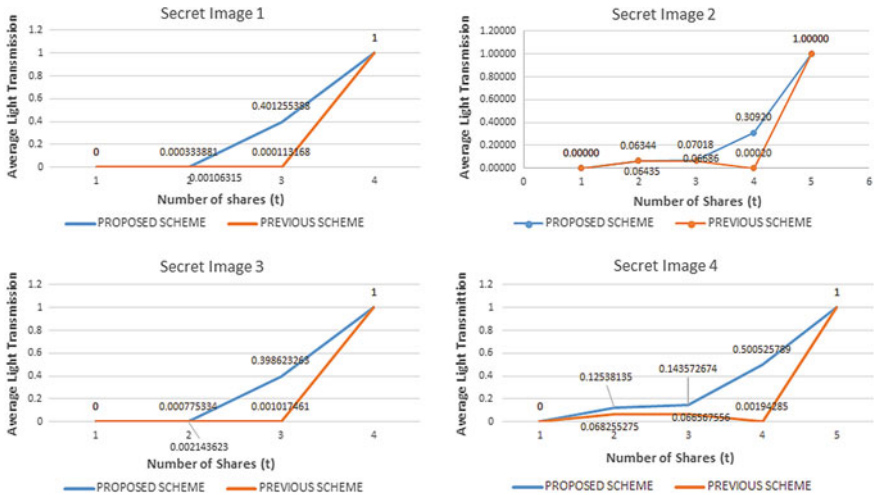
Fig. 5 Graphs showing the variation in average light transmission with different values of k and n

6 Conclusion

This paper proposes an enhanced random grid-based progressive secret sharing algorithms with the abilities of both XOR and OR operations. Our scheme is truly progressive with both XOR and OR operations which were not there in previous schemes. Moreover, in our scheme when a computational device having XOR operation is present then we can recover our lossless secret image when all the shares are present and in absence of such device, we can recover secret information by simply stacking the shares. In comparison with other schemes, there is no pixel expansion and basis

matrix requirement for the share generation phase as the shares generated are meaningful and it appears to the intruder that the shares are not carrying any information, hence it is more secure. The experimental results demonstrate the capability of our scheme and it can be applied to a variety of real-time applications as well. Comparison with previous schemes reveals that our scheme has several benefits that were lacking in previous schemes.

References

1. Shamir, A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979)
2. Naor, M., Shamir, A.: Visual cryptography, Advances in Cryptography. In: EUROCRYPT 94, LNCS, vol. 950, pp. 1–12 (1994, 1995)
3. Kafri, O., Keren, E.: Encryption of pictures and shapes by random grids. Opt. Lett. **12**(6), 377–379 (1987)
4. Chen, T.H., Tsao, K.H.: Visual secret sharing by random grids revisited. Pattern Recognit. **42**(11), 2203–2217 (2009)
5. Chen, T.-H., Tsao, K.-H.: Threshold visual secret sharing by random grids. J. Syst. Softw. **84**(7), 1197–1208 (2011)
6. Wu, X., Sun, W.: Random grid-based visual secret sharing with abilities of OR and XOR decryptions. J. Vis. Commun. Image Represent. **24**(1), 48–62 (2013)
7. Yan, X., Wang, S., Niu, X., Yang, C.N.: Random grid-based visual secret sharing with multiple decryptions. J. Vis. Commun. Image Represent. **26**, 94–104 (2015)

EEG Seizure Detection from Compressive Measurements

Meenu Rani, S. B. Dhok and R. B. Deshmukh

Abstract Electroencephalogram (EEG) signal is a measure of electrical activity across the brain. For patients suffering from brain disorders like epilepsy, coma, sleep disorders, etc., this electrical activity is continuously monitored. For this, a minimum of 21 electrodes are required, which are placed across the scalp. These electrodes generate a lot of data to be processed for diagnosing the brain disease. Compressive sensing (CS), which is a newer sensing modality, has proved itself to be a better candidate for handling large amount of data to be as compared to the traditional sampling mechanism. CS generates far fewer samples than that suggested by Nyquist rate and still allows faithful reconstruction. The CS reconstruction employs complex nonlinear methods, which are very costly. Compressed signal processing (CSP), which is an advancement over CS, gives a direction to solve certain signal processing tasks from compressive measurements itself, without the need for reconstructing the original signal at all. In this paper, CSP has been used for detecting the presence or absence of epileptic seizure in the EEG signal. For this purpose, a feature extraction method is proposed for extracting the features from compressed EEG measurements. The performance of proposed method has been found to be surprisingly effective in this regard. All the experiments are done on the EEG database taken from physionet CHB-MIT using MATLAB.

Keywords Compressive sensing · Compressed signal processing · EEG monitoring · Random demodulator · Feature extraction · SVM

M. Rani (✉) · S. B. Dhok · R. B. Deshmukh
Visvesvaraya National Institute of Technology, Nagpur 440010, MH, India
e-mail: meenubanait@gmail.com

S. B. Dhok
e-mail: sanjaydhok@gmail.com

R. B. Deshmukh
e-mail: mona1810@yahoo.com

1   Introduction

The brain diseases like epilepsy, coma, sleep disorders, etc., can be diagnosed by monitoring the electrical activity across the brain. The signal obtained measuring this activity is known as electroencephalogram (EEG) signal. An abnormality in the brains' electrical activity is an indicator of certain disease. So, an accurate diagnosis of this activity is very important. For this purpose, these signals are sampled at a much higher rate. Then, this oversampled data is compressed before storage and/or transmission. The problem with this traditional sampling and processing scheme is that both sampling and compression stages consume a lot of power. The situation becomes more problematic in case of power constrained environment, like, remote health monitoring, where battery operated devices are used. Then finally, at the receiver end, the original signal is recovered back to extract the relevant clinical states. In this scenario, compressive sensing (CS) is able to perform better than the traditional sampling. It lowers the power consumption by sampling at a much lower rate compared to the Nyquist rate and thus avoids the need for further compression [1–3].

CS is relatively new, which was introduced in 2006 by Donoho, et al., for sampling signals at a very low rate [4, 5]. CS can easily be applied to the sparse or compressible signals. The sparse signals are those, which can be represented using fewer significant components, either in the original domain or in some transformed domain. A signal is called compressible, if its sorted components obey power law decay, when represented in the original domain or in some transformed domain. The major differences between traditional sampling and CS are: (i) the sampling rate in case of traditional sampling is decided by the highest frequency component present in the signal, while in CS, the sampling rate is decided by the sparsity of the signal of interest. (ii) traditional sampling uses a uniform sampling mechanism, on the other hand, CS is based on random sampling mechanism. These differences make CS capable of generating fewer measurements, which means, CS compresses the signal during sensing only. From these measurements, the original signal can be recovered faithfully by complex nonlinear techniques [6–9].

Compressed signal processing (CSP), is a technique to get rid of costly CS reconstruction for solving certain signal processing tasks. CSP is an advancement over CS, which efficiently solves inference problems like classification, detection, and estimation, by directly utilizing the compressive measurements of input signal. CSP is based on the information preserving property of compressive measurements, which is normally utilized in reconstructing the original signal back. The same property can also be utilized to solve the inference problems. For this purpose, either the raw compressive measurements can be directly used or certain feature extraction technique can be applied to extract features from these measurements [10–14]. A possible representation of CSP framework is shown in Fig. 1.

In this paper, CSP has been used for identifying the presence or absence of seizure in the EEG signal to diagnose the brain diseases. For this, a feature extraction method from compressive measurements is proposed. The efficiency of the proposed method

Fig. 1 A possible representation of CSP framework

for EEG seizure detection is evaluated using the simulations done in MATLAB, on the CHB-MIT database. The organization of the remainder of the paper is as: section II presents the background of CS acquisition and CSP. Section III describes the proposed scheme for EEG seizure detection. Section IV presents the supporting results, followed by discussion.

2 Background

CS samples and compresses a signal simultaneously by randomly subsampling the input signal. Therefore, CS does not require a separate compression stage. The number of compressive measurements which are necessary for further processing can be minimized by choosing incoherent bases for signal acquisition and for sparsifying the signal. For example, if the sparsifying basis is frequency domain then the acquisition domain should be time, because a signal having sparse representation in frequency domain, spreads out in the time domain [6–8]. The CS acquisition process can be represented in mathematical form by (1) as:

$$y = \varphi x, \tag{1}$$

where, $x \in \mathbb{R}^n$ is the input signal, $\varphi \in \mathbb{R}^{m \times n}$ or $\mathbb{C}^{m \times n}$ is the CS measurement matrix and $y \in \mathbb{R}^m$ or $\mathbb{C}^m$ is the output vector. Here, n represents number of samples in input signal, m represents number of measurements and $m \ll n$.

CSP is an advancement over CS, in which features are directly extracted from compressive measurements. The features so extracted can be used for solving the inference problems directly. CSP is based on the information preserving property of compressive measurements. These measurements give a unique signature at the output, corresponding to the input signal frequencies. These unique signatures enables the reconstruction and can also be used for solving some signal processing tasks which do not require reconstruction [10–14].

3 Proposed Model

The proposed model for detecting the seizure from the EEG signal is shown in Fig. 2. The input EEG signal is first acquired segment wise using random demodulator (RD) technique of CS. Although, other CS acquisition strategies are available, but, RD is
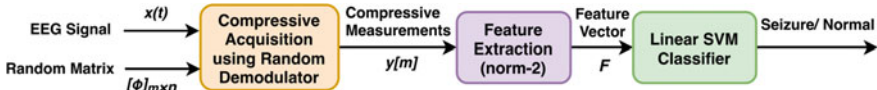
Fig. 2 The proposed model for EEG seizure detection
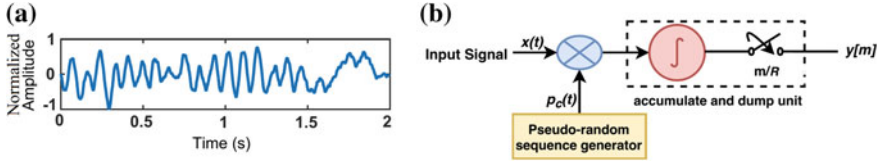
Fig. 3 EEG signal and its acquisition method: **a** segment of EEG signal taken from physionet CHB-MIT database, **b** EEG signal acquisition via random demodulator

chosen as a sampling scheme here because of its simpler and efficient architecture [9]. The input EEG signal is taken from the physionet CHB-MIT database and is shown in Fig. 3a [15]. The RD sampling mechanism is shown in Fig. 3b. RD first randomizes the input signal by multiplying with a pseudorandom sequence of ±1s. The randomized signal is then passed through an integrator, which accumulates the signal samples and serves as a low pass filter (LPF). The output of the integrator, which is a low frequency signal, is then passed through the sample and hold circuit. The output of this stage is the compressive measurements, which can be easily stored or transmitted [16].

The above described operation of RD can be represented in matrix form by (2) and (3). In this, P is a diagonal matrix having diagonal elements as pseudorandom sequence, $p_c(t)$ and H is the accumulate and dump matrix. The number of terms that needs to be accumulated to generate single measurement is given by $R = \lfloor (n/m) \rfloor$.

$$P = \begin{bmatrix} p_1 & & \\ & \ddots & \\ & & p_n \end{bmatrix}; \quad H = \begin{bmatrix} 111\cdots & & \\ & 111\cdots & \\ & & 111\cdots \end{bmatrix} \tag{2}$$

$$\left.\begin{aligned} \tilde{x} &= Px \\ y &= H\tilde{x} = \varphi x \\ \varphi &= HP \end{aligned}\right\}. \tag{3}$$

After getting the compressive measurements, features are extracted from these measurements. This step is required in order to obtain a dataset for the classifier, on which classification can be easily performed. Although, one directly perform the classification on compressive measurements as it is, to reduce the cost of extracting the features. But, then the cost has to be paid in terms of low accuracy, even at the lower undersampling factors. So, in order to obtain higher accuracy and to support higher undersampling, features need to be extracted from compressive measurements. In

this paper, the norm-2 of compressive measurements is computed segment wise and is proposed as a feature for classifying the EEG dataset. This process is then repeated for all the EEG channels, in order to obtain the feature set.

For detecting the seizure from the feature set obtained above, a linear SVM classifier is used. The training and testing of this classifier have been done using k-fold cross validation. The classifier used here is a binary classifier, which detects either the presence or absence of seizure in the EEG signal.

4   Results and Discussion

The EEG data is taken from physionet CHB-MIT database [15] and is processed in the segments of duration 2 seconds each. In this database, the 2 second EEG segment is equivalent to 512 samples and the total number of samples processed are 921600. The compressive measurements are obtained using RD for each segment of EEG signal, for various undersampling factors like 2, 4, 8, 16, and 32. Corresponding to these undersampling factors, the number of compressive measurements (m), generated are 256, 128, 64, 32, and 16, respectively. For each set of compressive measurements so obtained, the norm-2 is computed to generate a feature set. After applying this process on the complete EEG signal, the size of the feature vector obtained corresponding to a single undersampling factor is 1800. Then, on this feature set, a linear SVM classifier is trained and tested using tenfold cross validation. In the tenfold cross validation, the data is divided into 10 equal parts. Each of ten parts recursively becomes test set with the remaining nine parts used as training sets. For different undersampling factors the testing accuracy obtained is shown in Table 1. For comparison purpose, the accuracy obtained by directly classifying the compressive measurements without feature extraction is also shown. The accuracy obtained by directly classifying the compressive measurements is much less compared to the case when feature extraction is used. This proves the effectiveness of the proposed method.

Table 1   Comparison of testing accuracy for the case of classification done directly on compressive measurements with the case when features are extracted, for different undersampling ratios

Undersampling factor	Classification accuracy obtained from compressive measurements	
	Without feature extraction (%)	With feature extraction (%)
2	30.5	99.1
4	28.2	99.1
8	27.8	99.1
16	24.6	99.1
32	20.8	98.9
64	20.1	98.8

5 Conclusion

Compressed signal processing is an efficient way of solving inference problems like classification, detection, and estimation. CSP is an advancement over CS, which gives a way of solving some signal processing problems directly using compressive measurements without reconstruction at all or without complete reconstruction. These problems can be solved in two ways: by extracting feature from compressive measurements or by using the raw compressive measurements to reduce the cost of feature extraction. In this paper, CSP is used for EEG seizure detection. The features are extracted from compressed EEG measurements using norm-2 and then classification is done using linear SVM classifier. The testing accuracy obtained using this proposed method is compared with the accuracy obtained by directly classifying the compressive measurements. It has been found that the proposed method astonishingly outperforms the direct method.

References

1. Aviyente, S.: Compressed Sensing Framework for EEG Compression. IEEE/SP 14th Workshop on Stat. Sig. Proc., Madison, WI, USA, pp. 181-184, 2007. DOI: https://doi.org/10.1109/SSP.2007.4301243
2. Zhang, Z., et al.: Compressed Sensing of EEG for Wireless Telemonitoring With Low Energy Consumption and Inexpensive Hardware. IEEE Trans. on Biomed. Engg. **60**(1), 221–224 (2013). https://doi.org/10.1109/TBME.2012.2217959
3. Abdulghani, A.M., et al.: Compressive sensing scalp EEG signals: implementations and practical performance. E. Med Biol Eng Comp **50**, 1137–1145 (2012). https://doi.org/10.1007/s11517-011-0832-1
4. Candès, E.J., et al.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Inf. Theory **52**(2), 489–509 (2006). https://doi.org/10.1109/TIT.2005.862083
5. Donoho, D.L.: Compressed sensing. IEEE Trans. on Inf. Theory **52**(4), 1289–1306 (2006). https://doi.org/10.1109/TIT.2006.871582
6. Baraniuk, R.G.: Compressive Sensing [Lecture Notes]. IEEE Signal Process. Mag. **24**(4), 118–121 (2007). https://doi.org/10.1109/MSP.2007.4286571
7. Candès, E.J., Wakin, M.B.: An Introduction to Compressive Sampling. IEEE Sig. Process. Mag. **25**(2), 21–30 (2008). https://doi.org/10.1109/MSP.2007.914731
8. Baraniuk, R., et al.: An Introduction to Compressive Sensing. OpenStax-CNX. April 2, 2011. http://legacy.cnx.org/content/col11133/1.5/
9. Rani, M., et al.: A Systematic Review of Compressive Sensing: Concepts. Implementations and Applications. IEEE Access **6**, 4875–4894 (2018). https://doi.org/10.1109/ACCESS.2018.2793851
10. Haupt, J., et al.: Compressive Sampling for Signal Classification. Fortieth Asilomar Conf. on Sig., Sys. and Comp., Pacific Grove, CA, pp. 1430-1434, 2006
11. Duarte, M. F., et al.: Sparse Signal Detection from Incoherent Projections. IEEE Int. Conf. on Acou., Speech and Sig. Process. Proceed., Toulouse, 2006, pp. III-III
12. Haupt, J. and Nowak, R.: Compressive Sampling for Signal Detection. IEEE Int. Conf. on Acou., Speech and Sig. Process. - ICASSP07, Honolulu, HI, pp. III-1509- III-1512, 2007
13. Davenport, M. A., et al.: Signal Processing With Compressive Measurements. IEEE J. of Sel. Top. in Sig. Proces., vol. 4, no. 2, pp. 445-460, April 2010

14. Park, J.Y., et al.: Modal Analysis With Compressive Measurements. IEEE Trans. on Sig. Process. **62**(7), 1655–1670 (2014)
15. Goldberger AL, et al.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13)
16. Tropp, J.A., et al.: Beyond Nyquist: Efficient Sampling of Sparse Bandlimited Signals. IEEE Trans. on Inf. Theory **56**(1), 520–544 (2010). https://doi.org/10.1109/TIT.2009.2034811

A Medical Diagnostic Information System with Computing with Words Using Hesitant Fuzzy Sets

Rajkrishna Mondal, **Akshay Verma** and **Pushpendra Kumar Gupta**

Abstract   Handling uncertainty in a medical diagnostic information system is a challenging and difficult problem. The recent Hesitant Fuzzy set has been introduced to control hesitant situations in which experts have hesitated about their opinion for rating a state of the system. Most of the diagnostics is discussed in a quantitative setting. But in some situations, it is difficult for rating quantitatively and calculation complexity is higher than the qualitative setting. In this article, to overcome such difficulty, we have used the Hesitant Fuzzy Linguistic Approach to design this diagnostic information system.

Keywords   Computing with words · Diagnostic information system · Hesitant Fuzzy sets · Linguistic decision-making

1   Introduction

Nowadays, due to the lack of appropriate data sets dealing with uncertainty, vague, imprecise information is a challenging problem in many decision-making fields. In mathematics, a lot of tools like Probability Theory, etc. are used to solve such difficulty. But in nature, uncertainty sometimes is not probabilistic; in such situations fuzzy system [13] is the most acceptable mathematical tool.

R. Mondal (✉)
Department of Mathematics, Motilal Nehru National Institute
of Technology Allahabad, Allahabad 211004, India
e-mail: rajkrishna.math@gmail.com

A. Verma · P. K. Gupta
Department of Electronics and Communication Engineering,
Motilal Nehru National Institute of Technology Allahabad,
Allahabad 211004, India
e-mail: akshayv.verma49@gmail.com

P. K. Gupta
e-mail: pushpendraucer@gmail.com

But there have been many situations where imprecise information occurs simultaneously many times. If we take an example like in a project there have been many experts to evaluate a final decision and each expert has different opinions (it may occur simultaneously in linguist form like one gives 'good' and another gives 'medium', etc.) from each other; in such situations, ordinary fuzzy system fails to answer this question.

So, Hesitant Fuzzy Sets [9] has been introduced to handle this kind of situations. Most of the fuzzy systems are based on the quantitative setting, but in nature uncertainty, vague, imprecise information is rather qualitative. In such qualitative situations, using linguistic approaches with fuzzy set theory [14] is given the more appropriate result with less computational complexity. In other languages, sometimes it is called 'Computing with Words' [4].

Through the review of some literature on fuzzy linguistic approaches [3, 5, 10, 11, 15], it is clear that for modelling some complex systems, there are some limitations because single and simple linguistic terms are enclosed here. Sometimes, we have seen that there are some systems where the decision maker cannot provide easily a single linguistic term for stating his knowledge information, where he needs to provide several linguistic terms like suppose, in a group of patients, we want to know who the most dangerous patient for a certain disease is and for each criterion for each patient, there are different opinions from different doctors. Therefore, to overcome such situations in the medical diagnostic information system, we have used Hesitant Fuzzy Set theory through linguistic approaches and we have developed an algorithm for taking more appropriate decisions for such situations.

We have arranged this article as follows. In Sect. 2, we discussed some preliminary definitions about Hesitant Fuzzy Linguistic Term Sets (HFLTS) and some basic properties, operations on Computing with Words (CW). In Sect. 3, we modelled a medical diagnostic information system for multi-criteria decision-making. In Sect. 4, we have discussed some case studies using the proposed algorithm. At last, in Sect. 5, we have justified our results and drawn a proper conclusion.

2 Preliminary

2.1 *Hesitant Fuzzy Set (HFS)* [9]

Let U be a set of universe and $X \subset U$. Then Hesitant Fuzzy Set (HFS) of X over U is defined by

$$Hes_U(X) = \{< u, h_X(u) >| \ u \in U\} \qquad (1)$$

where $h_X(u)$ is set of possible membership degrees of u in [0, 1] to the set X.

Example 1 If $U = \{1, 2, 3, 4, 5\}$ and $X_1 = \{1, 2, 3\}$ and $X_2 = \{2, 3, 4\}$ be two subsets of U. In this example, we randomly choose the membership values in [0, 1] for showing two different HFSs in U.

$$Hes_U(X_1) = \left\{ \begin{array}{c} < 1, \{0.9, 0.4, 0.7, 0.8\} > \\ < 2, \{0.4, 0.5, 0.9\} > \\ < 3, \{0.1, 0.9, 0.2, 0.8, 0.4\} > \end{array} \right\}$$

$$Hes_U(X_2) = \left\{ \begin{array}{c} < 2, \{0.2, 0.5\} > \\ < 3, \{0.3, 0.5, 0.6\} > \\ < 4, \{0.2, 0.5, 0.7, 0.9\} > \end{array} \right\}$$

2.2 Hesitant Fuzzy Linguistic Term Set (HFLTS) [7]

Let for a state of a system, $S = \{s_0, s_1, \ldots, s_n\}$ be a universe set of the ordered linguistic terms. Then an HFLTS with respect to S is denoted by H_S, a finite order consecutive terms from S.

Example 2 Let for a system state '**Age**', $S = \{s_1 : VeryYoung, s_2 : Young, s_3 : MiddleAged, s_4 : Aged, s_5 : VeryAged, s_6 : Old, s_7 : VeryOld\}$ be the universe linguistic term set.

χ_1 and χ_2 be two different linguistic expression on S, then for two different HFLTS may occur as
$H_S^1 = H_S(\chi_1) = \{s_3 : MiddleAged, s_4 : Aged, s_5 : VeryAged\}$
$H_S^2 = H_S(\chi_2) = \{s_5 : VeryAged, s_6 : Old, s_7 : VeryOld\}$

Let H_S, H_S^m, H_S^n, where m,n are two different integers, be the three different HFLTS,

1. Complement: $(H_S)^c = \{s_i \mid s_i \in S \text{ and } s_i \notin H_S\}$
2. Union: $H_S^m \cup H_S^n = \{s_i \mid s_i \in H_S^m \text{ or } s_i \in H_S^n\}$
3. Intersection: $H_S^m \cap H_S^n = \{s_i \mid s_i \in H_S^m \text{ and } s_i \in H_S^n\}$
4. Envelope: $\text{env}(H_S) = [H_{S^-}, H_{S^+}]$,
   where $H_{S^-} = min(s_i), H_{S^+} = max(s_i), s_i \in H_S$

2.3 Context-Free Grammar

The context-free grammar(G) is used for generating the linguistic expressions, i.e. the sentences [1, 2]. G is mainly defined in 4-tuple (V_N, V_T, I, P) where
V_N : Non-terminal symbols, V_T : Terminal symbols, I : Starting symbols,
P : Production Rules. $P = \{(PrimaryTerm) \text{ or } (UnaryRelation)$
$(PrimaryTerm) \text{ or } (BinaryRelation)(PrimaryTerm)(Conjunction)$
$(PrimaryTerm)\}$
$(PrimaryTerm) = s_i \in S$
$(UnaryRelation) = LowerThan \text{ or } GreaterThan$
$(BinaryRelation) = Between$
$(conjunction) = And$

Example 3 Let $S = \{s_i \mid i = 1, 2, \ldots, 7\}$ be defined as in Example 2, then there are various types of linguistic expressions produced by a context-free grammar(G) as

$$\Theta_1 = \underbrace{\text{Aged}}_{\text{primary term}}$$

$$\Theta_2 = \underbrace{\text{lower than}}_{\text{unary}}\underbrace{\text{Aged}}_{\text{primary}}$$

$$\Theta_3 = \underbrace{\text{between}}_{\text{binary}}\underbrace{\text{Aged}}_{\text{primary}}\underbrace{\text{and}}_{\text{conjunction}}\underbrace{\text{Old}}_{\text{primary}}$$

2.4 Transformation of Linguistic Expressions into HFLTS [7]

The linguistic expressions Θ_i generated by G(a context-free grammar) is transformed into HFLTS by the function E_G.

Let $\Theta = \{\Theta_i \mid i \in \mathbf{N}\}$ be the set of linguistic expressions; $S = \{s_i \mid i \in \mathbf{N}\}$, where $\mathbf{N}$ is the set of +ve integers, be the set of ordered linguistic terms and H_S be set of HFLTS on S. Then the function $E_G : \Theta \mapsto H_S$ is defined by

$$E_G(\Theta_i) = \{s_k, s_{k+1}, \ldots, s_{k+l} \mid s_m \in S, m = k, k+1, \ldots, k+l; k, l \in \mathbf{N}\}$$

where E_G follows these rules

1. $E_G(s_m) = \{s_m \mid s_m \in S\}$
2. $E_G(LessThan\ s_m) = \{s_p \mid \forall s_p \in S \text{ and } s_p \preceq s_m\}$
3. $E_G(GreaterThan\ s_n) = \{s_p \mid \forall s_p \in S \text{ and } s_p \succeq s_n\}$
4. $E_G(Between\ s_m\ And\ s_n) = \{s_k \mid \forall s_k \in S \text{ and } s_m \preceq s_k \preceq s_n\}$

Example 4 As in Example 2, let our state of the system be 'Age' and $S = \{s_i \mid i = 1 : 7\}$ be as defined in Example 2.

Here, we have taken three different linguistic expressions Θ_i, $i = 1, 2, 3$ that have been taken in Example 3. After applying E_G on Θ_i, $i = 1, 2, 3$, the output will be
$E_G(\Theta_1) = E_G(Aged) = \{s_4\}$
$E_G(\Theta_2) = E_G(LowerThan\ Aged) = \{s_1, s_2, s_3, s_4\}$
$E_G(\Theta_3) = E_G(Between\ Aged\ and\ Old) = \{s_4, s_5, s_6\}$

2.5 HFLTS Aggregation Operators

There are two kinds of linguistic expressions, one is optimistic and the other is pessimistic. For balancing these two kinds of expressions, here we have used **Min-Upper** and **Max-Lower** aggregation operators on HFLTS.

Let $S = \{s_i \mid i \in \mathbf{N}\}$ be the set of linguistic terms and $H_S = \{H_S^i \mid i = 1, 2, \ldots, n\}$ be the collection of HFLTS on S where $H_S^i = \{s_{k_i}, s_{k_{i+1}}, \ldots, s_{k_{l(H_S^i)}} \mid \forall s_{k_i} \in S; k_i, k_{i+1}, \ldots, k_{l(H_S^i)} \in \mathbf{N}\}$ where $l(H_S^i)$ is the length of $H_S^i, i = 1, 2, \ldots, n$. H_{S^+}. Each aggregation operator is computed in two steps:

Min-Upper:

1. $H_{S^+}^i = Max\{s_{k_i}, s_{k_{i+1}}, \ldots, s_{k_{l(H_S^i)}}\}, i = 1, 2, \ldots, n$
2. $H_{S_{Min}^+} = Min\{H_{S^+}^1, H_{S^+}^2, \ldots, H_{S^+}^n\}$

Max-Lower:

1. $H_{S^-}^i = Min\{s_{k_i}, s_{k_{i+1}}, \ldots, s_{k_{l(H_S^i)}}\}, i = 1, 2, \ldots, n$
2. $H_{S_{Max}^-} = Max\{H_{S^-}^1, H_{S^-}^2, \ldots, H_{S^-}^n\}$

Example 5 Let $H_S^1 = \{s_1, s_2, s_3, s_4\}$ and $H_S^2 = \{s_4, s_5, s_6\}$ be two different HFLTSs. Then, $H_{S^+}^1 = \{s_4\}, H_{S^+}^2 = \{s_6\}$ and $H_{S^-}^1 = \{s_1\}, H_{S^-}^2 = \{s_4\}$
So, $H_{S_{Min}^+} = \{s_4\}$ and $H_{S_{Max}^-} = \{s_4\}$

3 Medical Diagnosis Decision-Making Model

3.1 Problem Formulation

Suppose in our diagnostic information system, there are n alternatives $X = \{x_1, x_2, \ldots, x_n\}$ and each alternative is defined in m criteria $C = \{c_1, c_2, \ldots, c_m\}$. Let, in this system, the set of linguistic term be $S = \{s_1, s_2, \ldots, s_g\}$.

A context-free grammar G, produces linguistic expressions $\Theta_{ij}(x_i, c_j)$ for alternative $x_i, i = 1, \ldots, n$ corresponding to criteria $c_j, j = 1, \ldots, m$ (shown in Fig. 1).

Alternatives
$X = \{x_1, x_2, \ldots, x_n\}$

A Context free Grammar (G)
$S = \{s_1, s_2, \ldots, s_g\}$

Linguistic Expression
$\theta_{ij}(x_i, c_j); i = 1, 2, \ldots, n, j = 1, 2, \ldots, m$

Criteria
$C = \{c_1, c_2, \ldots, c_m\}$
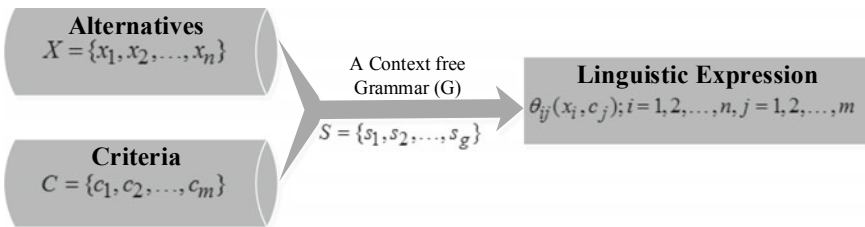
Fig. 1 Generating the linguistic expressions in diagnosis model

3.2 Model Designing

We have divided our model into three phases: (1) Transformation, (2) Aggregation, (3) Exploitation (shown in Fig. 2).

Transformation Phase: In this phase, through G produced linguistic expressions $\Theta_{ij}(x_i, c_j)$ is converted to HFLTS by the transformation function E_G, i.e.

$$E_G(\Theta_{ij}(x_i, c_j)) = \{s_k \mid s_k \in S\}$$

Aggregation Phase: From the previous phase generated HFLTS, there are two points of view: *optimistic* and *pessimistic* [8], to balance both approximations; here, we have used **Min-Upper** and **Max-Lower**: two aggregation operators. Let $E_G(\Theta_{ij}(x_i, c_j)) = H_S^j(x_i); i = 1 : n; j = 1 : m$.

Min-Upper:

(a) $H_{S+}(x_i) = \{H_{S+}^j \mid j = 1 : m\}, i = 1 : n$
(b) $H_{S_{Min}^+}(x_i) = Min\{H_{S+}^j \mid j = 1 : m\}, i = 1 : n$

Max-Lower:

(a) $H_{S-}(x_i) = \{H_{S-}^j \mid j = 1 : m\}, i = 1 : n$
(b) $H_{S_{Max}^-}(x_i) = Max\{H_{S-}^j \mid j = 1 : m\}, i = 1 : n$

After calculating aggregation operators, we need to create linguistic intervals for each alternative $x_i, i = 1, \ldots, n$.
$$H_{Max}^* = Max\{H_{S_{Min}^+}, H_{S_{Max}^-}\}$$
$$H_{Min}^* = Min\{H_{S_{Min}^+}, H_{S_{Max}^-}\}$$

So finally, the generated linguistic interval for the alternative x_i will be

$$H^*(x_i) = [H_{Min}^*(x_i), H_{Max}^*(x_i)], i = 1, \ldots, n$$
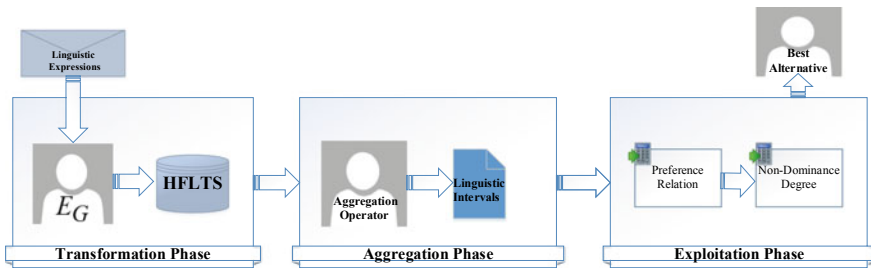
Fig. 2 Decision-making model

Exploitation Phase: In this phase, we will decide the best alternative. Here, we have taken the linguistic intervals for each alternative from the previous phase and through the following three steps, we have performed this phase:

Step 1: Here, we have built the preference relation P for each pair of alternatives [6, 12].

Let $x_i = [s_{iL}, s_{iR}]$, $x_j = [s_{jL}, s_{jR}]$; $i \neq j$ where $s_{iL}, s_{iR}, s_{jL}, s_{jR} \in S$, be two linguistic intervals for alternatives x_i and x_j and $ind(x_i) = i$ be the indexing function.

$$P = \begin{pmatrix} - & p_{12} & \cdots & p_{1n} \\ p_{21} & - & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & - \end{pmatrix}_{n \times n}$$

where $p_{ij} = P(x_i \succeq x_j) = \frac{Max\{0, ind(S_{iR}) - ind(S_{jL})\} - Max\{0, ind(S_{iL}) - ind(S_{jR})\}}{[ind(S_{iR}) - ind(S_{jL})] - [ind(S_{iL}) - ind(S_{jR})]}$

Step 2: Here, we calculated non-dominance degree for each alternative x_i such that

$$NDD_i = Min\{1 - p_{ji} \mid i \neq j, j = 1 : n\}, i = 1 : n$$

Step 3: The best alternative is selected as

$$x^{Best} = \{x_j \mid NDD_j = Max(NDD_i \mid i = 1 : n)\}$$

4 Case Studies

Suppose in a diagnostic centre, there are four patients $P = \{p_1, p_2, p_3, p_4\}$. Doctor of this diagnostic centre has detected five symptoms as the criteria for each patient $C = \{c_1 : Sugar\ Level, c_2 : BMI, c_3 : Cholesterol, c_4 : Daily\ Burning\ Cal., c_5 : Age\}$.

Let the set of linguistic terms be $S = \{s_1 : Nothing(n), s_2 : VeryLow(vl), s_3 : Low(l), s_4 : Medium(m), s_5 : High(h), s_6 : VeryHigh(vh), s_7 : Excessive(e)\}$ for detecting their diabetes level.

Through their medical report and according to their opinion, we have built the linguistic expression table Θ of order 4×5; that has been shown in Table 1.

Table 1 Linguistic expression for each patients

Θ	c_1	c_2	c_3	c_4	c_5
p_1	LessThan l	Between vl and m	between l and m	Between m and vh	Between vl and m
p_2	LessThan vl	Between l and h	m	Between m and h	Between m and h
p_3	GreaterThan h	GreaterThan vh	Between h and vh	LessThan l	GreaterThan vh
p_4	Between h and vh	h	Between m and vh	LessThan l	Between h and vh

Table 2 Transformed formed into HFLTS

$H_S^j(p_i)$	c_1	c_2	c_3	c_4	c_5
p_1	$\{n, vl, l\}$	$\{vl, l, m\}$	$\{l, m\}$	$\{m, h, vh\}$	$\{vl, l, m\}$
p_2	$\{n, vl\}$	$\{l, m, h\}$	$\{m\}$	$\{m, h\}$	$\{m, h\}$
p_3	$\{h, vh, e\}$	$\{vh, e\}$	$\{h, vh\}$	$\{n, vl, l\}$	$\{vh, e\}$
p_4	$\{h, vh\}$	$\{h\}$	$\{m, h, vh\}$	$\{n, vl, l\}$	$\{h, vh\}$

Table 3 $H_{S_{Min}^+}$ calculation table

$H_{S^+}^j(p_i)$	c_1	c_2	c_3	c_4	c_5	$H_{S_{Min}^+}(p_i)$
p_1	$\{l\}$	$\{m\}$	$\{m\}$	$\{vh\}$	$\{m\}$	$\{l\}$
p_2	$\{vl\}$	$\{h\}$	$\{m\}$	$\{h\}$	$\{h\}$	$\{vl\}$
p_3	$\{e\}$	$\{e\}$	$\{vh\}$	$\{l\}$	$\{e\}$	$\{l\}$
p_4	$\{vh\}$	$\{h\}$	$\{vh\}$	$\{l\}$	$\{vh\}$	$\{l\}$

Table 4 $H_{S_{Max}^-}$ calculation table

$H_{S^-}^j(p_i)$	c_1	c_2	c_3	c_4	c_5	$H_{S_{Max}^-}(p_i)$
p_1	$\{n\}$	$\{vl\}$	$\{l\}$	$\{m\}$	$\{vl\}$	$\{m\}$
p_2	$\{n\}$	$\{l\}$	$\{m\}$	$\{m\}$	$\{m\}$	$\{m\}$
p_3	$\{h\}$	$\{vh\}$	$\{h\}$	$\{n\}$	$\{vh\}$	$\{vh\}$
p_4	$\{h\}$	$\{h\}$	$\{m\}$	$\{n\}$	$\{h\}$	$\{h\}$

Table 5 Linguistic intervals for each alternatives

	p_1	p_2	p_3	p_4
$H^*(p_i)$	$[l, m]$	$[vl, m]$	$[l, vh]$	$[l, h]$
In $s_i \in S$ format	$[s_3, s_4]$	$[s_2, s_4]$	$[s_3, s_6]$	$[s_3, s_5]$

In Table 1, we have used E_G function to transform it into HFLTS, that has been shown in Table 2.

Now in Table 2, we have applied **Min-Upper** and **Max-Lower** operators to calculate $H_{S_{Min}^+}$ and $H_{S_{Max}^-}$, that has been given in Tables 3 and 4, respectively.

Now for Tables 3 and 4, we create the linguistic intervals $H^*(p_i)$ for each alternative p_i, $i = 1 : 4$, as shown in Table 5.

Now in the Exploitation Phase, we build the preference relation P for each different pair of alternatives and non-dominance degree NDD_i for each alternative p_i, $i = 1 : 4$, that has been shown in Table 6.

Finally, we see that the maximum NDD_i value is 0.6000, which is responsible for p_3 alternative.

Table 6 Preference relation matrix P and *NDD* value

p	p_1	p_2	p_3	p_4
p_1	NaN	6.666667e-01	2.500000e-01	3.333333e-01
p_2	3.333333e-01	NaN	2.000000e-01	2.500000e-01
p_3	7.500000e-01	8.000000e-01	NaN	6.000000e-01
p_4	6.666667e-01	7.500000e-01	4.000000e-01	NaN
NDD_i	0.2500	0.2000	0.6000	0.4000

From the above output result, we can conclude that Patient 3 is in the most dangerous level among these four patients.

5   Conclusion and Future Work

In this article, we have discussed a medical diagnostic information system using the concept of Hesitant Fuzzy Linguistic Term Set. When the medical experts are confused about levelling specific criteria of a patient, in such a hesitant situation the proposed diagnostic information system is more useful. Using the Hesitant Fuzzy Set with the Linguistic approach makes this model increase flexibility.

The designed diagnostic information system and its performance have been cross-verified by the medical experts and therefore we are of the view that the proposed diagnostic information system will be highly useful to the patient to make him aware of the current status of disease level.

Still this medical information system needs some improvements; here, we have just introduced the algorithm and discussed a case study to justify its output results and verified it with the help of doctor suggestions. In future, we will improve it to make this model more flexible and will discuss its complexity and benefits from other existing techniques.

References

1. Bonissone, P.P.: A fuzzy sets based linguistic approach: theory and applications. In: Approximate Reasoning in Decision Analysis, pp. 329–339 (1982)
2. Bordogna, G., Pasi, G.: A fuzzy linguistic approach generalizing boolean information retrieval: a model and its evaluation. J. Am. Soc. Inf. Sci. **44**(2), 70–82 (1993)
3. Dong, Y., Xu, Y., Yu, S.: Computing the numerical scale of the linguistic term set for the 2-tuple fuzzy linguistic representation model. IEEE Trans. Fuzzy Syst. **17**(6), 1366–1378 (2009)

4. Herrera, F., Alonso, S., Chiclana, F., Herrera-Viedma, E.: Computing with words in decision making: foundations, trends and prospects. Fuzzy Optim. Decis. Mak. **8**(4), 337–364 (2009)
5. Mendel, J.M.: An architecture for making judgments using computing with words. Int. J. Appl. Math. Comput. Sci. **12**, 325–335 (2002)
6. Orlovsky, S.: Decision-making with a fuzzy preference relation. Fuzzy Sets Syst. **1**(3), 155–167 (1978)
7. Rodriguez, R.M., Martinez, L., Herrera, F.: Hesitant fuzzy linguistic term sets for decision making. IEEE Trans. Fuzzy Syst. **20**(1), 109–119 (2012)
8. Sengupta, A., Pal, T.K.: On comparing interval numbers. Eur. J. Oper. Res. **127**(1), 28–43 (2000)
9. Torra, V.: Hesitant fuzzy sets. Int. J. Intell. Syst. **25**(6), 529–539 (2010)
10. Türkşen, I.B.: Type 2 representation and reasoning for CWW. Fuzzy Sets Syst. **127**(1), 17–36 (2002)
11. Wang, J.H., Hao, J.: A new version of 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans. Fuzzy Syst. **14**(3), 435–445 (2006)
12. Wang, Y.M., Yang, J.B., Xu, D.L.: A preference aggregation method through the estimation of utility intervals. Comput. Oper. Res. **32**(8), 2027–2049 (2005)
13. Zadeh, L.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965). https://doi.org/10.1016/S0019-9958(65)90241-X
14. Zadeh, L.: Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. part 1 inform. sci. 8, 199–249; part 2. inform. sci. 8, 301-353; part 3. inform. sci. 9, 43-80 **8**, 199–249 (01 1975)
15. Zhou, S.M., John, R.I., Chiclana, F., Garibaldi, J.M.: On aggregating uncertain information by type-2 owa operators for soft decision making. Int. J. Intell. Syst. **25**(6), 540–558 (2010)

Noise Cancelation Using Adaptive Filter

Akhilesh Kumar Ravat, Amit Dhawan and Manish Tiwari

Abstract Adaptive filtering creates one of the core technologies in the field of the digital signal processing and finds various applications in the area of science and technology, viz., adaptive noise cancelation, echo cancelation, channel equalization, bio-medical signal processing, etc. The principal objective of the noise cancelation is based on elimination of noise from audio as well as ECG (Electrocardiogram) signals. In this paper, an adaptive ECG filter is introduced to reduce the noise originated by body artifacts and exterior systems. The type of noises include interference caused by power line, interference caused by other electronic equipment, noise from electrode contact, and removing of movement of patient by adaptive filter to produce best results.

Keywords Noise · Adaptive algorithm · Adaptive filter · ANC · LMS · NLMS · ECG

1 Introduction

When the information is transmitted from source to destination, different types of noise are added automatically to the signal from the surrounding. The noisy signal has two components: one carries the information of the useful signal, i.e., the interested signal and the other carries random error or noise which is superimposed on the interested signal. These random errors are unwanted because they decrease the accuracy and precision of the measured signal. The transfer function of an adaptive filter is controlled by variable parameters which are adjusted according to optimization

A. K. Ravat (✉) · A. Dhawan · M. Tiwari
Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad 211004, Uttar Pradesh, India
e-mail: akravat1991@gmail.com

A. Dhawan
e-mail: dhawan@mnnit.ac.in

M. Tiwari
e-mail: mtiwari@mnnit.ac.in

algorithms. Adaptive filter is generally digital filter depending on the complexity of optimization algorithms [1–6]. The adaptive filter has become one of the most proposed solutions to minimize the signal degradation due to predictable and unpredictable noisy signal.

Due to the self-modifying behavior of the adaptive filter, it can change its frequency response to change its action with respect to time. Due to this ability, the adaptive filter has been used in different application like radar signal processing, navigation systems, communication, channel equalization, biomedical signal processing, and echo cancelation [5]. The transfer function of adaptive filter is controlled by variable parameters (stability, tracking, steady-state error, computational complexity, and convergence) [1].

An adaptive filter is clarified by following points:

- First, the signal is processed by filter.
- The structure clarifies that the output signal is related to its input signal.
- It also indicates how the parameter is changed for the adjustment of its transfer function.
- Lastly, the adaptive algorithm also narrates how parameters are modified due to varying of the time instant [1].

The functional diagram of an adaptive filter is given in Fig. 1.

$x(n) = \text{noisy signal} + \text{desired signal}$
$x(n) = [x(n), x(n-1), x(n-2)\ldots\ldots\ldots x(n-K-1)]^T$
where K is the filter order.
The coefficients for the filter are given as
$w(n) = [w_n(0), w_n(1), w_n(2)\ldots\ldots w_n(K-1)]^T$
$w(n)$ is the tap weight vector.
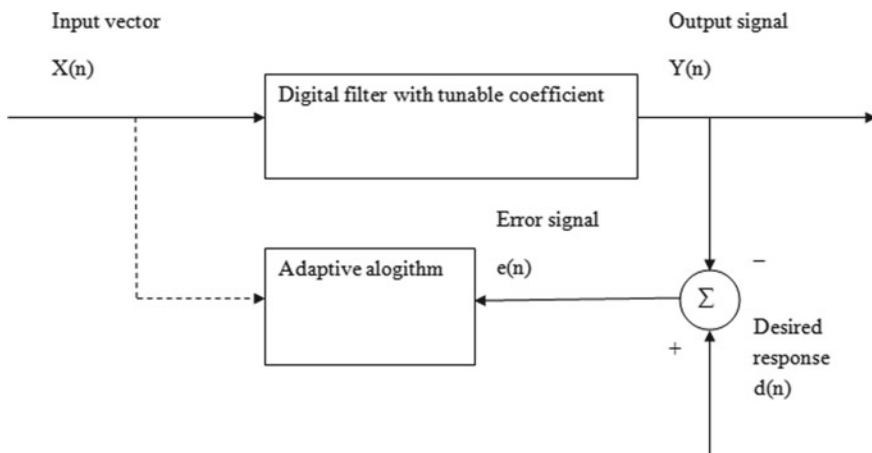
Fig. 1 Functional diagram of adaptive filter

Output $= w(n)^T x(n)$

Error signal $e(n) =$ desired signal—output

$w(n + 1) = w(n) + \delta.x(n).e(n)$

Electrocardiogram

Electrocardiogram is the physical performance of the electrical behavior generated by the heart muscles. ECG is the electrical activity of the heart. ECG is used to determine the rate and regularity of heartbeats, and the presence of any injury to the heart. An electrocardiogram is a graphical record taken by an electrocardiograph which records the electrical activity of the heart over a period of time [7]. The signal is obtained by measuring electrical potentials between different positions of the body.

Adaptive filter technique is required to diminish the noise from noisy ECG signal. ECG is an essential parameter for observing heart activity.

Interference caused by different sources are given in the example:

- Interference caused by power line
- Noise from muscle contraction
- Movement of patient
- Electrosurgical noise
- Noise from electrode contact (Fig. 2).

In this diagram, the **P-Wave** is produced by muscle contraction of atria, the **QRS Wave** denotes the ending of atria contraction and the beginning of ventricular contraction and **T Wave** denotes the ending of the ventricular contraction.
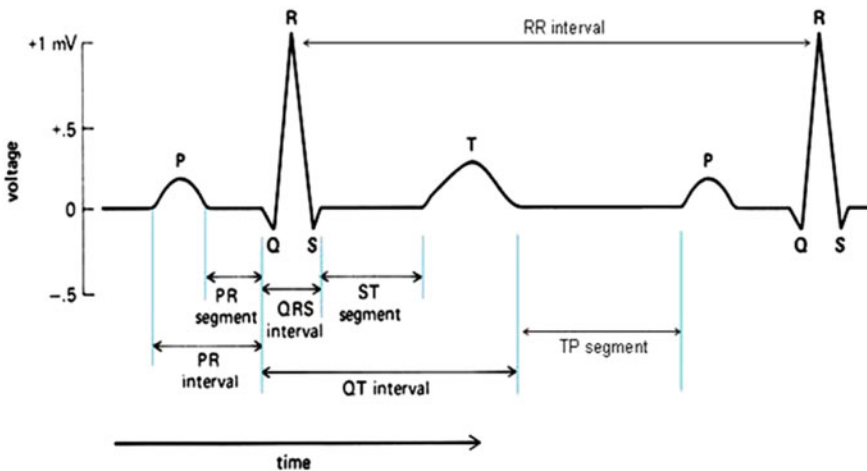
Fig. 2 Standard ECG signal

2 Adaptive Algorithm

2.1 The LMS Algorithm

The principle of this algorithm is established on modifying of the filter coefficient to diminish the mean square error between its desired response and its output; such system output is well suited for the estimation of the useful signal [5]. The equation for least mean square algorithm is given below.

$$w(n + 1) = w(n) + \delta.x(n).e(n) \tag{1}$$

In Eq. (1), e(n) is the error signal.
$w(n)$ is the vector of coefficient of adaptive filter.
δ is the step size.
The input vector is $x(n)$.

In many real-time systems, the LMS algorithm is preferable because of ease of implementation and its simplicity. This process is robust in the environment. In 1959, the LMS algorithm was proposed by Widrow and Hoff.

2.2 Normalized LMS Algorithm

If the convergence factor becomes high, the LMS algorithm has to face gradient noise amplification issues. To remove this problem, the NLMS algorithm is employed. The NLMS algorithm is also said to be a time-varying step-size algorithm as the step-size parameter is normalized. The convergence factor δ is given in the below equation.

$$\delta(n) = \frac{\beta}{\sigma + \|u(n)\|^2} \tag{2}$$

where $\|u(n)\|^2$ is defined as the squared Euclidean norm of the input $u(n)$.
β is defined as adaption constant. It also optimizes the convergence rate.
β lies $0 < \beta < 2$.
σ is known as normalization constant and valid for $\sigma < 1$.

The updated filter tap weight vector and step size are defined as in the following standard equation:

$$w(n + 1) = w(n) + \frac{\beta}{\sigma + ||u(n)||^2}e(n).u(n) \tag{3}$$

3 Adaptive Noise Cancelation

The adaptive noise cancelation systems are used for reducing the noise portion in order to obtain the desired signal without disturbing. The block diagram of ANC is given in Fig. 3. This system contains the first input as the primary signal and the second input as the reference signal. Desired signal corrupted with undesired noise is known as primary input. Reference signal is undesired noise which is filtered in the system. The motive of adaptive filtering system is to diminish the noise part and to determine the original signal which is not corrupted.

The primary input contains two portions, first is the desired signal and the other is the noise signal. The reference signal contains a solid reference of noise which is present in the primary signal.

The noise present in the reference signal is to be filtered to compensate for phase, time delay, and amplitude. Now noise in the reference signal will be subtracted from the primary signal [1, 2].

4 Results and Discussion

4.1 Simulink Model

The noise is designed as Gaussian noise which is added with the audio signal. The added signal is passed into an LMS filter for simulation. The Simulink model is shown in Fig. 4.
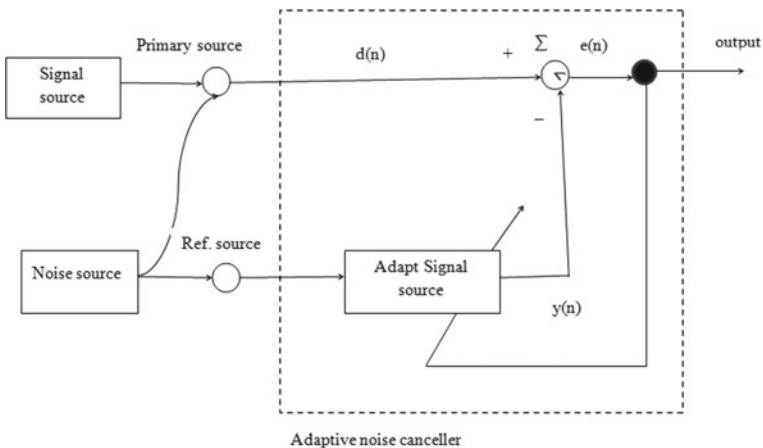
Fig. 3 Basic block diagram of adaptive noise cancelation system
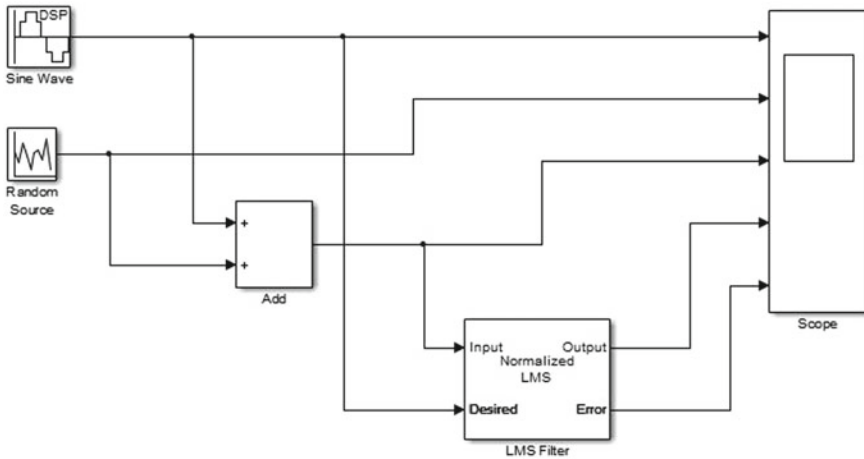
Fig. 4 Model for noise cancelation

The added signal is the input of the systems and the output of the system obtained from the model is sinusoidal signal after adaptive filtering.

When the step-size parameter is made at the high constant value, the speed of the filter is fast but it provides less accuracy. On the other hand, if the step size is made at a lower constant value, the speed of the filter may be slow but it provides more accurate performance.

The difference between the desired and input signal is known as an error signal which is determined by the output port of LMS filter. From this, port noise can be adaptively removed out.

Performance comparison of Adaptive Algorithms

S. No.	Algorithms	Complexity	Stability	Speed	MSE
1.	LMS	$2N+1$	Less stable	Medium	High
2.	NLMS	$3N+1$	stable	High	low
3.	RLS	$2N^2$	High stable	Very high	Very low

4.2   Adaptive Noise Cancelation Set for Fetal ECG

The maternal heartbeat signal creation:
The mother and fetal electrocardiogram shapes have been implemented. Sampling rate is fixed at 4 kHz. The peak voltage of the signal is kept at 3.5 mV and heart rate of the mother is also fixed at 89 beats per minute [7].

Fetal heartbeat signal creation:

Fetal heart rate is faster than the mother that is 120–160 beats/min. The amplitude of the mother heart is much stronger than fetal electrocardiogram. Here, we set peak voltage of fetal as 0.25 mV and heart rate of the fetal as 139 beats/minute.

The measured maternal electrocardiogram:

From the chest of the mother, the ECG signal is observed. The maternal heartbeat signal is eliminated from the fetal ECG signal with the help of a used technique. Some additive noise is contained by the fetal ECG signal.

The measured fetal electrocardiogram:

The fetal ECG signal which is observed from the mother abdomen is governed by the maternal heartbeat. The signal is passed from the chest cavity to abdomen.

The adaptive noise canceller:

To reduce complexity, the adaptive filter of step size 0.00006 has been used with 15 coefficients. To perform this work, adaptive noise canceller utilizes an adaptive process. The adaptive noise canceller is converged very well.

Recovered fetal heartbeat signal:

At the output, adaptive filter consists of the maternal heartbeat signal. The error signal e(n) of the system that contains the fetal heartbeat signal and noise. With the help of this error signal, the fetal heart rate can be determined [7–14].

4.3 Results

For Audio Signal:

Figure 5 shows the output of the model output. In this figure, the first waveform is defined as the input signal which enters the system. Noise is defined as the second waveform. The third waveform represents the addition of noise and input signal. The
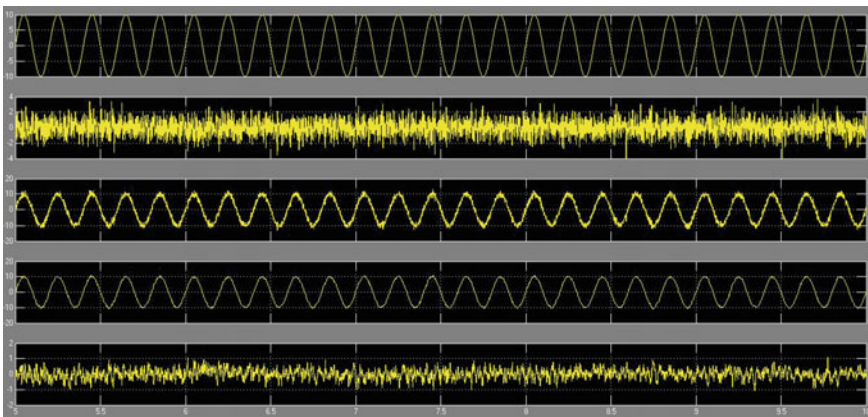
Fig. 5 Adaptive noise canceller Simulink model output

fourth waveform shows the desired signal and the last waveform shows the error signal which is defined as the subtraction of the desired signal and the input signal.
Implemented result of ECG signal:
The input of the filter is noisy ECG signal which is filtered by an adaptive algorithm. Extraction of fetal heartbeat from noisy ECG signal is shown below (Figs. 6, 7, 8, 9).

Fig. 6 Maternal heartbeat
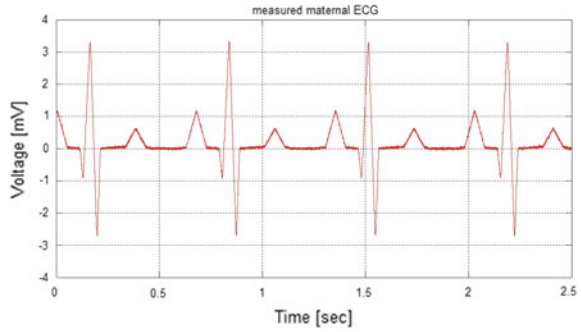
Fig. 7 Fetal heartbeat
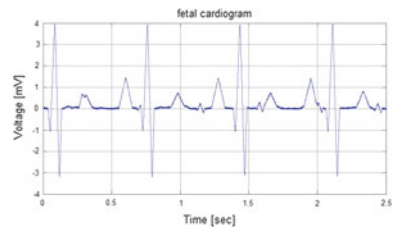
Fig. 8 Adaptive noise canceller
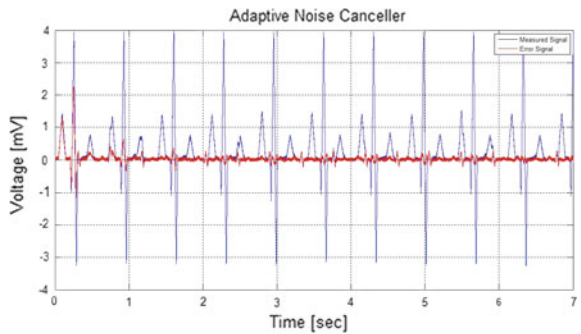
Fig. 9 Recovering the fetal heartbeat signal
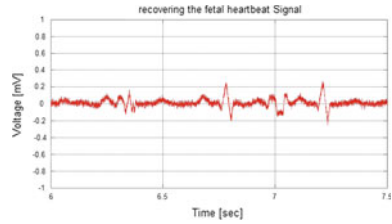
5 Conclusion

The adaptive noise cancelation method for cancelation of noise is discussed in this paper. This method has a basic advantage of adaptive capability and low signal distortion. Error performance using NLMS algorithm is found to be good. Fetal heart rate from ECG signal is an essential condition before and during birth of the baby. The developed model using this algorithm has been simulated using MATLAB environment and simulation results have been studied.

References

1. Sahu, K., Sinha, R.: Simulation of NLMS adaptive filter for noise cancellation. Int. J. Eng. Appl. Sci. (IJEAS), ISSN: 2394-3661
2. He, Y., He, H., Li, L., Wu, Y., Pan, H.: The applications and simulation of adaptive filter in noise canceling. In: 2008 International Conference on Computer Science and Software Engineering, vol. 4, pp. 1–4. IEEE (2008)
3. Ardalan, S., Moghadami, S., Jaafari, S.: Motion noise cancelation in heartbeat sensing using accelerometer and adaptive filter. IEEE Embed. Syst. Lett. **7**(4), 101–104 (2015)
4. Ramos, R., Mànuel-Làzaro, A., Del Río, J., Olivar, G.: FPGA-based implementation of an adaptive canceller for 50/60-Hz interference in electrocardiography. IEEE Trans. Instrum. Meas. **56**(6), 2633–2640 (2007)
5. Haykin, S.S.: Adaptive Filter Theory. Pearson Education India (2008)
6. Proakis, J.G.: Digital Signal Processing: Principles Algorithms and Applications. Pearson Education India (2001)
7. Sehamby, R., Singh, B.: Noise Cancellation using Adaptive Filtering in ECG Signals: Application to Biotelemetry. Int. J. Bio-Sci. Bio-Technol. **8**(2), 237–244 (2016)
8. Zarzoso, V., Nandi, A.K.: Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. IEEE Trans. Biomed. Eng. **48**(1), 12–18 (2001)
9. Khamene, A., Negahdaripour, S.: A new method for the extraction of fetal ECG from the composite abdominal signal. IEEE Trans. Biomed. Eng. **47**(4), 507–516 (2000)
10. Li, C., Zheng, C., Tai, C.: Detection of ECG characteristic points using wavelet transforms. IEEE Trans. Biomed. Eng. **42**(1), 21–28 (1995)
11. Kanjilal, P.P., Palit, S., Saha, G.: Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. IEEE Trans. Biomed. Eng. **44**(1), 51–59 (1997)

12. Moody, G.B., Mark, R.G.: QRS morphology representation and noise estimation using the Karhunen-Loeve transform. In: Computers in Cardiology 1989, Proceedings, pp. 269–272. IEEE (1989)
13. Barros, A.K., Mansour, A., Ohnishi, N.: Removing artifacts from electrocardiographic signals using independent components analysis. Neurocomputing **22**(1–3), 173–186 (1998)
14. He, T., Clifford, G., Tarassenko, L.: Application of independent component analysis in removing artefacts from the electrocardiogram. Neural Comput. Appl. **15**(2), 105–116 (2006)

Abrupt Scene Change Detection Using Spatiotemporal Regularity of Video Cube

Rupesh Kumar [ORCID]**, Sonali Ray, Meenakshi Sharma and Basant Kumar**

Abstract In this paper, we propose the detection method of abrupt scene change using spatial as well as spatiotemporal frames of video cube. Most of the methods use either intensity or motion of pixels for the scene change detection methodology. Unlike to the existing methods, both the intensity and flow vector of video frames are used simultaneously in this paper to propose a general abrupt scene change detection method. For a spatial frame, flow energy function is used for detection. Flow energy function, defined by the spatiotemporal regularity flow model, is the combinatorial form of intensity and flow vectors of the frames. In the spatio-temporal frames, abrupt scene change appears as a vertical line which is detected by the edge detection method. Combined results of spatial and the spatio-temporal frames provide the location of scene change. The proposed method detects almost all the locations of scene change with negligible false detection.

1 Introduction

To retrieve the semantic content in a video is a cumbersome task owing to the large size of video data. The retrieval problem in the video can be solved through indexing of video frames at the cost of high processing cost. Generally, scene change detection [2, 4, 6, 8–10] is a method of content detection for the application of annotation, scene analysis, fast searching, and indexing. The other use of scene change detection is in video compression where it is used to estimate the key frame. Manual indexing and annotation of large multimedia data are time-consuming tasks that encourage

R. Kumar (✉)
IIT, Kanpur, India
e-mail: rupeshkr@iitk.ac.in

S. Ray · M. Sharma · B. Kumar
MNNIT, Allahabad, India
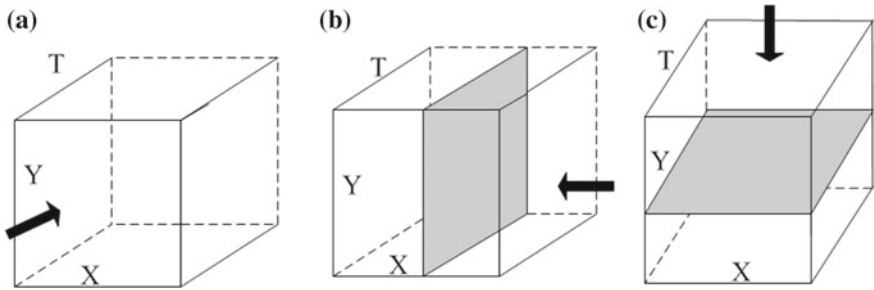
Fig. 1 Video cube. **a** $V_t XY$, **b** $V_x TY$, **c** $V_y TX$

the researcher to make an automatic scene change detection algorithm. Location of scene change in a video sequence appears as a boundary [13], because continuous scene appears as a flow of continuous action that depends on the foreground and background contents. Background and foreground of continuous scene possesses similar contents that make a shot [7, 12] or content of similarity. Scene change appears at the boundary of two shots or in between two scenes. Therefore, for the content retrieval purpose, the video is organized into a groups of shots. The aim of the scene change detection method is to partition the video sequences into the meaningful and manageable segments (shots) [10] for video indexing. The key frame is extracted from each segment of the scene that represents spatial and temporal features of that scene segment. A scene change [3] in a video stream can also be explained as a change of feature points or change of pixel intensity between two consecutive frames up to a remarkable limit. The term limit is interpreted as Thresholding [5, 14] which is widely used for the detection of scene change. Threshold may be fixed or dynamic in nature and the value of dynamic threshold [11, 18] is always updated according to the content of scene segment. Similarity measure between two consecutive frames is the basic idea of scene change detection technique and most of the prior work uses such methodology.

This paper is organized as follows. Section II explains the proposed scene change detection method. Experimental results are given in Section III, and Section IV concludes the paper. Notation of upper case capital X, Y, T is used for the respective axes and lower case x, y, t are for flow direction in the whole paper. $V_t XY$ represents video cube with XY frame along t direction (Fig. 1). Similarly, $V_x TY$ and $V_y TX$ represent video cubes with frames TY and TX along x- and y- directions.

2 Proposed Detection Method

Proposed scene change detection method uses spatial frame XY. The novelty of the proposed techniques is to consider the video as a cube and process the entire cube simultaneously. Unlike the other existing techniques, the proposed method does not

use frame-by-frame processing. The proposed detection method is a hybrid approach [16] which incorporates both the spatial frame and spatiotemporal frames (*TY* and *TX*). In the spatial frame, SPREF [1]-based frame energy is used to detect the abrupt scene change and the obtained result is fused with the method reported in [15] that uses both the spatiotemporal frames. In the next section, the proposed SPREF-based detection method has been explained.

2.1 SPREF-Based Detection Method

SPREF (spatiotemporal regularity flow) [1] is a general framework to model the video. Assumption of the video as a cube is one of the advantages of this model. SPREF (Spatiotemporal Regularity Flow) is a 3D vector field and it proposes a regular flow direction as a path in which the intensity of the pixel varies the least. If the scene is continuous, then intensity as well as flow vectors of frames vary regularly, but on the other hand, they show large deviation at the location of abrupt scene change. Using the SPREF model, we detect the deviation at the boundary of scene change with the help of flow energy function. In this paper, the translational-SPREF model is used and the flow energy is defined as

$$E\left(t\right) = \sum_{\Omega} |\left(I \star \frac{\partial H}{\partial x}\right)c_1'(t) + \left(I \star \frac{\partial H}{\partial y}\right)c_2'(t) + I \star \frac{\partial H}{\partial t}|^2 \qquad (1)$$

where $[c_1'(t), c_2'(t)]$ are the flow vector components in x- and y directions of the video frame *XY* with flow direction t. H is defined as a Gaussian filter and intensity of the image is I. Temporal size of the video cube is Ω and term c is used for translational flow. Flow energy function (Eq. 1) is solved by using translated box spline functions $b(u)$ of the first degree. Due to smoothness of spline, it approximates the regular flow direction by minimizing the flow energy function. Flow vectors in terms of spline coefficients are explained as

$$c_m'(u) = \sum_{n} \alpha_n^m b(2^{-l} u - n) \qquad (2)$$

where $m \in (1, 2)$ and $u \in (t)$. Term α_n is the nth spline coefficient. Length of temporal axis of video cube region Ω is $= 2^k$. Scaling factor of video cube is taken as l and its value has been taken as $l = 1, 2, \ldots, k$. Value of n is defined as $n = 2^{k-l}$. The spline function used here is defined as

$$b(z) = \begin{cases} 1 - |z| & \text{if } |z| < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Since flow energy function is the combinatorial form of intensity and the flow vectors of the frame, it estimates the regularity of frame efficiently. If no scene change occurs in a sequence of frames, then all the frames are regular on the basis of their frame energy. Flow energy of video defined by SPREF model combines both the features and therefore, it models the regularity of frame contents effectively than either of the pixel or flow vectors of the frame. Abrupt scene change in *XY* frames creates large deviation in their flow energy and the location of deviation is by the proposed threshold value:

$$Threshold = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (E_t - \mu)} \qquad (4)$$

$$where \quad \mu = \frac{1}{N} \sum_{t=1}^{N} E_t \qquad (5)$$

where E_t represents the flow energy of *t*th frame and N is the total number of frames. Abrupt scene change is detected with the help of the following condition:

$$E_{detected} = \begin{cases} 1 & \text{if } E_t >= 5 \times Threshold \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

$E_{detected}$ gives only the location of abrupt scene change with value 1. It has been investigated that all the peaks of abrupt scene change are greater than the mean value of the flow energy. How much the maximum peak value of flow energy is deviated from the mean value is defined by the standard deviation and this is the reason to select standard deviation for thresholding. The next section explains the detection approach by using spatiotemporal frames.

2.2 Boundary Detection in Spatiotemporal Frames

Apart from the use of flow energy for scene change, spatiotemporal frames are also considered for the detection of abrupt scene change which has been proposed in [15]. In this method, both spatiotemporal frames *TY* and *TX* are considered. Abrupt scene change produces large pixel intensity variation between two *XY* frames, but in spatiotemporal frames such variation appears as a vertical line or vertical boundary. The aim of this work is to detects the location of scene change as a vertical line and to combine the result with the results obtained from the previous section. Spatiotemporal frame-based abrupt scene change detection method [15] is summarized as

– Select four sampled *TY* and *TX* frames:

$$S_{TY} = [TY_{s_1}, TY_{s_2}, TY_{s_3}, TY_{s_4}]$$
$$S_{TX} = [TX_{s_1}, TX_{s_2}, TX_{s_3}, TX_{s_4}]$$

(7)

where s_1 is the first frame and sampled interval for other three frames is taken as

$$s' = \lfloor N/4.5 \rfloor$$

(8)

N is the total number of frames along the flow direction.

– Canny edge detection method is used to detect the edges of all the sampled frames and it produces binary images. Binary image of sampled frames are represented as

$$S_{TYedge} = [TY'_{s_1}, TY'_{s_2}, TY'_{s_3}, TY'_{s_4}]$$
$$S_{TXedge} = [TX'_{s_1}, TX'_{s_2}, TX'_{s_3}, TX'_{s_4}]$$

(9)

– In a binary image, the pixel value of the detected edges is assigned $'1'$ and only vertical lines are considered because they are part of the scene change location. As discussed earlier, abrupt scene change appears as a vertical line which is occupied by the column in both *TY* and *TX* frames.

– Spatiotemporal frames are considered as noisy images and hence, sometimes, the boundary of scene change might be distorted. Therefore, the length of the boundary is defined as a scene change location when

$$Length = 40\% of\,(frame \quad height).$$

(10)

where frame height is defined as the height of *TY* or *TX* frames.

– Condition: (*number of* $1's$ *in boundary line*) $>= Length.$
If the pixel value ($'1'$) of any boundary or vertical line follows the above condition in both the *TY* and *TX* frames, then it is interpreted as the location of abrupt scene change.

– The above procedure is repeated for all the sampled frames.

– Now look up Table 1 that has been generated. In this table, all the detected locations obtained from all the frames (*TY*, *TX*, *andXY*) are tabulated.

– Only those locations are considered that appeared at least twice among these spatial and spatiotemporal frames.

3 Experimental Results

Four natural test videos are taken from [17] for experiments and these are gstennis, anni002, anni003, and anni006. Video sequence gstennis (in Fig. 2) has 64 frames with one scene change and detection of scene change in spatiotemporal *TY* frame is

Table 1 Detected locations in spatial and spatiotemporal frames of anni006

Sampled frame	Detected Locations																						
TY'_{s_1}	–	77	–	215	277	349	413	530	577	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TY'_{s_2}	–	77	–	215	277	349	413	530	–	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TY'_{s_3}	–	77	–	215	–	349	413	530	577	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TY'_{s_4}	30	77	–	215	–	349	413	530	–	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TX'_{s_1}	–	77	–	215	277	349	413	530	–	581	702	–	865	936	–	1029	1142	1318	1554	–	1774	1890	1976
TX'_{s_2}	–	77	–	215	277	349	413	530	–	–	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TX'_{s_3}	–	77	–	215	277	–	413	530	–	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976
TX'_{s_4}	–	77	116	215	–	349	413	530	–	581	702	769	865	936	962	1029	1142	1318	1554	1656	1774	1890	1976
Frame XY	–	77			277	349	413	530	–	581	702	–	865	936	–	1029	1142	1318	1554	1656	1774	1890	1976

Fig. 2 Scenes of gstennis

(a) TY frame   (b) TY edge   (c) TY boundary

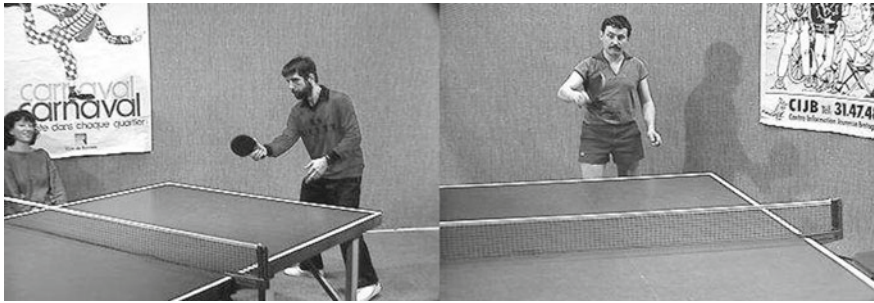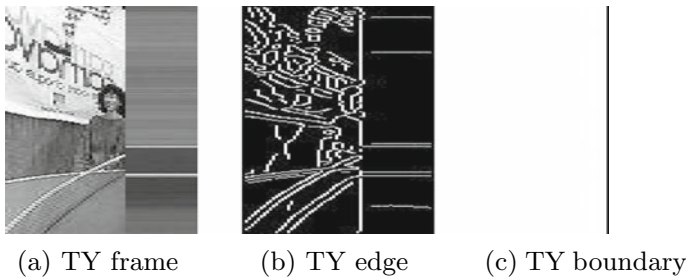**Fig. 3** Scene change detection in gstennis video

shown in Fig. 3. Since one scene change appears in gstennis video sequence, only one boundary or vertical line appeared in their binary image (Fig. 3b). Number of frames taken in videos anni002, anni006, and anni003 are 2048, 2048, and 1024, respectively, and all the scenes are represented in Fig. 4. The total number of abrupt scene changes in videos anni002, anni003, and anni006 are 12, 7, and 19, respectively. All the spatial and spatiotemporal frames have been processed and the obtained results by both the methods have been combined so as to obtain optimal results. Flow energy function of videos anni002 and anni003 are shown in Fig. 5 and 6, respectively. The vertical axis of the energy plot is the magnitude of flow energy and the horizontal axis represents the frame numbers. Flow energy of *XY* frames of video anni002 is shown in Fig. 5a and deviations in the plot show the location of scene change. Location of scene change obtained through threshold (6) is now represented by Fig. 5b and magnitude 1 is assigned to the detected location. Similarly, Fig. 6 represents the scene change detection of video anni003. Spatiotemporal frames *TY* and *TX* of video annoi006 have been shown in Figs. 7 and 8. In their binary images, vertical lines represent the location of scene change and all of them appear in the columns of the frame. Now these columns (in *TY* and *TX* frames) are directly converted into frame numbers (*XY*) with scene change locations. Flow energy of frames *XY* of video anni006 has been shown in Fig. 9a. Detected locations of abrupt scene change obtained from the proposed method are shown in Fig. 9b. Scene change location in video anni006

(a) Number of scenes in video anni002.

(b) Number of scenes in video anni003.

(c) Number of scenes in video anni006.
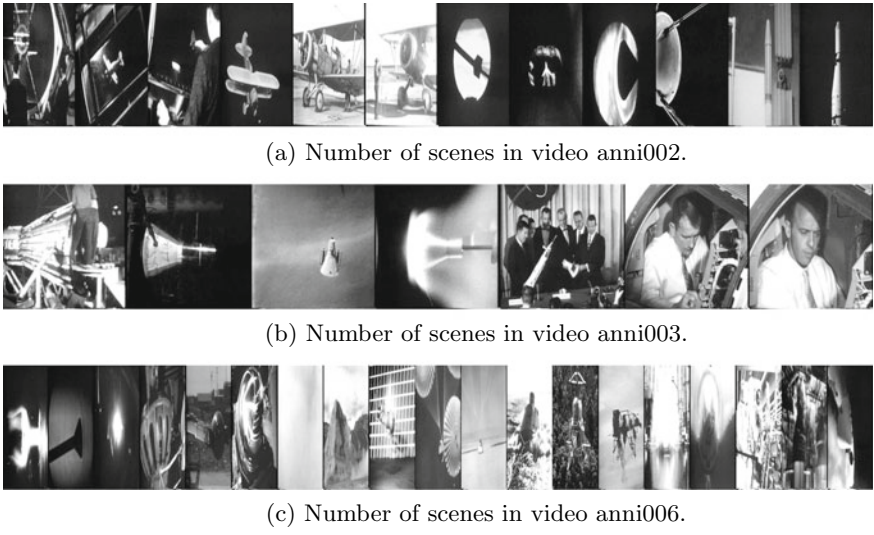
Fig. 4 Frames of different scenes

(a) Energy plot
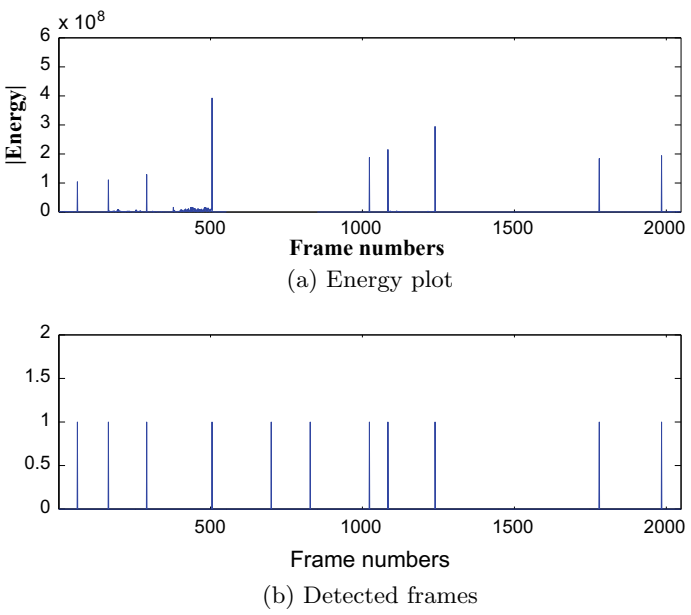
(b) Detected frames

Fig. 5 Detection of scene change in anni002

obtained by both the methods are combined together and tabulated in Table 1. There are three horizontal sections in Table 1 for the results obtained from frames TY, TX, and XY.
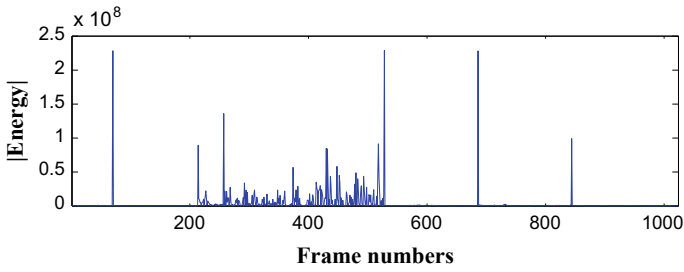
(a) Energy plot
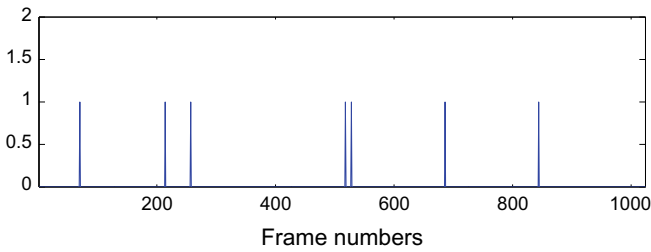
(b) Detected frames

Fig. 6 Detection of scene change in anni003

(a) TY frame
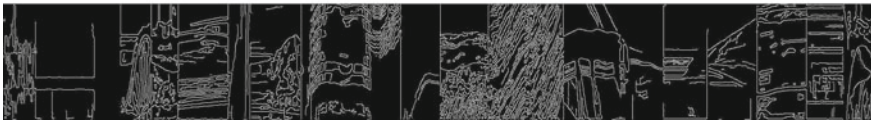
(b) TY'_{s_1} (binary) frame

Fig. 7 TY frame of video anni006

Among the three sections of spatial and spatiotemporal frames, location that appeared for at least two sections is considered as the approximated location of abrupt scene change. As shown in Table 1, locations obtained from the proposed method are [77, 215, 277, 349, 413, 530, 581, 702, 865, 936, 1029, 1142, 1318, 1554, 1774, 1890, 1976] and it is the same for the actual location. No missed or false detection are found and the proposed method detected all the abrupt scene changes. For the performance evaluation, the proposed method has been compared with the

(a) TX frame

(b) TX'_{s_1} (binary) frame
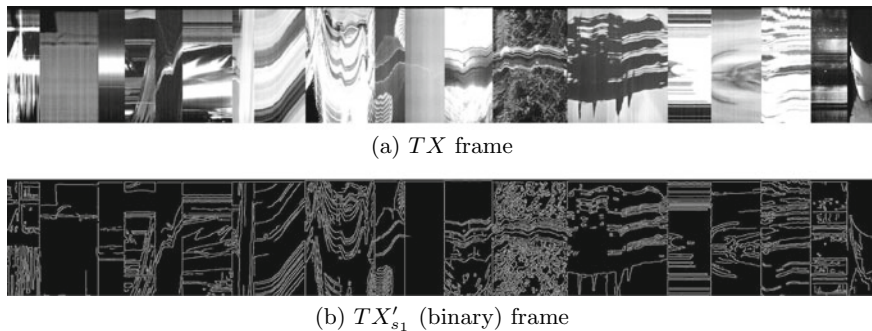
Fig. 8 *TX* frame of video anni006

(a) Energy plot

(b) Detected location
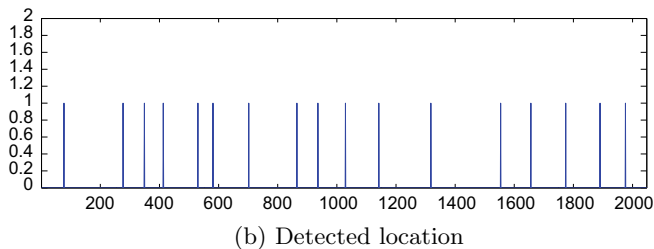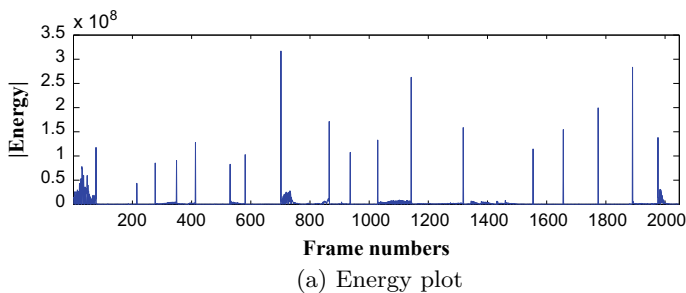
Fig. 9 Detection of scene change in anni006

method given in [2, 8] and the comparison result has been shown in Table 2. The accuracy of the proposed method has been evaluated with the help of precision, recall, and $F1$ score. $F1$ score is used for the evaluation of accuracy and it is defined as

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{11}$$

The proposed method detects all the scene changes and therefore the $F1$ score obtained is high as compared to the methods reported in [2, 8].

Table 2 Comparison

Test video	Methods	No. of scene change	F1
anni002	Method [8]	11	0.91
	Method [2]		0.83
	Proposed method		1
anni003	Method [8]	6	0.90
	Method [2]		0.86
	Proposed method		1
anni006	Method [8]	18	0.93
	Method [2]		0.82
	Proposed method		1

4   Conclusion

The proposed abrupt scene change detection method incorporates intensity as well as the flow energy of the frames. Using spatial and spatiotemporal frames reduces the false or missed detection. Edges of spatiotemporal frames and flow energy of spatial frames have been used for the detection. Detection accuracy of the proposed method is high with no false or missed detection as compared to the other methods.

References

1. Alatas, O., Yan, P., Shah, M.: Spatio–temporal regularity flow (SPREF): its estimation and applications. IEEE Trans. Circuits Syst. Video Technol. **17**(5), 584–589 (2007). https://doi.org/10.1109/TCSVT.2007.893832
2. Birinci, M., Kiranyaz, S.: A perceptual scheme for fully automatic video shot boundary detection. Signal Process.: Image Commun. **29**(3), 410–423 (2014)
3. Cyganek, B., Woźniak, M.: Tensor-based shot boundary detection in video streams. New Gener. Comput. **35**(4), 311–340 (2017)
4. Faernando, W., Canagarajah, C., Bull, D.: Scene change detection algorithms for content-based video indexing and retrieval. Electron. Commun. Eng. J. **13**(3), 117–126 (2001)
5. Hong, S., Cho, B., Choe, Y.: Adaptive thresholding for scene change detection. In: IEEE Third International Conference on Consumer Electronics, pp. 75–78. IEEE (2013)
6. Huang, C.L., Liao, B.Y.: A robust scene-change detection method for video segmentation. IEEE Trans. Circuits Syst. Video Technol. **11**(12), 1281–1288 (2001)
7. Jang, S.W., Byun, S.: Hough transform-based robust shot change detection in digital video images. Int. Inf. Inst. (Tokyo), Inf. **20**(2B), 1245 (2017)
8. Kang, S.J.: Adaptive luminance coding-based scene-change detection for frame rate up-conversion. IEEE Trans. Consum. Electron. **59**(2), 370–375 (2013)
9. Kang, S.J., Cho, S.I., Yoo, S., Kim, Y.H.: Scene change detection using multiple histograms for motion-compensated frame rate up-conversion. J. Disp. Technol. **8**(3), 121–126 (2012)
10. Koprinska, I., Carrato, S.: Temporal video segmentation: a survey. Signal Process.: Image Commun. **16**(5), 477–500 (2001)

11. Li, H., Liu, G., Zhang, Z., Li, Y.: Adaptive scene-detection algorithm for VBR video stream. IEEE Trans. Multimed. **6**(4), 624–633 (2004)
12. Majumdar, J., Aniketh, M., Abhishek, B., Hegde, N.: Video shot detection in transform domain. In: 2nd International Conference for Convergence in Technology (I2CT), pp. 161–168. IEEE (2017)
13. Prabavathy, A.K., Shree, J.D.: Histogram difference with fuzzy rule base modeling for gradual shot boundary detection in video cloud applications. Clust. Comput. 1–8 (2017)
14. Rosin, P.L., Ioannidis, E.: Evaluation of global image thresholding for change detection. Pattern Recognit. Lett. **24**(14), 2345–2356 (2003)
15. Rupesh, K., Gupta, S., Venkatesh, K.S.: Cut scene change detection using spatio temporal video frame. In: International Conference on Image Information Processing (ICIIP) (2015)
16. Shen, R.K., Lin, Y.N., Juang, T.T.Y., Shen, V.R., Lim, S.Y.: Automatic detection of video shot boundary in social media using a hybrid approach of HLFPN and keypoint matching. IEEE Trans. Comput. Soc. Syst. **5**(1), 210–219 (2018)
17. The open video project.: http://www.open-video.org/results.php?genre=Documentary
18. Youm, S., Kim, W.: Dynamic threshold method for scene change detection. In: International Conference on Multimedia and Expo, ICME'03, vol. 2, pp. II–337. IEEE (2003)

A Novel Approach for Compensation of Light Variation Effects with KELM Classification for Efficient Face Recognition

Virendra P. Vishwakarma and Sahil Dalal

Abstract A novel technique for compensating the effect of light variations is proposed here for robust person identification using human face images. The proposed technique is adaptive and efficient for face recognition under varying illuminations generated due to light incident from different angles. Illumination variation's effect is due to varying lighting conditions which are smoothly changing in nature. Therefore, illumination normalization is performed over some of the low-frequency discrete cosine transform (DCT) coefficients depending upon the illumination variations in the face image. A fuzzy modifier has been used to suppress the illumination variations on these low-frequency DCT coefficients. The number of low-frequency DCT coefficients is computed adaptively based upon the magnitude of these coefficients. The proposed approach utilizes KELM for the recognition of normalized face images and is tested over Extended YALE B face database. The experimental results clearly reveal that the proposed approach is significantly better than the existing approaches of illumination normalization for face recognition. With the proposed approach, the percentage error rate of 0%, 0.75%, and 1.11% on Subset 3, 4, and 5 of this database have been achieved, respectively.

Keywords DCT · Face recognition · Adaptive illumination normalization · Kernel extreme learning machine (KELM)

1 Introduction

Differentiating one person from another is generally done with the help of human faces. But illumination variations, on the face of the person, make the recognition

V. P. Vishwakarma (✉) · S. Dalal
University School of Information Communication & Technology, Guru Gobind Singh Indraprastha University, Sector 16-C, Dwarka 110078, New Delhi, India
e-mail: virendravishwa@rediffmail.com; vpv@ipu.ac.in

S. Dalal
e-mail: dalalsahil22@yahoo.co.in

difficult. Illumination can be from different directions and hence, it should be normalized so that face recognition can be done appropriately. Various algorithms have already been proposed for minimizing the effect of illumination on the face images. Illumination normalization can be done by using a representation for the image which can be insensitive to these variations. This was done by utilizing intensity derivatives by Y. Adini et al. in 1997 in which images were convolved with filters like Gabor filter, etc. [1]. Other algorithms which are presently in use to normalize the illuminated images are as follows: Simple Fuzzy filter with AHE + Log [2], Tetrolet Transform [3], S&L (discard-LFDCT Coeff.) [4], S&L (NPL-QI) [4]. Apart from these algorithms, M. Savvides and B. V. K. Kumar proposed an algorithm based on Logarithm transformations that was implemented on CMU PIE database for the recognition of face images [5]. In the year 2007, S.-I. Choi et al. introduced a unique algorithm to handle the illumination variation for face recognition. They utilized shadow characteristics, which occur due to illumination from different angles, to normalize the images and hence, face recognition was performed with comparable recognition rate [6]. H. Shim et al. proposed the algorithm which was tested on CMU PIE and Yale database. They utilized the concept of face relighting in which reflectance function (sin reflectance properties and facial hair), pose, and lighting were jointly estimated. This was helpful in creating face images under any illumination variations giving better results compared to other approaches [7]. In 2010, Adaptive region-based image preprocessing technique was proposed by S. Du and R. K. Ward for simplifying the illumination invariant face recognition. The proposed technique segmented the face image into different regions based on the varying lighting conditions so that enhancement can be done regionally to normalize the illumination variations. The technique was tested over Yale B + Extended Yale B database [8] and the CMU PIE databases with significant results [9]. In 2011, B. Wang introduced a ratio named as "Weberface" in which ration of local intensity variation and background was evaluated. This was based on Weber's law that stated that the ratio between the smallest perceptual change and background is constant for a stimulus. It was evaluated and tested over CMU PIE and Yale B face databases with very good results of face recognition [10]. Local binary pattern (LBP) was also utilized as a descriptor in illumination normalization by F. Juefei-Xu and M. Savvides. They had used the LBP concept in which it can remove illumination variation's effects. In the technique, illumination normalization was done in LBP domain and face image was retrieved back by reverse mapping from LBP to pixel domain. This approach was tested on Extended Yale B face database [11]. Recently, in the year 2017, Y.-F. Yu et al. proposed a discriminative multi-layer illumination-robust (DMI) feature extraction model to state illumination variation problem. As a linear combination, they decomposed the large-scale features into DMI features and a discriminant filter was utilized to improve the statistical discriminative ability and robustness of the reconstructed illumination-robust face image. This algorithm was tested over Extended Yale B, FRGC-v2, CAS-PEAL-R1, and Oulu-CASIA NIR-VIS databases with very good recognition rate [12]. No work has been carried over so far which can handle light variations adaptively.

In this paper, a novel approach has been introduced to solve this problem along with nonlinear classification, i.e., kernel extreme learning machine (KELM) for face

recognition of the images which are affected by the varying illumination. The remaining part of the paper is organized as follows: Sect. 2 gives a brief idea about the preliminaries used during this work. Section 3 provides a detailed explanation of the proposed method. Section 4 explains the database used with experimental results and comparison followed by conclusion in Sect. 5.

2 Preliminaries

2.1 AHELT

Contrast stretching is considered as a preprocessing approach for the face images under varying illumination. It is done by using a combination of adaptive histogram equalization (AHE) and logarithmic transform (LT) named as AHELT [13]. After processing using AHE, the number of low gray-level pixels increases in the resultant image. This effect is compensated by performing LT followed by AHE.

2.2 DCT

The discrete cosine transform (DCT) is a tool to transform the data from time to frequency domain and gives only the real part of discrete Fourier transform. For a $V \times W$ size face image, the 2-D DCT [14] can be defined as

$$D(k, l) = \phi(k)\phi(l) \sum_{v=0}^{V-1} \sum_{w=0}^{W-1} z(v, w) \times \cos\left[\frac{\pi(2v+1)k}{2V}\right] \times \cos\left[\frac{\pi(2w+1)l}{2W}\right]$$

(1)

for $k = 0, 1, 2, 3, \ldots, V - 1$ and $l = 0, 1, 2, 3, \ldots, W - 1$.

The inverse discrete cosine transform (IDCT), now, can be defined as

$$g(v, w) = \sum_{k=0}^{V-1} \sum_{l=0}^{W-1} \phi(k)\phi(l)D(k, l) \times \cos\left[\frac{\pi(2v+1)k}{2V}\right] \times \cos\left[\frac{\pi(2w+1)l}{2W}\right]$$

(2)

for $v = 0, 1, 2, 3, \ldots, V - 1$ and $w = 0, 1, 2, 3, \ldots, W - 1$.

where

$$\phi(k) = \begin{cases} \frac{1}{\sqrt{V}}; & k = 0 \\ \sqrt{\frac{2}{V}}; & k = 1, 2, 3, \ldots, V - 1 \end{cases}$$

(3)

$$\phi(l) = \begin{cases} \frac{1}{\sqrt{W}}; & l = 0 \\ \sqrt{\frac{2}{W}}; & l = 1, 2, 3, \ldots, W - 1 \end{cases} \tag{4}$$

v and w represent the rows and columns of the face image matrix, respectively, and k and l denote the rows and columns of the DCT matrix.

2.3 Fuzzy Filter

The concept of fuzzy is used to generate a model between the real-world applications and approximate reasoning of mathematics. Using this concept, recognition performance is increased by utilizing membership functions (MFs) to relate the real-time problems with each other. A function (say μ_ψ) can be expressed as follows:

$$\mu_\psi(x) = \begin{Bmatrix} 1; & z \in \psi \\ 0; & otherwise \end{Bmatrix} \tag{5}$$

The abovementioned equation is the case for classical sets in which element z either belongs or does not belong to the class Ψ. But in real-time problems, there is fuzziness present which does not completely separate the situations from one another. Such situations are solved by fuzzy MFs. It contains the values from 0 to 1. Universal set (y) of the real-world F then maps the values as

$$\mu_F : y \rightarrow [0, 1] \tag{6}$$

Here, the low-frequency DCT coefficients are the universal set [15].

2.4 KELM

Extreme learning machine (ELM) is an algorithm that is simply a single hidden-layer feed-forward neural network (FFNN). Unlike NN, ELM randomly assigns its hidden layer weights and need not be learned iteratively. The input weights and bias selected, for the hidden layer, are randomly initialized. The output ρ for this hidden layer matrix is computed as follows:

$$\Theta = \sum_{s=1}^{S} \lambda_s \rho(A_s, B_s, x) = \lambda \cdot \Omega(x) \tag{7}$$

where $\Omega(x) = [\rho(A_1, B_1, x) \ldots, \rho(A_S, B_S, x)]$ is the output matrix obtained from the hidden layer with respect to the input x. A_S and B_S are the input weights and

biases, respectively. The output weight λ which connects the output to hidden nodes is obtained analytically as

$$\lambda = \Omega^{-1}\Theta = \Omega^{\Gamma}\left(\frac{I}{\varsigma} + \Omega\Omega^{\Gamma}\right)^{-1}\Gamma \tag{8}$$

where Γ is the target class and ς is the user-defined parameter for regularization. Therefore, the ELM model can be formulated as

$$\Theta_{ELM}(x) = \Omega(x)\Omega^{\Gamma}\left(\frac{I}{\varsigma} + \Omega\Omega^{\Gamma}\right)^{-1}\Gamma \tag{9}$$

ELM is to minimize the training errors and also allows to reduce the computational time. With these advantages, ELM also has some disadvantage as it has local minima issue and issue of easy over-fitting. So, to overcome the limitations of ELM, another algorithm is proposed in terms of the kernel matrix. The kernel matrix can be obtained using (10):

$$\zeta = \sum_{\varsigma=1}^{S} \Omega(A_\varsigma, B_\varsigma, x_s) \cdot \Omega(A_\varsigma, B_\varsigma, x_s) \tag{10}$$

where x_s denotes the training data and S represents the number of training data used. Therefore, the formulation for KELM can be done as

$$\Theta_{KELM}(x) = \Omega(x)\Omega^{\Gamma}\left(\frac{I}{\varsigma} + \zeta\right)^{-1}\Gamma \tag{11}$$

This approach is named as kernel extreme learning machine (KELM) and can be used as a binary as well as for multi-class classification [16].

3 Proposed Approach

The proposed approach gives an efficient method of illumination normalization by modifying the low-frequency DCT coefficients with the help of a fuzzy filter. The database used to evaluate the proposed approach is briefly given in the next section. The database contains images with illumination on the faces from different directions. From the database, images with frontal (uniform) or with very small illumination variation are considered for training set and the remaining images are used as test set.

All these face images are first preprocessed using AHELT. Results obtained after AHE and AHELT are also shown in Fig. 1 (2nd and 3rd column). As illumination varies slowly and mainly affects the low-frequency DCT coefficients, DCT of the

Database	ORIGINAL IMAGE	AHE	AHELT	ILLU. NORM. [2]	ILLU. NORM. (Adaptive)
Extended YALE B Subset 1				CC=0.5198	CC=0.5404, N=946
Extended YALE B Subset 4				CC=0.2379	CC=0.2521,N=842
Extended YALE B Subset 5				CC=0.1767	CC=0.1985, N=742
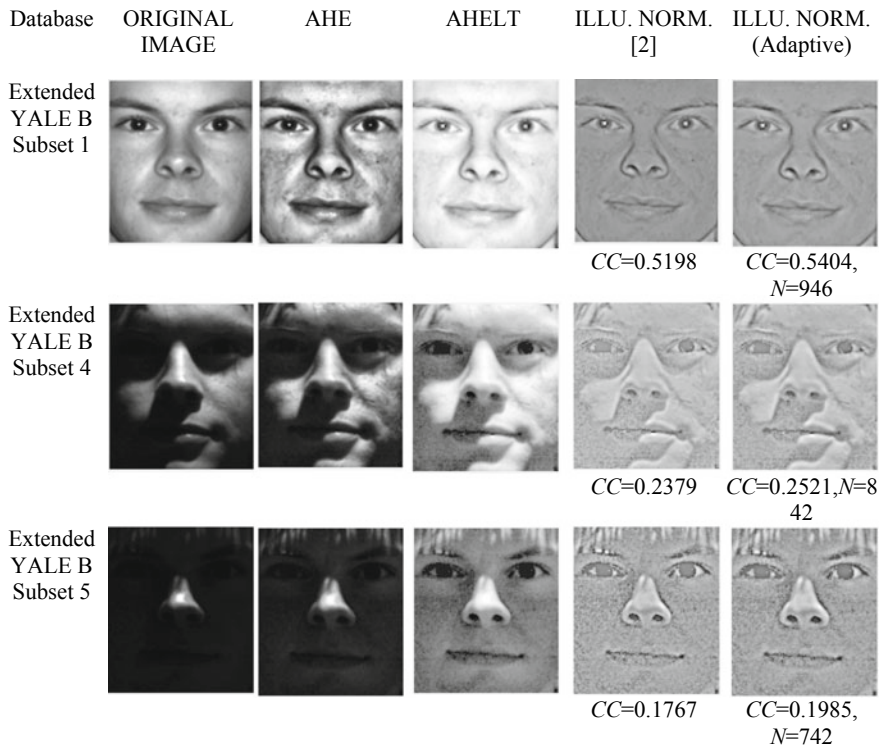
Fig. 1 Effect of illumination normalization on training and testing images for Extended YALE B database

face images are taken. All of these coefficients are arranged in an increasing order of frequency by selecting them in a zigzag manner [2]. This vector form *Vec* can be represented in the form of an equation as

$$Vec = ZigzagScan(D) \tag{12}$$

where *D* is the DCT matrix of $V \times W$ size. *ZigzagScan* is telling about the selection process of coefficients from the DCT matrix *D*. Now, in *Vec*, only those coefficients are considered for further operation which are highly affected by illumination variations and these DCT coefficients are those which have a magnitude greater than some constant (say ω). Let *NDctCoef* be the number of these DCT coefficients in *Vec*. Then *NDctCoef* can be obtained as

$$NDctCoef = count(Magnitue(Vec) > \omega) \tag{13}$$

This count is not giving the number of low-frequency DCT coefficients which are slightly affected by illumination variations. So, to obtain that exact number of

low-frequency DCT coefficients affected by illumination variations, *NDctCoef* is multiplied by η as

$$N = \eta \times NDctCoef \tag{14}$$

Then these low-frequency DCT coefficients (N) are modified using the fuzzy filter. For this modification, fuzzy membership function μ_F which can be selected is

$$\mu_F = \frac{g^\xi}{G^\xi} \tag{15}$$

where G is the index of the last low-frequency DCT coefficient considered for fuzzy filtering and ξ is a positive constant. As g varies from 1 to G, a vector is obtained which contains fuzzy membership grades as their elements. It is then used as a fuzzy filter for the illumination normalization in the proposed approach. This vector is represented as M and can be expressed as

$$M = [M_1, M_2, M_3, \ldots, M_n, \ldots, M_N]^T \tag{16}$$

where M_n represents the membership grade for the nth low-frequency DCT coefficient.

The DCT coefficients after illumination normalization are obtained as follows:

$$FzyMdfd = invZigZagScan(Vec(0\ldots N-1) \times M + Vec(N\ldots VW-1)) \tag{17}$$

invZigzagScan is representing the reverse back process of *ZigzagScan* and results in a modified DCT vector to a matrix *FzyMdfd*. Now, by applying IDCT on the *FzyMdfd* (un-zigzag matrix form), illumination normalized face image is obtained. The training and test images before and after normalization are shown in Fig. 1. This can be seen using values of similarity between the training image (first image of the first column in Fig. 1) and the illumination normalized images based on correlation coefficient (*CC*) where the results of illumination normalization are much better compared to [2] which are also shown in the fourth column of Fig. 1. After applying the proposed approach of illumination normalization, classification is performed using KELM. For the classification process, regularization coefficient, kernel parameter, and kernel type are some parameters in KELM which need to be optimally selected as there is no appropriate method existing in the literature for finding them so that minimum percentage error rate can be achieved.

4   Experimental Results and Analysis

4.1   Database

Extended YALE B database is a large database which has 38 subjects with each having 5 subsets each. These subsets contain different number of images and different sets of illumination variations [8]. Subset 1 consists of 70 face images (7 of each subject) with 0°–12° of illumination variations. Similarly, Subset 2, Subset 3, Subset 4, and Subset 5 consist of 120, 120, 140, and 190 face images with 13°–25°, 26°–50°, 51°–77° and above 77° of illumination variations, respectively. The original size of each image is 640×480 pixels which reduced to 120×105 pixels by cropping of hairs and background with further reduction by 1.6 in our experiments. Subset 1 images are used for training and the remaining are used for testing.

4.2   Result

To test the performance of the proposed approach on a large database, it is applied to the Extended YALE B database. By varying the values of ω, η, and ξ, optimized values are obtained which can give the minimum value of percentage error rate. This is evident from the results that $\psi = 0.9$, $\eta = 2$, and $\xi = 0.5$ are the best values which can give promising results on Extended YALE B database using KELM.

Regularization coefficient is utilized in KELM to remove the problems of overfitting [17] or to solve the ill-posed problems. It is selected optimally as 1 for the proposed approach. Kernel parameter of KELM is related to the nonlinear mapping from lower dimensional to higher dimensional feature space and its value is selected as 550 (Subset 4) and 1100 (Subset 5) optimally. The kernel type, which includes linear, polynomial, wavelet, and RBF, RBF kernel is selected for the proposed approach. Minimum values of error obtained using KELM for Subset 3, 4, and 5 are shown in Table 1.

Table 1 Percentage error rate on Extended YALE B database and comparison with existing approaches

Approach	Subset 3	Subset 4	Subset 5
Proposed approach	**0**	**0.75**	**1.11**
Simple Fuzzy filter with AHE + Log [2]	0	1.13	1.94
Tetrolet transform [3]	5.04	9.96	16.07
S&L (discard-LFDCT Coeff.) [4]	14.0	14.7	15.2
S&L (NPL-QI) [4]	3.3	13.0	30.9

5 Conclusion

The drawbacks of the existing techniques are overcome by the proposed approach in this paper. The proposed approach computes the number of low-frequency DCT coefficients for each illuminated face image based on the level of illumination present in the images. After illumination normalization, use of KELM makes the face recognition even more accurate compared to the existing approaches. The time complexity of the proposed approach is of the same order as that of the existing approaches because of the use of DCT, IDCT, and some vector operations only as needed in the methods which operate in DCT domain. For example, Discard LFDCT Coeff. [18], Rescaling of DCT coefficients [13], and Simple Fuzzy filter with AHE + Log [2]. In future, the proposed approach will be explored over other databases as well to make it more general and adaptive for face recognition.

References

1. Adini, Y., Moses, Y., Ullman, S.: Face recognition: the problem of compensating for changes in illumination direction. IEEE Trans. Pattern Anal. Mach. Intell. **19**, 721–732 (1997)
2. Vishwakarma, V.P.: Illumination normalization using fuzzy filter in DCT domain for face recognition. Int. J. Mach. Learn. Cybern. **6**, 17–34 (2015)
3. Hui-xian, Y., Yong-yong, C.: Adaptively weighted orthogonal gradient binary pattern for single sample face recognition under varying illumination. IET Biom. **5**, 76–82 (2016)
4. Xie, X., Zheng, W.-S., Lai, J., et al.: Normalization of face illumination based on large-and small-scale features. IEEE Trans. Image Process. **20**, 1807–1821 (2011)
5. Savvides, M., Kumar, B.V.K.: Illumination normalization using logarithm transforms for face authentication. In: Audio-and Video-Based Biometric Person Authentication, p. 1055 (2003)
6. Choi, S.-I., Kim, C., Choi, C.-H.: Shadow compensation in 2D images for face recognition. Pattern Recognit. **40**, 2118–2125 (2007)
7. Shim, H., Luo, J., Chen, T.: A subspace model-based approach to face relighting under unknown lighting and poses. IEEE Trans. Image Process. **17**, 1331–1341 (2008)
8. Extended Yale B face database. http://vision.ucsd.edu/~leekc/ExtYaleDatabase/Ext
9. Du, S., Ward, R.K.: Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions. IEEE Trans. Circuits Syst. Video Technol. **20**, 1165–1175 (2010)
10. Wang, B., Li, W., Yang, W., Liao, Q.: Illumination normalization based on Weber's law with application to face recognition. IEEE Signal Process. Lett. **18**, 462–465 (2011)
11. Juefei-Xu, F., Savvides, M.: Encoding and decoding local binary patterns for harsh face illumination normalization. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 3220–3224 (2015)
12. Yu, Y.-F., Dai, D.-Q., Ren, C.-X., Huang, K.-K.: Discriminative multi-layer illumination-robust feature extraction for face recognition. Pattern Recognit. **67**, 201–212 (2017)
13. Vishwakarma, V.P., Pandey, S., Gupta, M.N.: Adaptive histogram equalization and logarithm transform with rescaled low frequency DCT coefficients for illumination normalization. Int. J. Recent. Trends Eng. Technol. **1**, 318–322 (2009)
14. Gonzalez, R.C., Woods, R.E., Masters, B.R.: Digital image processing, third edition. J. Biomed. Opt. 14, 29901 (2008). https://doi.org/10.1117/1.3115362
15. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)

16. Wong, C.M., Vong, C.M., Wong, P.K., Cao, J.: Kernel-based multilayer extreme learning machines for representation learning. IEEE Trans. Neural Netw. Learn. Syst. (2016)
17. Vishwakarma, V.P.: Deterministic learning machine for face recognition with multi-model feature extraction. In: 2016 Ninth International Conference on Contemporary Computing (IC3), pp. 1–6 (2016)
18. Chen, W., Er, M.J., Wu, S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. IEEE Trans. Syst. Man, Cybern. Part B **36**, 458–466 (2006)

A Hybridization of Fuzzy Logic and Deterministic Learning Machine for Face Recognition

Virendra P. Vishwakarma and Sudesh Yadav

Abstract In this chapter, a new method for face recognition (FR) is proposed by integrating fuzzy logic in deterministic learning machine called fuzzy deterministic learning machine (FDLM). The main steps of the proposed approach are the fuzzification and classification step. The fuzzification step is done using π-membership function (MF) to map the grades of association of each input feature to each subject in a fuzzy matrix form and classification is done with the help of fast learning, parameter-free deterministic learning machine. To show the efficacy and superiority of the proposed approach, we have conducted a comparison analysis of the proposed approach with other methods available in the literature on Georgia Tech face database. Experimental results obtained reveal that the proposed approach shows a significant improvement in recognition performance for FR.

Keywords Face recognition · Fuzzy logic · Single-hidden layer feedforward neural networks · Classification algorithms

1 Introduction

Face recognition is a growing research area in the field of machine learning because of its use and choice of application in various fields: commercial, law enforcement, personnel benefits, etc. In early years, FR systems are based on visible spectrum. These systems suffer a lot of challenges of non-linear non-convex variations, viz. expression changes, pose variation, and different illumination intensities used during capturing the database [1–5]. These challenges are divided mainly into two categories: intrinsic and extrinsic factors. Intrinsic factors are physical appearance of human face (facial expression, ageing) and extrinsic factors include those factors due

V. P. Vishwakarma (✉) · S. Yadav
University School of Information Communication & Technology, Guru Gobind Singh
Indraprastha University, Sector 16-C, Dwarka 110078, New Delhi, India
e-mail: virendravishwa@rediffmail.com; vpv@ipu.ac.in

S. Yadav
e-mail: yadavsudesh01@gmail.com

to which intrinsic factors change, such as pose and lightening conditions [1, 6]. To diminish the effect of all these challenges, infrared spectrum-based FR systems may be developed. So, it leads and encourages many researchers for continuous research in this field of FR. For solving the above-mentioned challenges, first, a general strategy is made to understand the issue, how it formed and how it can be resolved [7]. Any FR system mainly comprises three steps: preprocessing, feature selection and classification [8–10]. From the last few years, numerous techniques have been built and studied to improve the recognition accuracy of FR. In these techniques, computers have been used as a primary tool for detecting recognizing human face images like the human perceptron system. These techniques mainly focused on identifying important traits and features of an individual using various feature extraction techniques. In face recognition, feature extraction is the most important and primary step. A good feature extraction technique reduces the intra-class dissimilarities and increases the difference between inter-class. In the literature, a lot of research has been done on facial feature extraction techniques (holistic methods as well as local feature-based methods). In holistic feature extraction methods, a complete face image is used as an input and classification is done similar to principal component analysis (PCA) [11, 12]. Whereas in local feature-based techniques, local features are extracted and used in recognition tasks [13]. In the literature, various face recognition techniques are available, viz. Eigenfaces-based [11], Gabor filter [14], 2-D Gabor Filter, Extended Bunch Graph, Local Binary Pattern, Hidden Markov Model, Support Vector Machine, etc. [7].

In today's era of development, fuzzy logic and artificial neural network (a part of soft computing) have been emerged as an important tool in improving the performance of FR systems. A number of based classification algorithms based on ANN are available in the literature to solve the non-linearity in FR applications [15]. One such network is the single-layer feedforward neural network (SLFN) [16]. The most commonly used training algorithm to train these networks are iterative (gradient decent based) and non-iterative (extreme learning machine) in nature, having a lot of issues like out of memory, slow learning, under-fitting, over–fitting, etc. To overcome all these problems of training SLFN, authors [17] proposed a new algorithm which is fast, deterministic and parameter-free as weight and biases are derived from input space to solve the high-dimensional problem of FR. As studies states that in addition to precise cognition human brain is also does analysis based on imprecise reasoning. Fuzzy logic is a branch of soft computing which deals with uncertain or imprecise information [18]. We have seen uncertainty everywhere in real life. Face recognition applications also suffer the problem of uncertainty. A lot of work has been done using fuzzy logic in the field of FR [8, 19, 20]. Researchers [21] modified the Sugeno integral with the help of interval type-2 fuzzy logic and used in modular neural networks for FR. The same was tested and validated on a large- scale database (FERET) to show the efficacy and superiority of the approach. Researcher [15] proposed a new approach to feature extraction using fuzzy k-nearest neighbour and integrating obtained fuzzy membership values in intra- and inter-class scatter matrices. They also include Gaussian probabilistic distribution information in intra-class fuzzy scatter matrices. Researchers [8] proposed a new approach to feature

extraction using interval type-2 fuzzy logic for FR, in which they have used a new interval type-2π based fuzzy membership function for the generation of fuzzy features and dimension reduction, and classification of high dimension is done with the help of PCA and *k*-nearest neighbour classifier.

In this paper, we propose a hybrid approach integrating the merits of fuzzy logics and deterministic learning machine for improving the recognition accuracy of FR systems. Till now, there is no approach available for advancement in DLM—a fast learning algorithm for training SLFN using fuzzy logic. Therefore, to integrate the pluton of fuzzy logic in DLM, we use π-membership function (π-mf) for obtaining fuzzy feature vectors corresponding to each sample image. After fuzzification, DLM is used as a classifier for high-dimensional input vectors extracted in the feature extraction phase.

The organization of the rest of the paper is as follows: Sect. 2 presents the proposed approach, Sect. 3 discusses the experimental study of the proposed approach and then finally, we conclude our paper with some future directions.

2 Proposed Approach

As we know, the performance of classification algorithms for face recognition mainly depends on two factors: one is the better feature extraction technique and the second one is the better classification engine. The main problem associated with FR systems is associated with the input signals; which are highly non-linear and non-convex due to data prepared in a highly uncontrolled environment such as different alignments of faces, illumination conditions, pose variation, expression, etc. To suppress the effect of all these and removing the non-linearity occurring in the features and for better classification, we propose a technique which integrates these two in a single approach for FR.

2.1 Feature Extraction Phase

In the present study, features are extracted using fuzzy logic to find the pixelwise association of individual pixel of a face image to different classes. This operation uses π membership function (MF) to obtain grade of participation of a selected pixel to all classes. It takes a face image as an input signal and applies π-MF for the generation of grades of participation of individual pixels to different classes. A face image in appearance-based FR systems is two-dimensional in nature. Therefore, it can be represented in a two-dimensional matrix representation form as below.

$$p = \begin{bmatrix} p_{1,1} & \cdots & p_{1,s} \\ \vdots & \vdots & \vdots \\ p_{r,1} & \cdots & p_{r,s} \end{bmatrix} \tag{1}$$

where r is the number of rows and s is the number of columns in the matrix. For our experiment, face images are converted in a column vector form such as

$$p = \begin{bmatrix} p_{1,1}, p_{2,1}, \ldots, p_{r,1}, p_{1,2}, \ldots, p_{1,s}, p_{2,s}, \ldots, p_{r,s} \end{bmatrix}^T \tag{2}$$

where $r \times s = Q$ denotes the total number of pixels in a face image.

On these column vectors, we apply fuzzification process and obtain one fuzzy vector corresponding to each face image. Our fuzzification phase uses a π-type membership function, which consists of a parameter b which tunes the function as per the requirement of the problem. By changing the value of the parameter, we can control the steepness of the MF. Shape of this MF is very much similar to Gaussian MF. The MF is as below.

$$f(p; u, v, w) = \begin{cases} 0, & p \leq u \\ 2^{b-1} \left(\frac{p-u}{v-u} \right)^b, & u < p \leq d_1 \\ 1 - 2^{b-1} \left(\frac{v-p}{v-u} \right)^b, & d_1 < p \leq v \\ 1 - 2^{b-1} \left(\frac{w-p}{w-v} \right)^b, & d_2 < p \leq w \\ 2^{b-1} \left(\frac{p-v}{w-v} \right)^b, & v < p \leq d_2 \\ 0, & p \geq v \end{cases} \tag{3}$$

where d_1 and d_2 are the two crossover points and value of $b = 2$. u, v and w are the minimum, maximum and mean value of a particular pixel. The value of MF can be measured for a particular pixel in data pattern (t) using

$$u = \min(t) \tag{4}$$

$$v = \max(t) \tag{5}$$
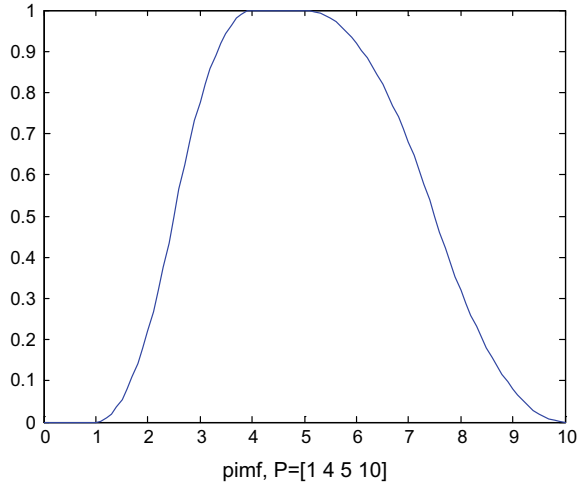
$$w = mean(t) \tag{6}$$

The grades of membership after applying the fuzzification process are as follows corresponding to a face image vector:

$$m = \begin{bmatrix} m_1, \ldots, m_n, \ldots, m_Q \end{bmatrix} \tag{7}$$

where m_n is the value of membership of nth pixel of a face image p (Fig. 1).

Fig. 1 Sample image of π-MF

pimf, P=[1 4 5 10]

2.2 Classification

Classification is a very important step of any pattern recognition algorithms. Choosing a good classification engine, not only improves the performance of the FR systems but also has applications in real-life fields such as finding fraudulent activities, improving security in various public areas like army, banks, hospitals, academic institutions, etc. In this paper, we use deterministic learning machine (DLM) as a classification engine for extracted fuzzy features. The reason behind choosing DLM is its learning speed, parameter independency and free from local minima, improper learning rate and under-fitting and over-fitting problems. In DLM, input weights and biases are obtained experimentally using input training patterns/data. This particular feature of training single-layer feedforward network (SLFN) makes it computationally efficient for classifying the above-extracted fuzzy features.

Assuming an SLFN for training a set of input specimens, $\{m_j, g_j\}_{j=1}^{Q}$ with Q number of fuzzy input specimens having $L(L = 1, 2, \ldots, k)$ number of classes, fw is the fuzzy input weight, r is the connection weight between hidden layer neurons and output neurons, e is the activation function and g is the output. The output for the same is given below.

$$\sum_{i=1}^{Q} r_i e_i(m)_j = \sum_{i=1}^{Q} r_i e(f w_i^T \cdot m_j + z_j) = g_j, j = 1, 2, \ldots, Q \tag{8}$$

where fw is fuzzy input vector which is derived experimentally from input matrix in DLM as follows:

$$f w_i = \frac{m_i}{cNR_i} \tag{9}$$

where NR_i is the norm of input fuzzy feature vectors, i.e. m_j and c is constant calculated by finding the norm of all NR_i.

And z_j is the bias of the jth hidden neuron which is computed by finding the mean value of its input weights. The same is calculated as follows:

$$z_j = \frac{1}{Q} \sum_{i=1}^{Q} f w_{j,i} \tag{10}$$

In this way, there are Q equations given by Eq. (8). These Q equations can be represented in the matrix form as follows:

$$FR = T \tag{11}$$

where

$$F = \begin{bmatrix} a\left(f w_1^T \cdot x_1 + b_1\right) & \dots & a\left(f w_Q^T \cdot x_1 + b_Q\right) \\ \vdots & \vdots & \vdots \\ a\left(f w_1^T \cdot x_Q + b_1\right) & \dots & a\left(f w_Q^T \cdot x_Q + b_m\right) \end{bmatrix} \tag{12}$$

$$R = \begin{bmatrix} r_1^T \\ \vdots \\ r_Q^T \end{bmatrix}, T = \begin{bmatrix} g_1^T \\ \vdots \\ g_Q^T \end{bmatrix} \tag{13}$$

Solution of Eq. (11) gives the value of output weight (R) and calculated as follows:

$$R = F^{-1} T \tag{14}$$

3   Experimental Results

In this section, we discuss the results obtained after applying the proposed technique. We have investigated our technique on standard face rec dataset Georgia tech face dataset. This dataset consists of 15 colour sample images of 50 different subjects taken at Georgia Tech Institute at Georgia. The images were taken in an uncontrolled environment with different facial expressions, pose variations, illumination variation, frontal and upward directions, etc. Some of the face images of Georgia tech face dataset are shown in Fig. 2.

Fig. 2 Example face images of Georgia tech face dataset

For our experiment purpose, we first convert each colour face image to grayscale image and resize to dimension of 40 × 30. Next, we convert each face image to a column vector and on these vectors, we apply fuzzification operation. In this way, we obtain fuzzy feature vectors corresponding to each image. After feature extraction, classification is done through non-iterative, fast deterministic learning machine, which is parameter-dependant and uses radial basis function as an activation function of network model.

We have tested our experiment for 4–10 no of images as training specimens and rest of the images from the same subject are taken as test specimens. A comparison with principal component analysis (PCA), fuzzy-based pixelwise information extraction (FPIE), quaternion principal component analysis (QPCA), interval type-2 fuzzy-based pixelwise information extraction (IT2FPIE), fuzzy quaternion- based pixelwise information extraction (FQPIE) and DLM is done. The results obtained and comparison study done show that the proposed approach is better in terms of speed, accuracy, and solving non-linear problems associated with classification algorithms of FR. Experimental results obtained and comparison to other methods is shown in Table 1.

Table 1 Comparison of percentage recognition error rate on Georgia tech face database

Method	No. of sample images taken for training per subject					
	4	6	7	8	9	10
Proposed approach	**40.36**	**27.55**	**21.75**	**21.14**	**17.33**	**18.00**
PCA [11]	50.00	40.89	36.50	33.43	31.33	29.20
FPIE [22]	47.00	42.60	29.55	25.25	24.57	20.33
Quaternion PCA [23]	47.27	36.67	34.00	30.86	26.67	26.00
IT2FPIE [9]	48.36	33.56	27.50	24.57	21.66	22.00
DLM [17]	41.27	29.78	25.75	22.86	20.33	18.00
NIntTy2FPIE [8]	44.00	29.33	24.75	23.14	20.67	20.40

4 Conclusion

In this paper, we propose a hybrid approach called fuzzy deterministic learning machine for solving various issues associated with single-layer feedforward network for FR systems, viz. non-linearity, speed and interdependency of different parameters. The key point of the proposed approach is that it incorporates the benefit of fuzzy logic in collaboration with fast learning algorithm DLM. It provides a significant improvement to DLM; a learning algorithm for classification problem of FR. Experimental results prove the significance of fuzzy logics in DLM. In future, more advancement in classification algorithms can be done by using other fuzzy logics like type-2, interval type-2 and Sugeno integrals in hybridization with DLMs for FR.

References

1. Li, L., Peng, Y., Qiu, G., Sun, Z., Liu, S.: A survey of virtual sample generation technology for face recognition. Artif. Intell. Rev. **50**, 1–20 (2018)
2. Vishwakarma, V.P., Pandey, S., Gupta, M.N.: Pose invariant face recognition using virtual frontal view generation. In: Proceedings of the International Conference on Computing, Communication and Networking (ICCCN 2008), pp. 1–5 (2008)
3. Sharif, M., Mohsin, S., Hanan, R.A., Javed, M.Y., Raza, M.: Using nose heuristics for efficient face recognition. Sindh Univ. Res. J. SURJ (Science Ser. 43) (2011)
4. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. **35**, 399–458 (2003)
5. De Marsico, M., Nappi, M.: Face recognition in adverse conditions: a look at achieved advancements. In: Computer Vision: Concepts, Methodologies, Tools, and Applications, pp. 2184–2210. IGI Global (2018)
6. Ping Tian, D., et al.: A review on image feature extraction and representation techniques. Int. J. Multimed. Ubiquitous Eng. **8**, 385–396 (2013)
7. Rouhi, R., Amiri, M., Irannejad, B.: A review on feature extraction techniques in face recognition. Signal Image Process. **3**, 1 (2012)
8. Yadav, S., Vishwakarma, V.P.: A new interval type 2 fuzzy based pixel wise information extraction for face recognition. Int. J. Appl. Pattern Recognit. **5**, 171–190 (2018)
9. Yadav, S., Vishwakarma, V.P.: Interval type-2 fuzzy based pixel wise information extraction: an improved approach to face recognition. In: 2016 International Conference on Computational Techniques in Information and Communication Technologies, ICCTICT 2016—Proceedings (2016)
10. Li, S., Jain: Handbook of Face Recognition (2005)
11. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
12. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Pattern Anal. Mach. Intell. **12**, 103–108 (1990). https://doi.org/10.1109/34.41390
13. Aguerrebere, C., Capdehourat, G., Delbracio, M., Mateu, M., Fernández, A., Lecumberry, F.: Aguara: an improved face recognition algorithm through gabor filter adaptation. In: 2007 IEEE Workshop on Automatic Identification Advanced Technologies, pp. 74–79 (2007)
14. Sharif, M., Naz, F., Yasmin, M., Shahid, M.A., Rehman, A.: Face recognition: a survey. J. Eng. Sci. Technol. Rev. 10 (2017)
15. Dey, A., Chowdhury, S., Sing, J.K.: Feature extraction using fuzzy generalized two-dimensional inverse LDA with Gaussian probabilistic distribution and face recognition. In: Advanced Computational and Communication Paradigms, pp. 553–561. Springer (2018)

16. Tamura, S., Tateishi, M.: Capabilities of a four-layered feedforward neural network: four layers versus three. IEEE Trans. Neural Netw. **8**, 251–255 (1997)
17. Vishwakarma, V.P.: Deterministic learning machine for face recognition with multi-model feature extraction. In: 2016 Ninth International Conference on Contemporary Computing (IC3), pp. 1–6 (2016)
18. Klir, G., Yuan, B.: Fuzzy Sets and Fuzzy Logic. Prentice hall, New Jersey (1995)
19. Melin, P., Castillo, O., Gonzalez, C.I., Castro, J.R., Mendoza, O.: General type-2 fuzzy edge detectors applied to face recognition systems. In: 2016 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS), pp. 1–6 (2016)
20. Polyakova, A., Lipinskiy, L.: A study of fuzzy logic ensemble system performance on face recognition problem. In: IOP Conference Series: Materials Science and Engineering, p. 12013 (2017)
21. Melin, P., Mendoza, O., Castillo, O.: Face recognition with an improved interval type-2 fuzzy logic Sugeno integral and modular neural networks. IEEE Trans. Syst. Man Cybern. A Syst. Humans **41**, 1001–1012 (2011)
22. Vishwakarma, V.P., Pandey, S., Gupta, M.N.: Fuzzy based pixel wise information extraction for face recognition. Int. J. Eng. Technol. **2**, 117 (2010)
23. Le Bihan, N., Sangwine, S.J.: Quaternion principal component analysis of color images. In: 2003 International Conference on Image Processing, 2003. ICIP 2003—Proceedings, p. I–809 (2003)