

Akhtar Kalam
Swagatam Das
Kalpana Sharma
Editors

Advances in Electronics, Communication and Computing

ETAERE-2016

Lecture Notes in Electrical Engineering

Volume 443

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadhyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

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

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

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

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

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

Akhtar Kalam · Swagatam Das
Kalpana Sharma
Editors

Advances in Electronics, Communication and Computing

ETAERE-2016

 Springer

Editors

Akhtar Kalam
Smart Energy Research Unit, College
of Engineering and Science
Victoria University
Melbourne, VIC
Australia

Kalpana Sharma
Department of Computer Science
and Engineering
Sikkim Manipal Institute of Technology
Rangpo, Sikkim
India

Swagatam Das
Electronics and Communication Sciences
Unit
Indian Statistical Institute
Kolkata, West Bengal
India

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-4764-0 ISBN 978-981-10-4765-7 (eBook)
<https://doi.org/10.1007/978-981-10-4765-7>

Library of Congress Control Number: 2017940354

© Springer Nature Singapore Pte Ltd. 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

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

Foreword

Advances in Electronics, Communication and Computing is a collection of research articles and critical review articles presented in the International Conference on “Emerging Trends and Advances in Electrical Engineering and Renewable Energy —ETAEEERE 2016,” organized by the Department of Electrical and Electronics Engineering (EEE) of Sikkim Manipal Institute of Technology (SMIT), Majhitar, Sikkim, India during December 17–18, 2016. This was a very unique conference which combined renewable energy, electronics, computing, communication, systems, controls, and automations under one roof. Moreover, it is a matter of honor for SMIT to learn that Springer was associated with ETAEEERE-2016 as a major publication sponsor for the event. The proceedings of this conference came out with four different book volume titles under Lecture Notes in Electrical Engineering (LNEE). This book is a compilation of research work in the interdisciplinary areas of electronics, communication, and computing. The chapters of this book cover the different approaches and techniques for specific applications, such as particle swarm optimization, Otsu’s function and harmony search optimization algorithm, triple-gate silicon-on-insulator (SOI) MOSFET, micro-Raman and Fourier transform infrared spectroscopy (FTIR) analysis, high-k dielectric gate oxide, spectrum sensing in cognitive radio, microstrip antenna, ground-penetrating radar (GPR) with conducting surfaces, and digital image forgery detection.

Eminent speakers like Prof. A Chakrabarti (former Vice Chancellor) of Jadavpur University, Prof. A Rajaraman of IIT Chennai, Prof. Gyoo-Shee Chae of Baekshok University, South Korea, Prof. Avinash Konkani of University of Virginia, USA, Prof. Kamani KK (the global economic advisor of Karnataka), Prof. Manjesh of Bangalore University, and Dr. Amitanshu Patnaik of DRDO Delhi shared their knowledge and experience. The conference attended and presented by participants from institutes like IISc, IITs, NITs, NEHU, BIT, VIT, MIT Manipal, IEST Kolkata and abroad deliberated on their research works. Additionally, the paper presentations were accompanied by six keynote addresses from leading academic and industry researchers around the globe. The paper presentations took place in three different tracks with 18 parallel sessions. Through the platform of

ETAEEERE-2016, we got the opportunity to promote the national campaign “Make In India”.

The review committee has done an excellent job in reviewing the articles and approving the high-quality research articles to be published in the conference proceedings by LNEE—Springer. The editors are thankful to all the faculty and students of these various committees for their dedication in making this a very successful conference and also to the staff of Springer for making the compilation possible. We sincerely hope that this volume will inspire researchers.

Rangpo, India

Karma Sonam Sherpa
Akas Kumar Bhoi
Mohammed Nasir Ansari
Amit Kumar Singh

Editorial Board

Chief Patron

Prof. (Dr.) Somnath Mishra, Vice Chancellor, Sikkim Manipal University

Patron

Prof. (Dr.) Ashis Sharma, Registrar, SMU

Prof. (Dr.) Amik Garg, Director, SMIT

Prof. (Dr.) Sadasivan Thekkey Veetil, Joint Director, SMIT

Programme Chair

Dr. Rabindranath Bera, SMIT

Dr. Karma Sonam Sherpa, SMIT

Dr. Kalpana Sharma, SMIT

Dr. H.K.D. Sarma, SMIT

Dr. Tejbanta Singh Chingtham, SMIT

Dr. Utpal Deka, Physics, SMIT

Dr. B.B. Pradhan, SMIT

Dr. Samarjeet Borah, Dept. Of CA, SMIT

Dr. Gobinda Chandra Mishra, SMIT

Prof. Om Prakash Singh, SMIT

Special Session Chairs

Dr. Sabyasachi Sen Gupta, IIT Kharagpur
 Dr. Samrat Paul, NEHU, Shillong
 Dr. Swagatam Das, ISI, Kolkata
 Dr. Abhijit Chakrabarti, IEST, Shibpur
 Prof. Kamani K.K, Dept. of Higher Education, Govt. of Karnataka (Global Economic Adviser)
 Dr. G.S. Chae, Baekseok University, South Korea
 Prof. Natarajan Gajendran, Editor IJST, President (iSee)
 Dr. Manjesh, Dept. of Electronic Science, Bangalore University
 Dr. Amitanshu Patnaik, DTRL, DRDO

International Advisory Committee

Dr. Avinash Konkani, AHFP, Clinical Engineer, University of Virginia Health System, USA
 Dr. P. Sanjeevikumar, Dept. of EEE, University of Johannesburg, South Africa
 Dr. Ahmed Faheem Zobaa, BU, UK
 Dr. Akhtar Kalam, VU, Australia
 Dr. David YU, UWM, US
 Dr. Atilla Elci, Chairman, Dept. of EEE, Aksaray University, Turkey
 Dr. Dmitri Vinnikov, TUT, Estonia
 Dr. Hussain Shareef, UKM, Malaysia
 Dr. Seshadev Sahoo, Purdue University, US
 Dr. Anil Kavala, Sr. Engineer, Samsung Electronics, Seoul, South Korea
 Dr. Kamran Morovati, University of New Brunswick, Canada
 Dr. Joseph Olorunfemi Ojo, TTU, USA
 Dr. Mohamed A. Zohdy, OU, MI
 Dr. Murad Al-Shibli, Head, EMET, Abu Dhabi
 Dr. Nesimi Ertugrul, UA, Australia
 Dr. Omar Abdel-Baqi, UWM, USA
 Dr. Richard Blanchard, LBU, UK
 Dr. Shashi Paul, DM, UK

National Advisory Committee

Dr. Sabyasachi Sen Gupta, IIT Kharagpur
 Prof. Kamani K.K., Dept. of Higher Education, Govt. of Karnataka (Global Economic Adviser)

Dr. Manjesh, Dept. Of Electronic Science, Bangalore University
 Dr. Amitanshu Patnaik, DTRL, DRDO
 Dr. Swagatam Das, ISI Kolkata
 Dr. Ajoy Kr. Ray, SMIT
 Dr. Ajeya Jha, SMIT
 Dr. Rabindranath Bera, SMIT
 Dr. Karma Sonam Sherpa, SMIT
 Dr. Kalpana Sharma, SMIT
 Dr. B.B Pradhan, SMIT
 Dr. H.K.D Sarma, SMIT
 Dr. Debabrata Pradhan, IIT, Kharagpur
 Prof. C.J. Thomas, SMIT
 Dr. Bidita Khandelwal, SMIMS
 Dr. Sangeeta Jha, SMIT
 Dr. Vinod Kumar Sayal, SMIT
 Dr. Arun Baran Samaddar, Director, NIT Sikkim
 Dr. Gobinda Chandra Mishra, SMIT
 Dr. V.K. Manhoharr, CEO & MD, TechMaven Private Limited
 Dr. Anjan Kumar Ray, NIT Sikkim
 Dr. Ashok Kumar Pradhan, IIT Kharagpur
 Dr. Bhim Singh, IIT Delhi
 Dr. Ganapati Panda, IIT Bhubaneswar
 Dr. Debashish Jena, NITK, India
 Dr. N.P. Padhy, IIT Roorkee
 Dr. C. Subramani, IIT Roorkee
 Dr. Patnaik S.K., Anna University
 Dr. G. Uma, Anna University
 Dr. Avik Bhattacharya, IIT Roorkee
 Dr. Smarajit Ghosh, Thapar University
 Dr. Sandeep Chakravorty, Baddi University
 Dr. Krishnendu Chakraborty, KGEC
 Dr. Urmila Kar, NITTTR, Kolkata
 Dr. Abhijit Chakrabarti, IEST, Shibpur
 Er. Kunal Munshi, Managing Partner at Sunrator Technologies LLP
 Er. Bikash Rai, Assistant Engineering, Power & Energy Department, Govt. of Sikkim

Reviewer Committee

Dr. Ahmed Faheem Zobia, Dept. of ECE, BU, UK
 Dr. Avinash Konkani, AHFP, Clinical Engineer, University of Virginia Health System, USA

Dr. Atilla Elci, Chairman, EEE, Aksaray University, Turkey
Dr. Kamran Morovati, CS, University of New Brunswick, Canada
Dr. Karma Sonam Sherpa, HOD EEE, SMIT
Dr. Chandrashekar Bhuiyan, CE, SMIT
Dr. Seshadev Sahoo, ME, Purdue University, US
Dr. Akhtar Kalam, Leader–Smart Energy Research Unit, VU, Australia
Dr. Richard Blanchard, Renewable Energy, LBU, UK
Dr. Utpal Deka, Physics, SMIT
Dr. Nitai Paitya, ECE, SMIT
Dr. N.K. Bhattacharyya, Chemistry, SMIT
Dr. A. Senthil Kumar, EEE, VEC, Chennai
Dr. Samarjeet Borah, Dept. Of CA, SMIT

Organizing Committee (Sikkim Manipal Institute of Technology)

Overall Incharge

Prof. Karma Sonam Sherpa—Convenor
Prof. Akash Kumar Bhoi—Secretary
Prof. Mohammed Nasir Ansari—Convenor
Prof. Amit Kumar Singh—Convenor

Publicity

Prof. Pradeep Kumar Mallick
Prof. Himangshu Pal

Hall and Stage Arrangements

Prof. Shabbiruddin
Prof. Sandip Kumar Gupta

Caterings and Refreshment

Prof. Sunam Saha
Prof. Roshan Pradhan

Transport

Prof. Arunava Tikader
Prof. Arijit Ghosh

Help Desk and Registration

Prof. Chitragada Roy
Prof. Dibyadeep Bhattacharya

Sponsorship

Prof. Saikat Chatterjee
Prof. Anirban Sengupta

Guest Hospitality

Prof. Moumi Pandit

Prof. Bijay Rai

Prof. Rahul Kumar

Printing

Prof. Pratik Pradhan

Prof. Rajiv Pradhan

**Student Organizing Committee ((Sikkim Manipal
Institute of Technology))**

Yupphung Keimba Limbu

Kunglho Lepcha

Sagar Pradhan

Avinash Gupta

Avinash Pradhan

Tshering Sangmo Sherpa

Sisir Chettri

Deepika Chettri

Suyog Pradhan

Shristi Shresta

Pankaj Kumar

Biswas Pradhan

Ritika Pradhan

Aruna Chettri

Ashish Lamichaney

Neeta Lamichaney

Ashutosh Mukherjee

Amit kr. Singh

Gyan Shree

Dheeraj Kumar

Raval Parth Pradip

Srihari k.

Saurabh Bhowmik

Madhura Pardhe

Contents

Analysis of Authentication and Key Agreement (AKA) Protocols in Long-Term Evolution (LTE) Access Network	1
Mariya Ouaiassa, A. Rhattoy and M. Lahmer	
Impact of M2M Traffic in Random Access Channel over LTE Networks	11
Mariyam Ouaiassa, M. Benmoussa, A. Rhattoy, M. Lahmer and I. Chana	
Predictive Modeling of Students Performance Through the Enhanced Decision Tree	21
Subitha Sivakumar and Rajalakshmi Selvaraj	
A New Pulse Width Modulation Technique with Hybrid Carrier Arrangement for Multilevel Inverter Topology	37
N. Prabaharan, V. Arun, K. Palanisamy and P. Sanjeevikumar	
Watershed-Matching Algorithm: A New Pathway for Brain Tumor Segmentation	45
S.M. Kamrul Hasan, Yugoshree Sarkar and Mohiudding Ahmad	
A Compact Slotted Textile Patch Antenna for Ultra-wide Band Application	53
Nikhil Kumar Singh, Vinod Kumar Singh, Anurag Saxena, Akash Kumar Bhoi, Amik Garg and K.S. Sherpa	
Optimization of Hodgkin–Huxley Conductance-Based Model Using Particle Swarm Optimization and Firefly Method	61
Rashmi Deka, Kuntala Boruah and Jiten Ch. Dutta	
A Model to Demonstrate the Universality of DNA-NAND Gate	67
Kuntala Boruah, Rashmi Deka and Jiten Ch. Dutta	

Micro-Raman and FTIR Analysis of Silicon Carbo-Nitride Thin Films at Different H₂ Flow Rate	77
Dhruva Kumar, Umesh Rizal, Soham Das, Bhabani S. Swain and Bibhu Prasad Swain	
Applying Classification Methods for Spectrum Sensing in Cognitive Radio Networks: An Empirical Study	85
Nayan Basumatary, Nityananda Sarma and Bhabesh Nath	
Highly Negative Dispersive, Low Loss Single-Mode Photonic Crystal Fiber	93
Shahiruddin, Dharmendra K. Singh and Sneha Singh	
Triple Gate SOI MOSFET	103
Amit Agarwal, P.C. Pradhan and Bibhu Prasad Swain	
Wearable Antennas for Medical Application: A Review	115
Shreema Manna, Tanushree Bose and Rabindranath Bera	
Characterization of Silicon Carbo-Nitride Thin Films	131
Dhruva Kumar and Bibhu Prasad Swain	
Improved Normalization Approach for Iris Image Classification Using SVM	139
Mahaboob Shaik	
Investigation of Titanium Aluminium Nitride (TiAlN): A Review	147
Soham Das and Bibhu Prasad Swain	
The Study of GPR with Conducting Surfaces to Determine the Operating Frequency in Single-Fold Reflection Profiling	159
Amitansu Pattanaik and Rajat Sharma	
Investigation of Titanium Silicon Nitride: A Review	169
Spandan Guha, Asish Bandyopadhyay, Santanu Das and Bibhu Prasad Swain	
An Efficient MOS Class-E Power Amplifier for Wireless LAN Applications	181
M. Ananda and A.B. Kalpana	
Security in Cloud Computing-Based Mobile Commerce	191
Ferdousi Khatun and Md. Ruhul Islam	
From FET to SET: A Review	199
Amit Agarwal, P.C. Pradhan and Bibhu Prasad Swain	
Important Factors for Designing Assamese Prosody with Festival Frame Work	211
Parismita Sarma and Shikhar Kumar Sarma	

Location Privacy in Wireless Sensor Networks Using Anonymity 221
 Shriya Kesharwani and Manisha R. Dhage

A Compact CPW-Fed Defected Ground Microstrip Antenna for Ku Band Application 231
 Rashmi Singh, Vinod Kumar Singh and Puneet Khanna

A Study on Different Normalization Approaches of Word 239
 N. Chitrapriya, Md. Ruhul Islam, Minakshi Roy and Sujala Pradhan

Duplicate Resource Detection in RDF Datasets Using Hadoop and MapReduce 253
 Kumar Sharma, Ujjal Marjit and Utpal Biswas

TEECS: A Time-Based Energy Efficient Clustering Scheme in Wireless Sensor Networks 263
 Nirnaya Pradhan, Kalpana Sharma, Vikash Kumar Singh and Nima Donka Tamang

Packet Sniffing and Network Traffic Analysis Using TCP—A New Approach 273
 Aishwarya Bhandari, Samala Gautam, Tawal K. Koirala and Md. Ruhul Islam

Passive Techniques of Digital Image Forgery Detection: Developments and Challenges 281
 Santoshini Panda and Minati Mishra

m-GeoEduNet: Mobile SDI Model for Education Information Infrastructure Network 291
 R.K. Barik, R.K. Lenka, A.B. Samaddar, J. Pattnaik, B. Prakash and V. Agarwal

Investigation of Structural and Electronic Environments of Nitrogen-Doped CVD-Grown DLC Films 301
 Ranjan Kr. Ghadai, Soham Das, Subhash Chandra Mondal and Bibhu Prasad Swain

A Novel Region Growing Based Method to Remove Pectoral Muscle from MLO Mammogram Images 307
 Manasi Hazarika and Lipi B. Mahanta

Microstrip Patch Antenna: A Review 317
 Arun Kumar Singh, Rabindranath Bera and Bansibadan Maji

A Comparison Study of Face, Gait and Speech Features for Age Estimation 325
 Prachi Punyani, Rashmi Gupta and Ashwani Kumar

A Review on MIMO Based V2V Communication	333
Samarendra Nath Sur, Rabindranath Bera and Bansibadan Maji	
Energy-Efficient Packet Routing Model for Wireless Sensor Network	341
Madhu Patil and Chirag Sharma	
An ECC-Based Algorithm to Handle Secure Communication Between Heterogeneous IoT Devices	351
S. Sasirekha, S. Swamynathan and S. Suganya	
Enhanced Pixel Purity Index Algorithm to Find the Index Position of the Pure Pixels in Hyperspectral Images	363
S. Graceline Jasmine and V. Pattabiraman	
Assamese Character Recognition Using Zoning Feature	371
Kalyanbrat Medhi and Sanjib Kr. Kalita	
Improving Convergence in iBGP Route Reflectors	381
Shipra Shukla and Mahesh Kumar	
Performance Analysis of Various Eigenvalue-Based Spectrum Sensing Algorithms for Different Types of Primary User Signals	389
Pankaj Verma and Brahmjit Singh	
A Crown-Shaped Microstrip Patch Antenna for Wireless Communication Systems	399
Atanu Nag, Kousik Roy and Debika Chaudhuri	
Atrophy Measure of Brain Cortex to Detect Alzheimer’s Disease from Magnetic Resonance Images	407
Dulumani Das and Sanjib Kumar Kalita	
Study of Effect of High-k Dielectric Gate Oxide on the Performance of SB-GNRFETs	415
Bhubon Chandra Mech and Jitendra Kumar	
A Simple and Efficient Higher Order Finite Element Scheme for Helmholtz Waveguides	421
Tattwa Darshi Panda, K.V. Nagaraja and V. Kesavulu Naidu	
Point Feature Based Recognition of Handwritten Meetei Mayek Script	431
Chandan Jyoti Kumar and Sanjib Kumar Kalita	
Design of Circularly Polarized Antenna with Different Iterations for UWB Applications	441
Piyush Sharma, Ashok Yadav and Vinod Kumar Singh	

Weighted Wavelet Tree Sparsity Regularization for Compressed Sensing Magnetic Resonance Image Reconstruction 449
 Bhabesh Deka and Sumit Datta

A Score-Level Solution to Speaker Verification Using UBM Pooling and Adaptive Cohort Selection 459
 Pranab Das

A Novel Encryption Scheme for Secure SMS Communication 467
 Aeneaus Malla, Pranab Sahu and Minati Mishra

Reviewing the Performance of an Improved Seven-Level Multilevel Inverter for Various Pulse Width Modulation Techniques 479
 K. Aroul, S. Umashankar, K.R. Prabhu and P. Sanjeevikumar

Robust Multi-thresholding in Noisy Grayscale Images Using Otsu’s Function and Harmony Search Optimization Algorithm 491
 K. Suresh and U. Sakthi

An Approach for Visual Cryptography Scheme on Color Images 501
 Chinmoy Kar, Suman Kalyan Kar and Sreeparna Banerjee

A Novel Hybrid Approach for Influence Maximization in Online Social Networks Based on Node Neighborhoods 509
 Gypsy Nandi, Uzzal Sharma and Anjan Das

Diagnosis of Ground Glass Opacity in HRCT Lung Images Using High-Intensity Pixels 521
 Punal M. Arabi, Nanditha Krishna, T.P. Prathibha and N. Vamsha Deepa

Coefficient Random Permutation Based Compressed Sensing for Medical Image Compression 529
 R. Monika, S. Dhanalakshmi and S. Sreejith

A Flexible Scheme to Fault Detection for Electrical Assets Using Infrared Thermography 537
 Deepak Kumar, Amit Kumar and M.A. Ansari

Design of Double Fork Shaped Patch Radiator for Ultra Wide Band Applications 549
 Swarnaprava Sahoo, Mihir N. Mohanty and Laxmi P. Mishra

Subjective Analysis of Multimedia Traffic in Wireless Multimedia Sensor Networks 561
 A. Sivagami and S. Malarkkan

Performance Analysis of Collinear Beam-Steering Array Antenna 573
 Shuchismita Pani, Arun Kumar and Malay Ranjan Tripathy

E-Mail Spam Filtering: A Review of Techniques and Trends	583
Alexy Bhowmick and Shyamanta M. Hazarika	
Stealthy Attacks in MANET to Detect and Counter Measure by Ant Colony Optimization	591
Abhishek Ranjan, Rajalakshmi Selvaraj, Venu Madhav Kuthadi and Tshilidzi Marwala	
Inset Fed Circular Microstrip Antenna with Defected Ground	605
Nikhil Kumar Singh, Niraj Sharma, Zakir Ali, Vinod Kumar Singh and Akash Kumar Bhoi	
Analysis of Thermal Comfort and Visual Comfort—A Soft Computing Approach	613
Sandhyalaxmi G. Navada, Chandrashekara S. Adiga and Savitha G. Kini	
Spoken Dialog System in Bodo Language for Agro Services	623
Aniruddha Deka and Manoj Kumar Deka	
Random-Valued Impulse Denoising Using a Fast l^1-Minimization- Based Image inpainting Technique	633
Mayuri Kalita and Bhabesh Deka	
Comparative Analysis of DFT–DCT-Spreading Strategies-Based SC-FDMA	643
G.B.S.R. Naidu and V. Malleswara Rao	
A Review on Spread Spectrum Radar	653
Nima Donka Tamang, Samarendra Nath Sur, Soumyasree Bera and Rabindranath Bera	
An Empirical Analysis of Three Moments on Sattriya Dance Single-Hand Gestures Dataset	665
Mampi Devi and Sarat Saharia	
Implementation of Doppler Radar at 24 GHz	675
Manisha Das, Tarini Singh and Soumyasree Bera	
Efficient Use of Location of Unlicensed Users in Improving the Utilization of Licensed Spectrum	685
G. Verma and O.P. Sahu	
Parking Places Discovery and Reservation Using Vehicular Ad Hoc Networks	695
Alak Roy, Jayati Paul, Ratnadeep Baidya and Mampi Devi	
Design of 5.8 GHz Rectenna for Space-Based Solar Power	705
Deepak Kumar and Kalpana Chaudhary	

Dynamic 2-D Pattern-Based Passive Communication in Swarm Agents 713
 V. Navya Deepthi, Chayan Banerjee and Shubhra Saxena

Cloud-Based Cognitive Premise Security System Using IBM Watson and IBM Internet of Things (IoT) 723
 Shubhradeep Nandi

A MATLAB-Based Simulator for Amorphous Silicon and Polycrystalline Silicon Thin Film Transistor 733
 Suman Das, Subham Datta and Somenath Chatterjee

Relay Node Selection Using Analytical Hierarchy Process (AHP) for Secondary Transmission in Multi-user Cooperative Cognitive Radio Systems 745
 Jyoti Sekhar Banerjee, Arpita Chakraborty and Abir Chattopadhyay

Mutual Information-Based Hierarchical Band Selection Approach for Hyperspectral Images 755
 Sonia Sarmah and Sanjib Kumar Kalita

Petri Net: Design and Analysis of Parallel Task Scheduling Algorithm 765
 Sasmita Parida, Suvendu Chandan Nayak, Piyush Priyadarshi, Prasant Kumar Pattnaik and Gaurav Ray

Opportunistic Selection of Threshold Using Double Threshold in Cognitive Radio Networks 777
 G. Soumya and G. Verma

X-Band Patch Array Antenna 785
 Nitya Nirmal, Ayushi Agarwal, Arun Kumar Singh, Rabindranath Bera and Bansibadan Maji

A Compact Pentagonal Textile Microstrip Antenna for Wide Band Application 793
 Rachana Kushwaha, Vinod Kumar Singh, Nikhil Kumar Singh, Anurag Saxena and Deepak Sharma

Similarity Measurement of Gene Using Arc Tan Function in Gene Ontology 801
 K. Vengatesan, S.B. Mahajan, P. Sanjeevikumar, R.M. Kulkarni and Sana Moin

About the Editors

Prof. Akhtar Kalam has been working at Victoria University, Melbourne, since 1985 and was Deputy Dean of the Faculty of Health, Engineering and Science for 7 years. He is currently the Head of the Engineering Department. He is also the current Chair of the Academic Board and lectures in the Masters by coursework program in the Engineering Institute of Technology, Perth, Australia. Further, he has Distinguished Professorship position at the University of New South Wales, Sydney, Australia, and five Malaysian universities. He received his B.Sc. and B.Sc. Engineering degrees from University of Calcutta and the Aligarh Muslim University, India. He completed his MS and Ph.D. degrees at the University of Oklahoma, USA, and the University of Bath, UK.

Dr. Swagatam Das is currently serving as Assistant Professor at the Electronics and Communication Sciences Unit of the Indian Statistical Institute, Kolkata, India. His research interests include evolutionary computing, pattern recognition, multi-agent systems, and wireless communication. Dr. Das has published one research monograph, one edited volume, and more than 200 research articles in peer-reviewed journals and international conferences. He is the founding co-editor-in-chief of “Swarm and Evolutionary Computation”, an international journal from Elsevier. He also serves as the associate editor of the IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Computational Intelligence Magazine, IEEE Access, Neurocomputing (Elsevier), Engineering Applications of Artificial Intelligence (Elsevier), and Information Sciences (Elsevier). Dr. Das has 7500+ Google Scholar citations and an H-index of 44 till date.

Prof. (Dr.) Kalpana Sharma completed her Bachelor of Engineering (B.E.) from National Institute of Technology (NIT), Silchar, in 1998 and M.Tech. from the Department of Computer Science and Technology (CSE), Indian Institute of Technology Kharagpur (IIT, Kharagpur) and Ph.D. from Sikkim Manipal University (SMU), Sikkim. She is presently working as Professor and Head of the Department of Computer Science and Engineering (CSE). She has been actively participating at institute level and university level research and development (R&D) activities and has successfully executed the All India Council for Technical Education (AICTE) funded research project: “Design of an Integrated Security Scheme of Wireless Sensor Network”. Her research areas include sensor networks, software engineering, real-time systems and security. She has published a number of technical papers in national and international journals and conferences. She has also organized workshops, seminars, and conferences.

Analysis of Authentication and Key Agreement (AKA) Protocols in Long-Term Evolution (LTE) Access Network

Mariya Ouaisa, A. Rhattoy and M. Lahmer

Abstract Mobile communications systems have evolved considerably in recent years. Fourth generation networks (4G) allow to expand cellular coverage while improving accessibility to Internet services. Network access security includes security features that provide the subscriber with secure access to services of the EPS, and protects particularly against several attacks on the radio link. This area is the largest and most vulnerable among all EPS security domains since it ensures the security of the radio link, the weakest link of all mobile telephone networks. In the LTE architecture, Evolved Packet System Authentication and Key Agreement (EPS-AKA) procedure is used to realize mutual authentication between the subscriber and the network. However, the current authentication and key agreement protocol 3GPP LTE-AKA has some issues, including bandwidth consumption, traffic overload due to control message and vulnerabilities. Several protocols have been proposed to solve these problems. In this paper, we will analyze and compare several existing protocols: EPS-AKA, SE-AKA, EC-AKA, and EC-AKA2 according to different factors in order to estimate the performances in terms of security, cost, and delay of each one of these protocols.

Keywords LTE/SAE · AKA · Network security · Authentication

M. Ouaisa (✉)

ISIC, High School of Technology, LMMI Laboratory, ENSAM,
Moulay-Ismaïl University, Meknes, Morocco
e-mail: mariya.ouaisa@edu.umi.ac.ma

A. Rhattoy · M. Lahmer

Department of Computer Engineering, High School of Technology,
Moulay-Ismaïl University, Meknes, Morocco
e-mail: rhattoy@gmail.com

M. Lahmer

e-mail: mohammed.lahmer@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_1

1 Introduction

Mobile communications systems have evolved considerably in recent years. Fourth generation networks (4G) allow to expand cellular coverage while improving accessibility to Internet services. The new mobile network system called Evolved Packet System (EPS) comprises a new access network called LTE (Long-Term Evolution) and a new core network called Evolved Packet Core (EPC) or System Architecture Evolution (SAE). All services are offered by a packet domain. The mobile must first connect and authenticate to the EPS network before being able to send or receive Internet Protocol (IP) packets [1, 2]. The element that allows authentication is called Mobility Management Entity (MME), the Home Location Register/Authentication Center (HLR/AuC) is characterized by Home Subscriber Service (HSS) in the EPS architecture (Fig. 1). Security in the fourth generation mobile networks EPS includes security of the radio access network infrastructure, terminals, and applications running on it. Network access security includes security features that provide the subscriber with secure access to services of the EPS, and protects particularly against the attacks on the radio link. One of the most important security services is authentication, 3GPP LTE-AKA protocol has some problems, including the consumption of bandwidth, the traffic generated: control and authentication messages and several vulnerabilities such as disclosure of user identity, man-in-middle attack, etc. In recent years, many articles have been published on the subject of security, focusing mainly on the applicative field. Some recent works on cellular communications have nevertheless made significant advances and demonstrated realism of attacks previously reputed theoretical. These studies identify certain inertness in the consideration of threats. This article is primarily concerned with these works and presents an overview of the security of communications in LTE Network. The second section of this article presents first a brief description of the elements composing the 4G network. These elements are necessary for understanding the principles underlying the security of mobile networks and their vulnerabilities, and then explain the principle steps of Evolved Packet System Authentication and Key Agreement (EPS-AKA) procedure.

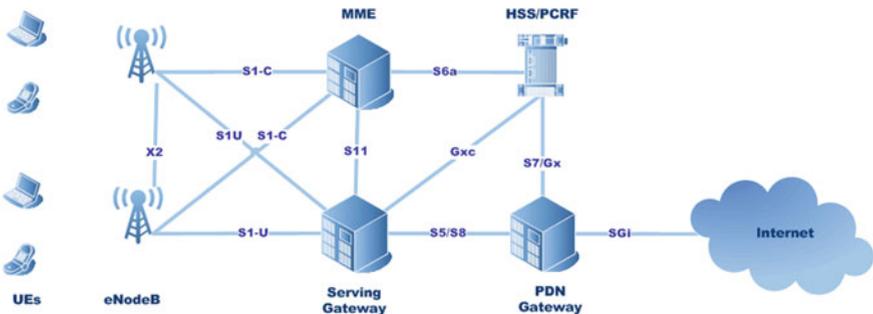


Fig. 1 LTE network architecture

Section 3 addresses in one hand the security analysis of AKA protocols of 4G and presents in the other hand, a comparison of performances between various existing security protocols in terms of communication overhead and delay, and finally conclusions are retained for Sect. 4.

2 LTE System Architecture and Security Mechanisms

This section presents an overview of the main elements of LTE network. We separate this description into two parts: the first is devoted to the infrastructure of the 4G network and the second to the mechanisms of security and the EPS-AKA procedure used to realize mutual authentication between the user and the network.

2.1 System Architecture

LTE is also referred to as the Evolved Packet System (EPS). The EPS is divided between the radio access network Evolved UMTS Terrestrial Radio Access Network (EUTRAN) and the EPC. In EUTRAN, mobile devices are called User Equipment (UE). The operation of base stations has been defined from the Universal Mobile Telecommunications System (UMTS) network architecture. That is why they bear the same designation than UMTS, Evolved Node B (eNodeB). As in Fig. 1, each eNodeB connects with the EPC via the S1 interface and can also be connected to the neighboring base stations by the X2 interface [3, 4].

EPC uses IP as a transport medium. The MME is used to connect users to the network and to localize them on the LTE network. For this, the MME accesses the HSS. If the terminal has a valid SIM card, the account of the participant is assigned to a Serving Gateway (SGW). From there, a connection is established with the Packet Data Network Gateway (PDN-GW), which assigns an IP address to the terminal and establishes a connection with the IP network of the operator. The EPC also includes the Policy and Charging Rules Function (PCRF).

2.2 EPS-AKA Procedure

Before you can transfer or receive IP packets, the mobile must first connect and authenticate to the EPS network [5]. The element that allows authentication is called MME. The HLR/AuC is replaced by an HSS in the EPS architecture. Authentication vectors (AV) are generated by the MME from the HSS, as displayed in Fig. 2, through the interface S6 (based on DIAMETER) when the MME receives from the UE the Attach Request or Service Request messages. The MME launches along with the UE the EPS-AKA.

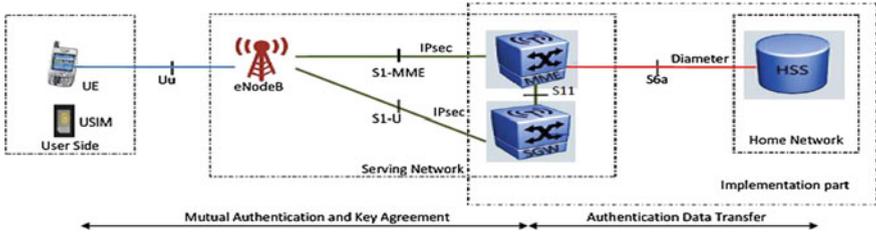


Fig. 2 EPS security architecture

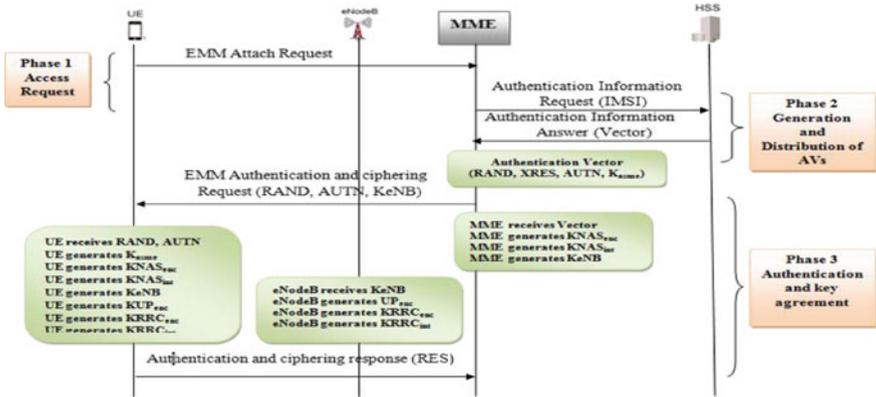


Fig. 3 EPS-AKA procedure

Figure 3 shows the progress of the EPS-AKA procedure which is initiated by the identification of the user through his IMSI or Globally Unique Temporary Identifier (GUTI) [6].

3 Performance Analysis and Simulations Results

In this section, we analyze and compare the performance of several authentication protocols: EPS-AKA, Security Enhanced Authentication and Key Agreement (SE-AKA) [7], Ensured Confidentiality Authentication and Key Agreement (EC-AKA) [8], and EC-AKA2 [9] protocols according to different parameters to estimate the performance and quality of service (QoS) in terms of security, cost, and delay of each one of these protocols. According to the survey we can see that in most existing 4G protocols, a high level of security can have a high processing cost and great added data rate (additional overhead).

To evaluate the performance of each protocol, we can consider the following parameters:

- **Security:** the security of a protocol is defined as its ability to resist attacks, while the risk is the probability that an attack could succeed in violating the protocol.
- **Overhead:** This is the added traffic on transmission interfaces in order to apply the relevant protocol.
- **Computational delays:** To calculate this delay, we consider the number and type of some operations required in each protocol.

3.1 Protocols Security

The security of a protocol is defined as its ability to resist attacks and the risk is the probability that an attack could succeed in violating the protocol. The more the cost and efforts to exploit vulnerability are high (increases), the more the probability of an attack's success decreases. In order to compare the security protocols' performance, EPS-AKA, SE-AKA, EC-AKA and EC-AKA2, we have compiled in Table 1, all of these studies are found in the literature.

After modeling the EPS-AKA protocol in High-Level Protocol Specification Language (HLPSSL) in order to be able to verify its security using Automated Validation of Internet Security Protocols and Applications (AVISPA) [10], the result indicates that it becomes insecure if the MME-HSS interface is not considered secure. When roaming, HSS and MME belong to different networks, so we believe it is open to attack if no closed network is used.

The SE-AKA protocol was analyzed using the authentication test method and it has been proved to be insecure, but the effort to exploit it exceeds one of the EPS-AKA. EC-AKA and EC-AKA2 protocols have been verified using AVISPA

Table 1 Comparison of different protocol security

Vulnerability	EPS-AKA	SE-AKA	EC-AKA and EC-AKA2
Ensure confidentiality of IMSI	No	Yes	Yes
Resistance against replay attack	No	No	No
Resistance against the DoS attack of UE	No	No	Yes
Resistance against the blocking of services by a man-in-the-middle (MITM)	No	No	Yes
Confidentiality of MME-HSS interface	No	Yes	Yes
Resistance against attacks on the responses of authentication data	No	Yes	No
Resistance against the DoS attack of HSS	No	No	No
Resistance against the usurpation of identity of MME	No	No	No

and transformed to be sure. The more protocol is secure, the lower the vulnerability it gets. The security level is inversely proportional to the level of vulnerability. From Table 1, we see that the two versions of EC-AKA protocol are the least vulnerable, and as such, they are the most secured.

3.2 Communication Overhead

In this section, we calculate the size of the transmitted messages, in order to estimate their communication overhead with regard to the EPS-AKA and other AKA protocols. The size of parameters in bit is shown in Table 2. On the basis of each parameter passed in the message, we can calculate the total number of bits in all messages during each protocol (where n is the number of authentication requests) [11].

Number of bits in step 1 = Σ messages (phase 1 + phase 2)

Number of bits in step 2 = Σ message (phase 3 * n)

Total number of transmitted bits = Σ messages (phase 1 + phase 2) + Σ message (phase 3 * n)

Figure 4 illustrates the overhead according to the number of AVs for different AKA protocols. According to this figure, it can be seen that EPS-AKA protocol generates the least overhead compared to other protocols.

Table 2 Size of parameters

Parameter	Size (bits)
IMSI/TMSI/GUTI	128
K/CK/EK/IK/TIK	128
AK	48
K_{ASME}	256
KSI_{ASME}	64
XRES/RES	128
AV	Variable
RAND/RANDIK/RANDEK/AUTN	128
SQN/XSQN/PSQN/NSQN	48/32
AMF	48
LAI	40
MACi/XMAC/MACue	64
SN id/IDhss	20
RandUESecCapab	6
UESecCapab	12

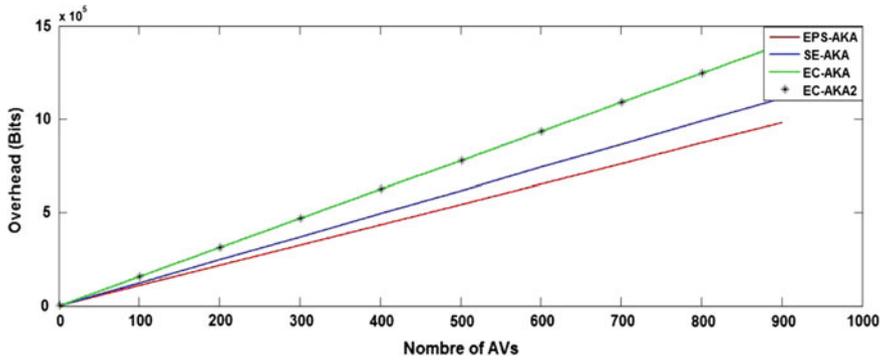


Fig. 4 Communication overhead for different AKA protocols

3.3 Computational Delays

The computation delays of all protocols are evaluated and compared to similar schemes. Corresponding calculations of the delays are due to calculations made at each of the following elements: UE, MME, and HSS. According to 3GPP, the functions f0, f1, f2, f3, f4, and f5 are Hash-based Message Authentication Code—Secure Hash Algorithm 256 (HMAC-SHA256). For digital signatures, we will study the overhead associated with Digital Signature Algorithm (DSA). For public key encryption, we will investigate the use of Rivest, Shamir, Adleman (RSA) and for symmetric encryption we will use the Advanced Encryption Standard (AES) algorithm.

The delay of calculation for each of the functions mentioned above is presented in Table 3. Delay values, available in [12], were obtained by measurements using Microsoft Visual C++ 2005 SP1 running on a processor Intel Core 21.83GHz Windows Vista32-bit.

To calculate the delay, we consider the number and type of some operations required in each protocol. The number and type of operations for each protocol are described in Table 4 [13].

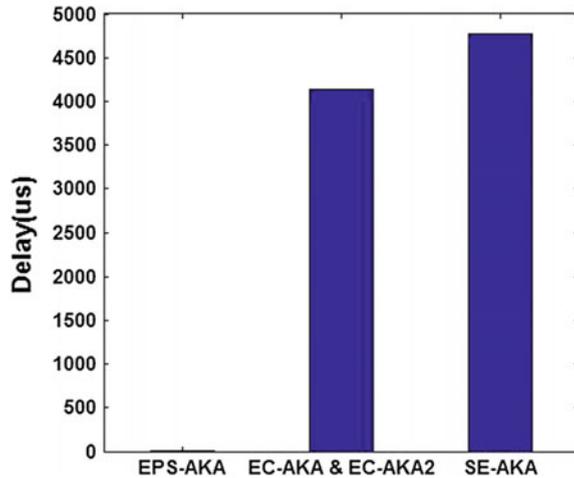
Figure 5 shows the different protocol delays. From the figure we see that SE-AKA protocol that uses public key cryptography has a great delay of calculation followed by EC-AKA hybrid protocol.

Table 3 Computational delay

Function	Delay (μ s)
HMAC-SHA-256	0.55
DSA-1024 (signature)	450
DSA-1024 (verification)	520
RSA-1024 (encryption)	80
RSA-1024 (decryption)	1460
AES-256	161

Table 4 Number of operations

Protocol	Operation	UE	MME	HSS
EPS-AKA	HMAC	6	–	6
	Encryption	–	–	–
	Decryption	–	–	–
SE-AKA	HMAC	6	–	6
	Encryption	1	2	1
	Decryption	1	1	1
EC-AKA	HMAC	6	–	6
	Encryption (asymmetric)	1	1	1
	Decryption (asymmetric)	–	1	1
	Encryption (symmetric)	1	2	–
	Decryption (symmetric)	2	1	–

Fig. 5 Computation delays

4 Conclusion

3GPP propose EPS-AKA for supporting authentication in the next generation mobile communication system. However, from the security analysis, the present paper has found divers vulnerabilities in EPS-AKA, several protocols have been proposed to solve these problems. The analysis that we did between existing AKA protocols EPS-AKA, SE-AKA, EC-AKA, and EC-AKA2 to estimate their performances in terms of security, cost, and delay showed the importance of each protocol according to the previous metrics. We conclude that the two versions of

EC-AKA protocol are the least vulnerable, thus the most secured. EPS-AKA protocol generates the least overhead compared to other protocols. SE-AKA protocol that uses public key cryptography has the greatest delay of calculation.

References

1. Fritze, G.: SAE: The Core Network for LTE. Ericsson (2012)
2. Netmanias Technical Document.: LTE Security I: LTE Security Concept and LTE Authentication (2013)
3. Bouguen, Y., Hardouin, E., Xavier Wolff, F.: LTE et les réseaux 4G (Chap 19). Eyrolles. ISBN: 978-2-212-12990-8 (2012)
4. Hu, H., Zhang, J., Zheng, X., Yang, Y., Wu, P.: Self configuration and self-optimization for LTE networks. *IEEE Commun. Mag.* 94–100 (2010)
5. Han, C., Choi, H.: Security analysis of handover key management in 4G LTE/SAE networks. *IEEE Trans. Mob. Comput.* **13**(2), 457–468 (2014)
6. 3GPP TS 35.206 V11.0.0. Technical Specification. 3G Security: Specification of the MILENAGE Algorithm Set: An Example Algorithm Set for the 3GPP Authentication and Key Generation Functions f1, f1*, f2, f3, f4, f5 and f5*. Document 2: Algorithm Specification (Release 11) (2012–09)
7. Li, X., Wang, Y.: Security enhanced authentication and key agreement protocol for LTE/SAE network. In: 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), pp. 1–4 (2011)
8. Bou Abdo, J., Chaouchi, H., Aoude, M.: Ensured confidentiality authentication and key agreement protocol for EPS. In: 3rd Symposium on Broadband Networks and Fast Internet (2012)
9. Bou Abdo, J., Chaouchi, H., Aoude, M., Pujolle, G.: EC-AKA2 a revolutionary AKA protocol. In: 2013 International Conference on Computer Applications Technology (ICCAT), pp. 1–6 (2013)
10. AVISPA Project. <http://www.avispa-project.org/>
11. Saxena, N., Thomas, J., Chaudhari, N.S.: ES-AKA: An Efficient and Secure Authentication and Key Agreement Protocol for UMTS Networks (2015)
12. Dai, W.: Crypto++5.6.0 Benchmarks (Online). Available: <http://www.cryptopp.com/benchmarks.html> (2009)
13. Hamandi, K., Sarji, I., Chehab, A., Elhadj, I.H., Kayssi, A.: Privacy enhanced and computationally efficient HSK-AKA LTE scheme. In: 27th International Conference on Advanced Information Networking and Applications Workshops (2013)

Impact of M2M Traffic in Random Access Channel over LTE Networks

Mariyam Ouaisa, M. Benmoussa, A. Rhattoy, M. Lahmer
and I. Chana

Abstract Nowadays, humanity is surrounded by many technological devices that are connected to the Internet (sensors, smart meters, etc.). These devices can transmit and receive data automatically via interfaces for wireless transmission of data, thus generating Machine-to-Machine (M2M) traffic. This massive connectivity provides attractive new services, usually transmitting fragments of small data. This is especially true for Long-Term Evolution (LTE), which was originally optimized for users Human-to-Human (H2H). The main problem caused by the introduction of M2M applications is the congestion that occurs at the access network because of the tremendous number of devices that attempt to simultaneously access the network. This article provides M2M communication methods based on the 3GPP standard. In this paper we consider concurrent access to radio resources in a M2M/H2H coexistence scenario based on a dynamic random access algorithm for LTE network. This work studies the impact of M2M terminals massively and simultaneously attempting to have a random access to LTE. First, we emphasize the problem of resources sharing by MTC devices. Then we assess the impact of the introduction of M2M services on the performance of applications H2H according to throughput metric.

Keywords M2M · MTC · H2H · RACH · Internet of Things · LTE

M. Ouaisa (✉) · M. Benmoussa
ISIC, High School of Technology, LMMI Laboratory, ENSAM,
Moulay-Ismaïl University, Meknes, Morocco
e-mail: mariyam.ouaisa@edu.umi.ac.ma

M. Benmoussa
e-mail: benmoussa.mariam@gmail.com

A. Rhattoy · M. Lahmer · I. Chana
Department of Computer Engineering, High School of Technology,
Moulay-Ismaïl University, Meknes, Morocco
e-mail: rhattoy@gmail.com

M. Lahmer
e-mail: mohammed.lahmer@gmail.com

I. Chana
e-mail: idrisschana@gmail.com

1 Introduction

Machine-to-Machine (M2M) is a trending concept that allows a huge number of machines to communicate independently via a network without human intervention. M2M allows the emergence of new services from different application areas such as transport, health, monitoring, and improving the human life. M2M was normalized in cellular networks Long-Term Evolution Advanced (LTE-A) by Machine-Type Communication (MTC) [1], their massive connectivity and ability to offer new interactive services, also raised significant challenges for managing a large number of devices, usually transmitting small fragments of data through wide range of emerging applications for the current cellular technology such as 3rd Generation Partnership Project (3GPP) LTE, which has been improved for the Human to Human (H2H) traffic. The community of 3GPP has been interested by various topics of study related to M2M communication.

LTE is the latest evolution of mobile telephony standards defined by the 3GPP. The network includes two parts: the radio access network Evolved UMTS Terrestrial Radio Access Network (EUTRAN) and the core network Evolved Packet Core (EPC). The base network EPC uses all Internet Protocol (IP) (Fig. 1) [2, 3].

Many challenges will face the operators of current mobile networks during the introduction of M2M applications for heterogeneous communication scenarios. The main problem caused by the introduction of these applications is the congestion that occurs at the access network because of the large number of devices that try to simultaneously access the network. The remainder of this paper is organized as follows. Section 2 introduces an overview of MTC in LTE Network. Section 3 explains the LTE random access procedure. Section 4 presents the impact of M2M in Random Access Channel (RACH) according to different metrics. In Sect. 5 we compare the throughput and collision probability of H2H and M2M Random Access. Finally, conclusions are retained for Sect. 6.

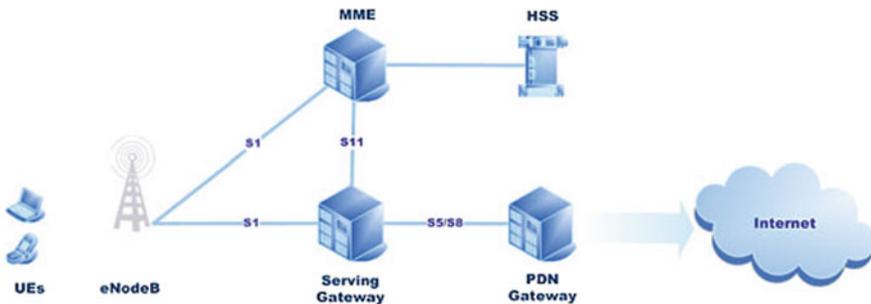


Fig. 1 LTE network architecture

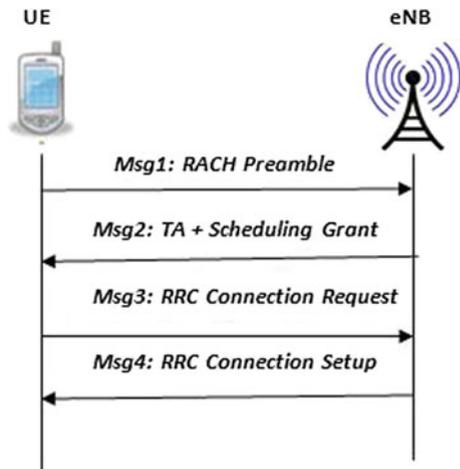
2 Machine Type Communication in LTE Networks

Machine-Type Communication (MTC) is the communication between machines without human intervention. This occurred because of the remarkable increase in the number of devices that require a connection wireless network, giving a large difference in the ratio of number devices to that of humans. This means that devices will communicate autonomously, complicating the direct control of devices by humans. MTC devices can communicate through a wired or wireless network [4, 5]. Existing cellular networks are supposed to support a wide range of MTC communications. This is due to their existing architecture and support roaming provided by cellular networks and next-generation LTE, the most suitable networks for cellular MTC. However, cellular networks are mainly intended for the Human–Human communication (H2H), on the other side, the MTC communication has various features such as the huge number of devices, small data loads and the way traffic (principally uplink). These previous characteristics of MTC communications can cause an immense problem of congestion in the radio access of a cellular network.

3 LTE Random Access Procedure

A crucial requirement for any cellular system is the ability for the device to request the establishment of connection. This is known by the “Random Access Procedure.” Random access (RA) is made not only during the initial access, but during the transition from inactive to active state and also after periods of inactivity in the uplink direction. The random access procedure is made of the following four steps (Fig. 2) [6, 7]:

Fig. 2 EPS security architecture



- Step 1: “Random access preamble transmission” (Msg1): The first step is based on the transmission of a random access preamble, allowing the Evolved Node B (eNodeB) of the cell to estimate the transmission time of the terminal. Synchronization is necessary as otherwise the device cannot transmit data in the uplink direction. Therefore, the terminal selects a preamble and outputs it on the Physical Random Access Channel (PRACH).
- Step 2: “Random access response” (Msg2): In the second step, the eNodeB sends a synchronization message to adjust the terminal emission timing, depending on the measurement performed in the first step. In addition to synchronizing the uplink, the eNodeB also allocates resources to the terminal for use in the third step of the procedure.
- Step 3: “RRC Connection Request (Msg3)” during this third step, the terminal transmits its identity to the network using the UL-SCH (Uplink Signaling Channel). The exact content of this signal depends on the state of the terminal, in particular if the terminal is already known by the network or not.
- Step 4: “RRC Connection Setup (Msg4)” The fourth and final step is based on the transmission of the network’s contention resolution message to the terminal on the DL-SCH (Downlink Signaling Channel). This step also resolves any contention caused by multiple devices trying to access the system using the same random access resource.

4 Impact of MTC on Random Access Channel

The RACH procedure was proposed long ago. We can find this procedure in the UMTS cellular technology, as in LTE. The paper [8] of 3GPP explains the approximation of collision probability of the RACH procedure. From this probability, the probability of transmission success and the probability of failure can be deduced. The RACH procedure approached the mechanism “Slotted ALOHA” because it is carried out within well-defined time slots: If a device tries to access the network in a slot time not specified for the RACH procedure, it will be prevented. Therefore, the calculation of the collision probability is similar to that of Slotted ALOHA mechanism. According to [9], the probability of collision is:

$$P_c = 1 - \exp\left(\frac{-\gamma}{L}\right) \quad (1)$$

With:

- γ random access intensity (number of random access test/s/cell)
- L Total number of RACH opportunities/s.

In [9], the author presents an analytical model to derive the probability of collision, the probability of success and the idle probability. In [10] the author analyzes the mathematical period of the RACH procedure, the probability of

success, failure, all based on the definition of the collision probability given by 3GPP, also this article summarizes an analysis of average delay.

MTC is characterized by a massive number of MTC devices in a cell. Therefore, some or all of these MTC terminals may attempt to access the base station simultaneously using an uncoordinated random access. In order to evaluate the impact of MTC in RACH, we choose three metrics already defined in 3GPP standards [11]: the probability of collision, the probability of success, and the access delay:

- Collision probability: the probability that two or more MTC devices realize a random access test using exactly the same preamble for the set period of time.
- Success Probability: the probability of successfully completing the RACH procedure respecting the maximum number of retransmissions.
- Access Delay: the average delay for each RACH procedure between the first access trial and the end of the RACH procedure for MTC devices that have successfully accessed the network.

Figure 3 illustrates respectively the collision probability (a) and the success probability (b) for different number of preamble sequences available according to the number of competing devices trying to access the same resources.

We observed in the two figures that the increase in MTC devices competing for access to the same resources in a cell causes an increase of packet collision rate in Fig. 3a, resulting a decrease in the probability of success (Fig. 3b). The increasing of the number of preambles translated better performance of the collision probability and success probability.

Figure 4 represents respectively the throughput (a) and the average delay (b) for a mobile terminal, according to the number of MTC devices competing for different numbers of preamble sequences available.

In Fig. 4a, when the number of MTC devices competing in a cell increases, the throughput increases, but from a certain number of terminals, the throughput decreases, this is due to the fact that when the number of MTC increases the collision rate increases as well (Fig. 3a). If the random access fails after a collision, the MTC terminal must wait for some time before starting a new random access,

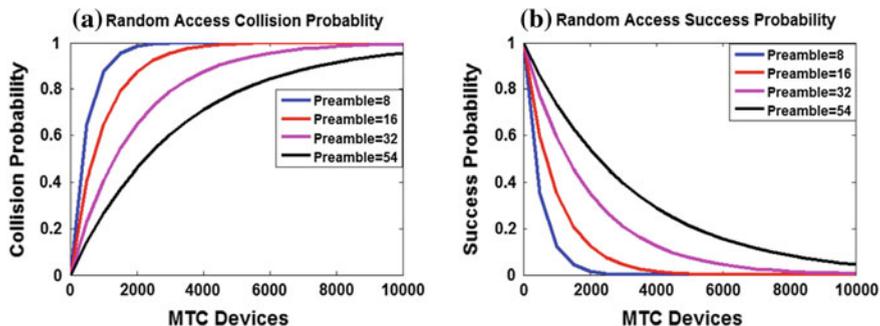


Fig. 3 Random access (RA) collision probability (a) and success probability (b)

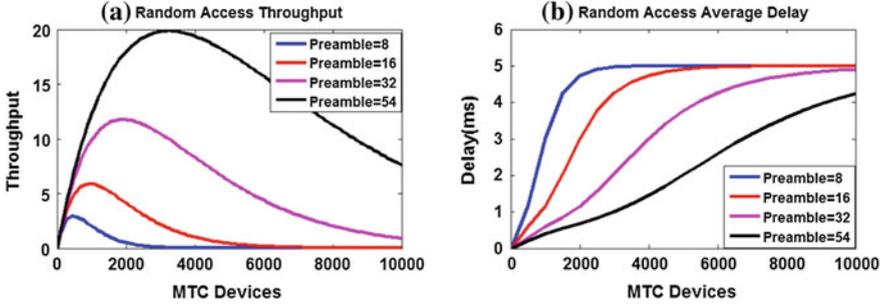


Fig. 4 Random access (RA) throughput (a) and average delay (b)

which introduces a latency of access to the channel and increases the size of the MTC queue requesting the channel and even causes an increase of the collision rate and the number of retransmission frames leading to a decrease in throughput.

In Fig. 4b, when the number of MTC devices in a cell increases, the access delay increases, when increasing the number of MTC rate of collisions increases, if the random access fails after collision, MTC terminal must wait for some time before starting a new random access, which introduces a large channel access latency (may be unacceptable for some real-time MTC) and increases the size of the MTC queue requesting the channel and causes the increase of the collision rate and the number of retransmission frames causing increasingly high-latency channel access. It is noted that increasing the number of preambles reflects the best performance of the access period.

5 Comparison Between M2M and H2H Traffics

The RACH throughput T , expressed by the arrival rate of RA attempts γ and the number of RACH opportunities S , is written as follows [12]:

$$T = \gamma \cdot \exp\left(\frac{-\gamma}{S}\right). \quad (2)$$

The RACH throughputs for H2H and M2M, using Poisson [13] as an arrival process, are expressed by the following equations:

$$T_{\text{H2H}} = \lambda_{\text{H2H}} \cdot \exp\left(-\frac{\lambda_{\text{H2H}}}{(N-m)}\right) \quad (3)$$

$$T_{\text{M2M}} = \lambda_{\text{M2M}} \cdot \exp\left(-\frac{\lambda_{\text{M2M}}}{m}\right) \quad (4)$$

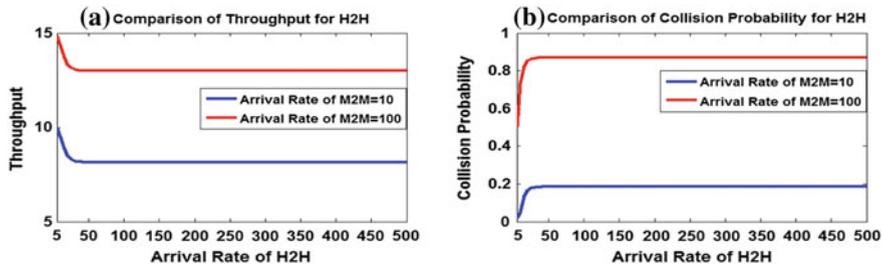


Fig. 5 Comparison of throughput (a) and collision probability (b) for H2H when $\lambda_{M2M} = 10, 100$

Therefore, the RACH throughput is given by:

$$T = \lambda_{H2H} \cdot \exp\left(-\frac{\lambda_{H2H}}{(N-m)}\right) + \lambda_{M2M} \cdot \exp\left(-\frac{\lambda_{M2M}}{m}\right). \quad (5)$$

In this section, the throughput and collision probability of M2M and H2H communication are compared for different parameters values. We consider that the eNB affects 54 preambles for User Equipment (UE), to use H2H traffics can use one of 5 preambles, whereas M2M traffics can use one of the other 49 preambles.

Figure 5 presents RACH throughput (a) and collision probability (b) for H2H UEs according to the variation of the arrival rate of RA attempts of H2H (λ_{H2H}) with the arrival rate of RA attempts of M2M UE's (λ_{M2M}) is fixed to 10 and 100.

It is shown in Fig. 5a that we have a better throughput when the arrival rate λ_{M2M} is 100 compared to 10. In Fig. 5b it can be seen that we have a high-collision probability with M2M arrival rate $\lambda_{M2M} = 100$.

Figure 6a illustrates the variation of the collision rate based on the number of M2M devices competing in a cell for a number of devices H2H set, the collision rate increases. As it is shown in this figure, the M2M users experience less number of RACH procedure in collision with $\lambda_{H2H} = 10$ compared to 100.

Figure 6b presents the RACH throughput for M2M according to the variation of RACH arrival rate λ_{M2M} with arrival rate $\lambda_{H2H} = 10, 100$. By increasing the arrival rate λ_{M2M} the throughput increases until 19 when $\lambda_{M2M} = 80$ in the case of arrival rate λ_{H2H} is 10.

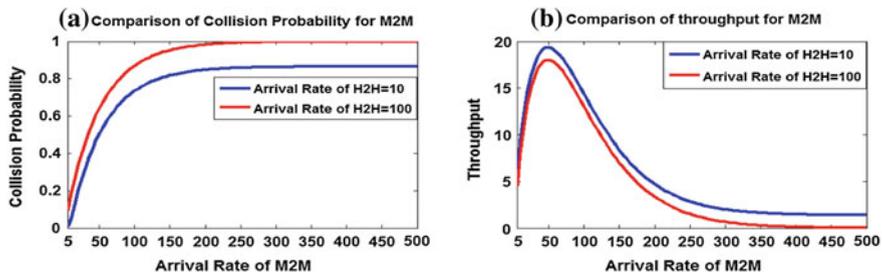


Fig. 6 Comparison of collision probability (a) and throughput (b) for M2M when $\lambda_{H2H} = 10, 100$

In Fig. 6b, when the number of M2M devices competing in a cell increases for a fixed number of H2H devices, the throughput increases, but from a given number of terminals, the throughput decreases, this is due to the fact that, when the number of M2M increases, the collision rate increases as in Fig. 6a. If the random access fails after collision M2M devices must wait for a certain period of time before starting a new random access, which introduces latency access to the additional channel.

6 Conclusion

M2M communications are characterized by high density of the machine by cell unlike H2H communications. In this article, we considered simultaneous access to the radio resources in a coexistence scenario M2M/H2H. First, we have formulated the resource sharing problem by MTC. Then we evaluated the impact of the introduction of M2M services on the performance of H2H applications in terms of collision and throughput. We can conclude that the system performance degrades when the number of M2M devices increases, and these performances depend on the resource allocation between M2M and H2H.

References

1. Kim, J., Lee, J., Kim, J., Yun, J.: M2M service platforms: survey, issues, and enabling technologies. *IEEE Commun. Surv. Tutorials* **16**(1), 61–76 (2014)
2. Bouguen, Y., Hardouin, E., Xavier Wolff, F.: *LTE et les réseaux 4G* (Chap. 19). Eyrolles. ISBN: 978-2-212-12990-8 (2012)
3. Fritze, G.: *SAE: The Core Network for LTE*. Ericsson (2012)
4. Kunz, A., Kim, L., Kim, H., Syed, S.: Husain: Machine Type Communications in 3GPP: From Release 10 to Release 12 (2012)
5. Taleb, T., Kunz, A.: Machine type communications in 3GPP networks: potential, challenges and solutions. *Commun. Mag.* **50**(3), 178–184 (2012)
6. Sesia, S., Toufik, I., Baker, M.: *LTE-The UMTS Long Term Evolution: From Theory to Practice*, 2nd edn. Wiley (2011)
7. Kouzayha, N., Taher, N.C., Ghamri-Doudane, Y.: Towards a better support of machine type communication in LTE-networks: analysis of random access mechanisms. In: *2nd International Conference on Advances in Biomedical Engineering* (2013)
8. 3GPP R1-061369.: LTE random-access capacity and collision probability. TSG-RAN WG1 #45 (2006)
9. Cheng, R.-G., Wei, C.-H., Tsao, S.-L., Ren, F.-C.: RACH collision probability for machine-type communications. In: *2012 IEEE 75th Conference on Vehicular Technology (VTC Spring)*, pp 1–5 (2012)
10. Tyagi, R.R., Lee, K.-D., Aurzada, F., Kim, S., Reisslein, M.: Efficient delivery of frequent small data for U-healthcare applications over LTE-advanced networks. *Mobile Health'12* (2012)
11. 3GPP TR 37.868 V11.2.0 (2011–09).: Study on RAN Improvements for Machine-type Communications (2011)

12. Lee, K.-D., Kim, S., Yi, B.: Throughput comparison of random access methods for M2M service over LTE networks. In: 2011 IEEE on GLOBECOM Workshops (GC Wkshps), pp. 373–377 (2011)
13. Luo, W., Ephremides, A.: Stability of N interacting queues in random-access systems. *IEEE Trans. Info. Theory* **45**(5), 1579–1587 (1999)

Predictive Modeling of Students Performance Through the Enhanced Decision Tree

Subitha Sivakumar and Rajalakshmi Selvaraj

Abstract Prognostic of student performance is one of the major issues in many institutions. The student's performance is based on many factors such as internal examinations, grade obtained in university examination, Academic, Extra Curricular and Co-Curricular activities and also concern with their activities in learning environment. Student performance prediction is used to model the students into any one of the four categories as excellent, good, average, and poor performance student. The category selection was determined using supervised classifiers. Academic institution can easily able to identify any academic problems and the corresponding solutions for their students through this predictive modeling approach. We have collected real world dataset related to student's academic performance from leading academic institution in India which consists of details about the students such as CGPA, Lab performance, History of arrears and so on. We have applied various supervised classifiers such as DT, SVM, KNN, NB, NN and Improved DT on student's academic performance dataset. Besides, the conventional decision tree is further improved by the use of normalized factor and Association Function. By comparing the accuracy of these two methods, the best result is exposed for Improved Decision Tree than all other supervised classifiers in the literature.

Keywords Data mining · Educational data mining · Supervised classifiers
Improved decision trees

S. Sivakumar (✉)

Faculty of Computing, Botho University, Gaborone, Botswana

e-mail: subitha.sivakumar@bothouniversity.ac.bw

S. Sivakumar · R. Selvaraj (✉)

Department of Information Systems, BIUST, Gaborone, Botswana

e-mail: selvarajr@biust.ac.bw

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication*

and Computing, Lecture Notes in Electrical Engineering 443,

https://doi.org/10.1007/978-981-10-4765-7_3

1 Introduction

Data Mining (DM) indicates to the nontrivial removal of indirect, earlier doubtful, possibly valuable data in databases. Both infer either filtering through a lot of material or clever investigating the material to precisely pinpoint where the qualities dwell. Data mining uses the wide availability of sizeable amounts of data, where those can be used as beneficial data for research. This provides a way toward finding canny, fascinating and novel examples, and enlightening reasonable, and prescient models from extensive volumes of information.

Educational Data Mining (EDM) is a separate research field started to mature a few years ago and to analysis the related student data from the dataset. EDM will provide an opportunity to collect, analysis, predict and forecast the student's performance from the student's academic performance dataset. It is used to derive new discoveries and hypotheses about how students learn. EDM is an inderdispilanry research field that uses the concept from Data Mining, Machine Learning and Statistics.

Student's Performance has been measured from their activities since when they had admitted to the college. Students academic performance datasets are collected for this study of corresponding student such as CGPA, attendance, lab performance, assignment work, internal marks, seminar performance, grade level, extracurricular, co-curricular activities, and so on. Based on these details, predictive modeling of student is carried out using supervised classifiers such as KNN, NB, NN, SVM, DT and Improved DT. We have used percentage split method as a test option. A further section of this paper contains detailed implementation of traditional and improved decision tree methods. Results reveal that, the proposed improved decision tree model is more accurate method for student's performance prediction.

2 Related Works

This section reviews what have been established in the literature about Educational Data mining, supervised classifiers for predicting student's performance, Predictive modeling of Students Performance and the interpretation of the findings are also covered in this section.

DM are classified according to the kind of knowledge mined from Correlation Analysis, Association Rule Mining, Classification, Clustering and Prediction. A detailed description about the applications of DM techniques for mining Student Retention and Students Academic Performance can be found in [1–4].

In recent years, there has been an increased focus on the role of EDM towards predicting the academic performance for students. Predicting student performance has been presented by the authors [5–9] considers Cumulative Grade Point Average (CGPA), Grade Point Average (GPA), student's score and grade as a predictive or dependent parameter and prediction of student's performance as an independent

parameter for predictive modeling of student's performance in particular courses or subjects.

In [10] the author used supervised classifiers for predicting student performance. In [11] the authors have proposed a Fuzzy K-Means Cluster approach to predict the student's academic performance. Various Data Mining techniques are applied to predict student academic performance such as tutoring, attitude, academic improvements [12–14].

In [15] the authors have concentrated on the use of information digging in educational databases for anticipating scholastic patterns and examples. In [16] the authors have developed a prediction model for predicting academic performance for first year students using various data mining techniques. In [17] the author has applied different supervised learning algorithms such as Decision Tree, K-NN Classifier, Support Vector Machine, Naïve Bayes and neural networks for student's performance predictions in a computer science curriculum. In [18] the author has discussed the several DM procedures like Association Rule Mining, Classification, prediction, correlation analysis and clustering techniques for discovering knowledge from educational data. In [19] the authors have used student database to derive the various performance indications towards better prediction of student's academic performance.

In [20] the authors have applied various Classification Algorithms to predict the students into the category of performer or underperformer using machine learning approaches. A Naïve Bayes Classifier to develop a predictive model in education system which uses the prior and posterior probability [21]. In [22] the author has discussed about the mining of educational data using classification algorithms to decrease dropout. In [23, 24] the authors have proposed the method of accurate rule sets without global optimization and different DM algorithms to classify students into various categories based on the students' academic performance.

3 Data Preparation and Discretization

The data set had been collected from leading academic institution in India. The information contains student basic details, academic details, and other related information. The value for the attributes such as SSLC MARK, HSC Mark and CGPA were collected through students profile database.

If the student could not pass all the subjects in the first attempt itself then it is considered as history of arrears. Other than academic performance the student active participation in intellectual skills enrichment activities like extra curriculum and co-Curriculum activities are also considered for the student's performance modeling. Time allotted for self study, attendance percentage, Submission of assignment, Internal mark, Seminar performance, Lab performance, Student participation in outside activities, Homework values are collected from the student database.

Table 1 Students database-attributes and their values

Attributes	Description	Values
MARK	Cumulative mark	{excellent, good, average, poor}
HOA	History of arrears	{no, average, poor, very poor}
ATTENDANCE	Attendance	{good, average, low}
SOA	Submission of assignment	{good, average, low}
IM	Internal mark	{good, average, low}
SP	Seminar performance	{good, average, low}
LP	Lab performance	{good, average, low}
STUDENTS PERFORMANCE	Class label	{excellent, good, average, low}

3.1 Data Discretization

Data discretization is the process of discretizing the attributes by the use of their values which will speed up the learning process. Student's database having attributes and their values are shown in Table 1 which are holding all necessary details for predicting student's performance.

4 Proposed Methods

Given a collection of records and features the process of classifying each object into one of the several pre-defined classes based on the value of record attribute is called as classification which aims to forecast a certain outcome. In classifier learning, will take place only on historical data and the outcome of learning is called as data mining model. The performance of model is evaluated by giving some unseen instance as a test data. The model usually takes any kind of data as an input and provides a prediction. The forecasting accuracy determines how "good" the generated model is.

KNN, NB, NN, SVM, DT and Improved decision tree are examined and studied. First we modeled the student's performance using historical training data and the performance of unseen record was predicted based on the generated model.

4.1 Traditional Decision Tree

Decision tree can be built from database attributes with the tree and set of rules. The base of tree structure is used to represent the decision tree structure. It splits the data set into small subsets meanwhile maintain the association between the subsets. Decision tree has the root and the leaf nodes. Datasets are mentioned with these nodes, the leaf denotes the classification, and the top most decision node called as a root node.

Split Point Evaluation Measures

Spontaneous to choose a split point which provides finest division and/or separation among diverse class marks.

Entropy

Entropy is a measure which is used to determine the amount of uncertainty in a system. If many of the instances having same class labels then a partition has least point or low order entropy and if there is a mixed class labels among instances then partition have higher or more disorder entropy.

Step 1

The entropy for set named D are characterized as:

$$H(D) = - \sum_{i=1}^k P\left(\frac{c_i}{D}\right) \log_2 P\left(\frac{c_i}{D}\right)$$

where, $P\left(\frac{c_i}{D}\right)$ = class possibility c_i in D , K = number of classes.

Step 2

Partitions D into D_Y and D_N using split point. The *split entropy* is defined for each resulting partition calculated as

$$H(D_Y, D_N) = \frac{n_Y}{n} H(D_Y) + \frac{n_N}{n} H(D_N)$$

where $n = |D|$ refers to number of points presented in D . $n_Y = |D_Y|$ and $n_N = |D_N|$ are the number of points in D_Y and D_N .

When the classes is mixed, then each will have the same probability $P\left(\frac{c_i}{D}\right) = \frac{1}{k}$, then the entropy has the highest value, $H(D) = \log_2 k$.

Step 3

The information gain for a given split point is calculated to ensure that whether the split point results to reduced over all entropy.

$$\text{Gain}(D, D_Y, D_N) = H(D) - H(D_Y, D_N)$$

In general we can get better split point for higher information gain value. Thereby, we can choose the split point which gives the highest information gain.

Step 4

Construct decision tree by considering the attribute which has highest information as a root node and which attribute need to be tested at node N is further determined by splitting criterion with the help of partition the instances into individual classes.

Table 2 Attributes with their corresponding entropy and gain values

Attribute name	Entropy value	Gain value
MARK	1.1783	0.7383
History of arrears (HOA)	0.7440	1.1726
ATTENDANCE	1.4556	0.4610
Submission of assignment (SOA)	1.5906	0.3260
Internal mark (IM)	1.3746	0.5420
Seminar performance (SP)	1.6249	0.2917
Lab performance (LP)	1.5998	0.3168

Step 5

If all tuples belong to the same class, then the root node attribute is get terminated and it is considered to dismiss the condition.

Step 6

When a dissimilar class label is there, the tree will further be expanded with the following node as trait to higher information gain.

Step 7

The previous stages will be repeated until the ending condition holds.

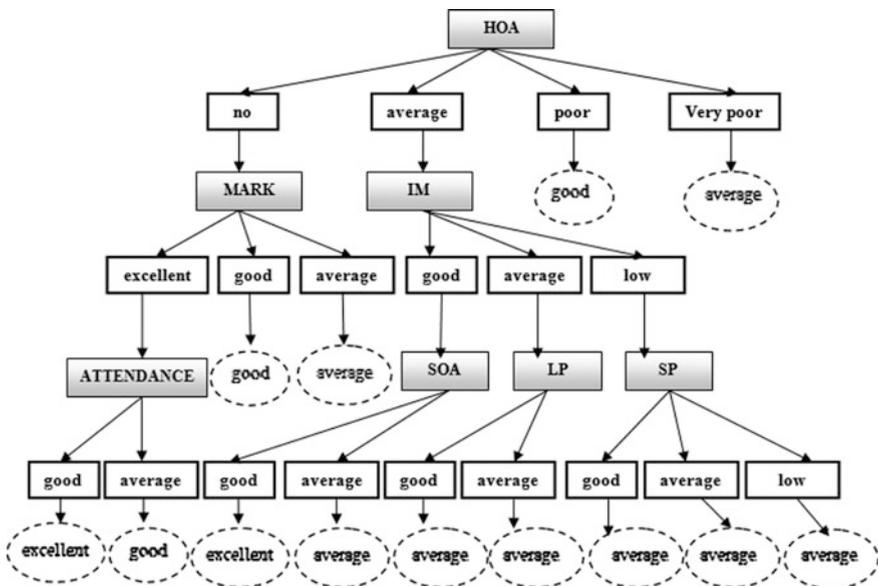


Fig. 1 Traditional decision tree structure

Computational Procedure for Traditional Decision Tree

We have the computed the entropy and information gain for all predictive attributes (MARK) and results are shown in Table 2. This gain value is the main key for arrange the attributes in their corresponding order. And also it is the dynamic value which means if the number of records is increased or decreased the gain value of attribute may be change which reflected on tree structure and the new rules will be generated.

Gained value showed as descending order, where the root node for the tree is formed by the attributes having the uppermost gain value are comes with their gain value order. Step 4, 5, 6, 7 had computed and the tree structure was shown in Fig. 1.

4.2 Improved Decision Tree

Various strategies are available to built decision tree and it is one of the simplest method so it can be used by wide range of application. Attribute selection is the main process in decision tree and that may hold many values. Improved decision tree used to get more effective accuracy results from these attribute values. Because the improved decision tree can significantly uses Association Function (AF) and Normalized Factor (NF) to further improve the model generation process (Fig. 2).

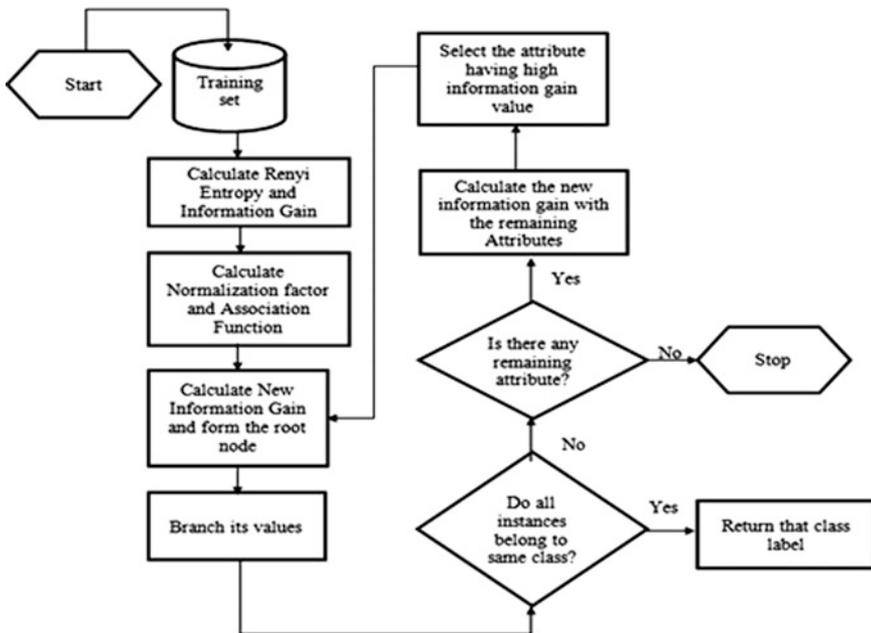


Fig. 2 Data flow procedure for improved decision tree

Step 1

As a first step split point partitions D into D_Y and D_N . Split entropy shows the weighted entropy for each of the subsequent segments.

$$H(D) = - \sum_{i=1}^k P\left(\frac{c_i}{D}\right) \log_2 P\left(\frac{c_i}{D}\right)$$

$$H(D_Y, D_N) = \frac{n_Y}{n} H(D_Y) + \frac{n_N}{n} H(D_N),$$

where $n = |D|$ is the number of points in D , and $n_Y = |D_Y|$ and $n_N = |D_N|$ are the number of points in D_Y and D_N .

The entropy is zero for pure region in which all the sample are belonging to the same class and for a mixed up positive and negative example every one appears with equal probability $P\left(\frac{c_i}{D}\right) = \frac{1}{k}$, then the entropy has the highest value, $H(D) = \log_2 k$.

Step 2

Compute information gain of each and every attribute with the following formula

$$\text{Gain}(D, D_Y, D_N) = H(D) - H(D_Y, D_N)$$

Ranking individually the split point to find out the split point that gives the highest information gain.

Step 3

Compute Association Function (AF): Assume A as attribute in Dataset D , C as class attributes for D . then relation degree function among A , C can calculate by

$$AF(D) = \frac{1}{n} * \sum_{i=1}^n |X_{i1} - X_{i2}|$$

where, $X_{ij} \Rightarrow$ attribute A in D has i th value. C (category attribute) \Rightarrow sample number in j th value and $n \Rightarrow$ number of value takes from attributes A .

Step 4

Compute Normalization Factor: Consider m attributes, for all attributes the relation degree function values are $AF(1)$, $AF(2)$... $AF(m)$ individually.

$$V(k) = \frac{A(k)}{A(1) + A(2) + A(3) + \dots + A(m)} \quad \text{as } 1 < k \leq m$$

Step 5

Calculate the new information gain attributes to build Decision Tree (DT)

$$\text{New gain}(D, D_Y, D_N) = \text{Gain}(D, D_Y, D_N) * V(k)$$

Step 6

Highest Information Gained attribute will form the root node in the tree.

Step 7

For root node, terminating condition done, when all samples are in the same class.

Step 8

Tree is further branched for different class label with higher information gain value.

Step 9

The previous stages will be repeated until the ending condition holds.

Computational Procedure for Improved Decision Tree

Step 1

Entropy calculation for class label

$$H(D) = - \sum_{i=1}^k P\left(\frac{c_i}{D}\right) \log_2 P\left(\frac{c_i}{D}\right)$$

$$\begin{aligned} \text{Info}(D) = H(D) &= - \sum_{i=1}^k P\left(\frac{c_i}{D}\right) \log_2 P\left(\frac{c_i}{D}\right) \\ &= \left(-\frac{34}{156} \log_2 \frac{34}{156}\right) + \left(-\frac{37}{156} \log_2 \frac{37}{156}\right) + \left(-\frac{61}{156} \log_2 \frac{61}{156}\right) + \left(-\frac{24}{156} \log_2 \frac{24}{156}\right) \\ &= 1.9166 \end{aligned}$$

$$\text{Info}(D) = H(D) = 1.9166$$

Entropy calculation for Attribute MARK

$$H(D_Y, D_N) = \frac{n_Y}{n} H(D_Y) + \frac{n_N}{n} H(D_N)$$

(Refer Table 1 for values)

$$\begin{aligned}
\text{Info}_{\text{MARK}}(DY, DN) &= H_{\text{MARK}}(DY, DN) \\
&= \frac{37}{156} \left(\left(-\frac{34}{37} \log_2 \frac{34}{37} \right) \left(-\frac{3}{37} \log_2 \frac{3}{37} \right) \left(-\frac{0}{37} \log_2 \frac{0}{37} \right) \left(-\frac{0}{37} \log_2 \frac{0}{37} \right) \right) \\
&\quad + \frac{108}{156} \left(\left(-\frac{0}{108} \log_2 \frac{0}{108} \right) \left(-\frac{34}{108} \log_2 \frac{34}{108} \right) \left(-\frac{55}{108} \log_2 \frac{55}{108} \right) \left(-\frac{19}{108} \log_2 \frac{19}{108} \right) \right) \\
&\quad + \frac{11}{156} \left(\left(-\frac{0}{11} \log_2 \frac{0}{11} \right) \left(-\frac{0}{11} \log_2 \frac{0}{11} \right) \left(-\frac{6}{11} \log_2 \frac{6}{11} \right) \left(-\frac{5}{11} \log_2 \frac{5}{11} \right) \right) \\
&= 1.1783 \\
\text{Info}_{\text{MARK}}(DY, DN) &= H_{\text{MARK}}(DY, DN) = 1.1783
\end{aligned}$$

Step 2

Gain calculation for Attribute MARK

$$\text{Gain}(D, D_Y, D_N) = H(D) - H(D_Y, D_N)$$

$$\text{Gain}(\text{STUDENT PERFORMANCE, MARK}) = 1.9166 - 1.1783$$

$$\text{Gain}(\text{STUDENT PERFORMANCE, MARK}) = 0.7383$$

Step 3

Association Function calculation for Attribute MARK

$$\text{AF}(D) = \frac{1}{n} * \sum_{i=1}^n |X_{i1} - X_{i2}|$$

Number of values in MARK attribute = 3, $n = 3$.

(Refer Table 1 for values)

$$\begin{aligned}
\text{AF}_{\text{MARK}}(D) &= \frac{1}{3} \left[\left(\left| \frac{34}{37} - \frac{3}{37} \right| + \left| \frac{3}{37} - \frac{34}{37} \right| + \left| \frac{0}{37} - \frac{37}{37} \right| + \left| \frac{0}{37} - \frac{37}{37} \right| \right) \right. \\
&\quad + \left(\left| \frac{0}{108} - \frac{108}{108} \right| + \left| \frac{34}{108} - \frac{74}{108} \right| + \left| \frac{55}{108} - \frac{53}{108} \right| + \left| \frac{19}{108} - \frac{89}{108} \right| \right) \\
&\quad \left. + \left(\left| \frac{0}{11} - \frac{11}{11} \right| + \left| \frac{0}{11} - \frac{11}{11} \right| + \left| \frac{6}{11} - \frac{5}{11} \right| + \left| \frac{5}{11} - \frac{6}{11} \right| \right) \right]
\end{aligned}$$

$$\text{AF}_{\text{MARK}}(D) = 2.6376$$

Association function value for all other attributes are displayed in Table 3.

Table 3 Attributes and their association function value

Attribute name	Association function value
MARK	2.6376
History of arrears (HOA)	3.4117
ATTENDANCE	2.3334
Submission of assignment (SOA)	2.2222
Internal mark (IM)	2.2264
Seminar performance (SP)	2.1222
Lab performance (LP)	2.6857

Step 4

Normalization Factor calculation for Attribute MARK

$$V(k) = \frac{A(k)}{A(1) + A(2) + A(3) + \dots + A(m)}$$

$$A(\text{MARK}) + A(\text{HOA}) + A(\text{ATTENDANCE}) + A(\text{SOA}) + A(\text{IM}) + A(\text{SP}) + A(\text{LP}) = 17.6392$$

$$V(\text{MARK}) = \frac{A(\text{MARK})}{A(\text{MARK}) + A(\text{HOA}) + A(\text{ATTENDANCE}) + A(\text{SOA}) + A(\text{IM}) + A(\text{SP}) + A(\text{LP})}$$

$$V(\text{MARK}) = 0.1495$$

Normalization Factor value for all other attributes values are displayed in Table 4.

Step 5

New Gain calculation for Attribute MARK

$$\text{New gain}(D, D_Y, D_N) = \text{Gain}(D, D_Y, D_N) * V(k)$$

Table 4 Attributes and their normalization factor value

Attribute name	Normalization factor value
MARK	0.1495
History of arrears (HOA)	0.1934
ATTENDANCE	0.1323
Submission of assignment (SOA)	0.1260
Internal mark (IM)	0.1262
Seminar performance (SP)	0.1203
Lab performance (LP)	0.1523

Table 5 Attributes and their new gain value

Attribute name	New gain values
MARK	0.1104
History of arrears (HOA)	0.2268
ATTENDANCE	0.0610
Submission of assignment (SOA)	0.0411
Internal mark (IM)	0.0684
Seminar performance (SP)	0.0351
Lab performance (LP)	0.0482

$$\text{New Gain}(\text{STUDENT PERFORMANCE, MARK}) = 0.7383 * 0.1495$$

$$\text{New Gain}(\text{STUDENT PERFORMANCE, MARK}) = 0.1104$$

From this calculation method we find out Normalization Factor value for all other attributes. Those values are displayed in Table 5.

New gain values are assigned with the descending order. Based on descending ordered new gain value we can calculate root node. Remaining node comes under with respect their values. Step 4, 5, 6, 7 had computed and the tree structure was made is shown in Fig. 3.

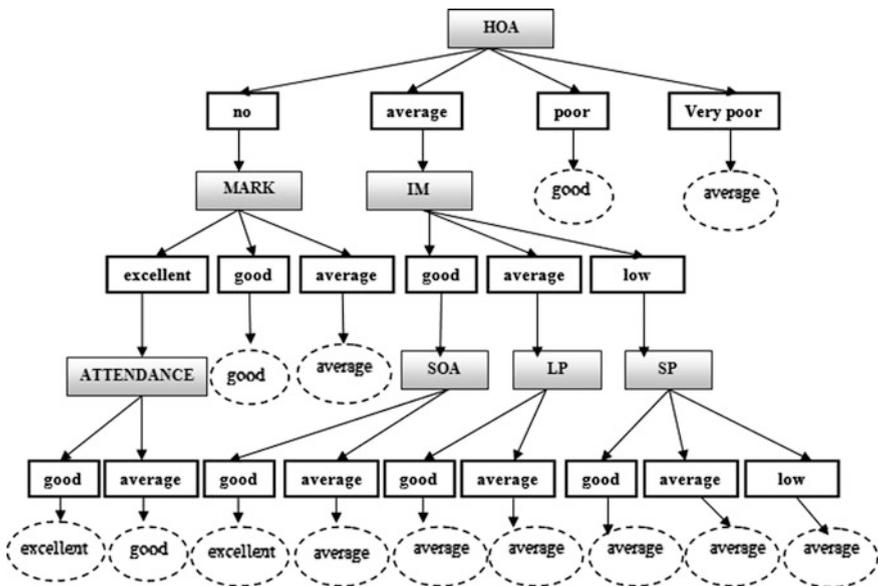


Fig. 3 Improved decision tree structure

5 Result and Discussion

Historical student’s data were collected for predicting the student’s performance. Most relevant attributes are successfully executing data mining process. The performance of KNN, SVM, DT, NN, NB and Improved DT were evaluated. Each and Every classifiers were produced the different level of accuracy with their

Table 6 Performance of classifiers with percentage split

Classifiers	Class	TP rate	FP rate	Precision	Recall	F measure	ROC area
KNN	Excellent	1.000	0.026	0.930	1.000	0.964	0.993
	Good	0.814	0.006	0.972	0.814	0.886	0.968
	Average	0.960	0.030	0.947	0.960	0.954	0.972
	Low	0.973	0.018	0.923	0.973	0.947	0.991
NB	Excellent	1.000	0.013	0.964	1.000	0.981	0.993
	Good	0.884	0.030	0.884	0.884	0.884	0.952
	Average	0.960	0.015	0.973	0.960	0.966	0.979
	Low	0.892	0.018	0.917	0.892	0.904	0.976
SVM	Excellent	1.000	0.013	0.964	1.000	0.981	0.994
	Good	0.953	0.006	0.976	0.953	0.965	0.956
	Average	0.973	0.008	0.986	0.973	0.980	0.983
	Low	1.000	1.000	1.000	1.000	1.000	1.000
DT	Excellent	1.000	0.039	0.898	1.000	0.946	0.987
	Good	0.907	0.000	1.000	0.907	0.951	0.982
	Average	0.973	0.000	1.000	0.973	0.986	0.988
	Low	1.000	1.000	1.000	1.000	1.000	1.000
Improved DT	Excellent	1.000	0.013	0.964	1.000	0.981	0.989
	Good	0.977	0.006	0.977	0.977	0.977	0.977
	Average	0.973	0.000	1.000	0.973	0.986	0.975
	Low	1.000	0.000	1.000	1.000	1.000	1.000

Table 7 Classifiers accuracy with percentage split

Classifier	Accuracy
KNN	94.23
NB	94.23
SVM	98.07
DT	97.11
Improved DT	98.56

Table 8 Time taken to build the model for each classifier

Classifier	Time taken (s)
KNN	0
NB	0.01
SVM	0.34
DT	0.03
Improved DT	0.03

corresponding time taken values. These results were compared to select the best classification method. The comparison result is shown in Table 6.

Predictive modeling of student's performance using traditional classifiers as well as improved DT is represented in this. How well the generated prediction model by different supervised classifiers can predict a new data is justified by calculating the accuracy generated by each and every individual classifier by considering the percentage split as a test method. To examine the generated model performance the entire dataset has spitted into two training: test set with the ratio of 75:25. Both conventional classifiers and improved decision tree is applied on training data and model is generated from historical data and performance of generated model is tested with how well the model perform for unseen data. The performance evaluation parameters such as TP Rate, FP Rate, Precision, Recall, F Measure and ROC Area are represented very clearly which contains the information about detection performance of each classifier (Tables 7, 8 and 9).

Table 9 Actual and predicted value for DT and improved DT

S. No.	Actual value	Predicted value	
		Traditional decision tree	Improved decision tree
157	Low	Low	Low
158	Excellent	Excellent	Excellent
159	Excellent	Good	Excellent
160	Excellent	Excellent	Excellent
161	Average	Average	Average
162	Low	Low	Low
163	Excellent	Excellent	Excellent
164	Low	Low	Low
165	Excellent	Excellent	Excellent
166	Low	Low	Low
167	Excellent	Excellent	Average
168	Excellent	Excellent	Excellent
169	Excellent	Excellent	Excellent
170	Average	Excellent	Average
171	Average	Average	Average
172	Excellent	Excellent	Excellent
173	Average	Excellent	Average
174	Good	Good	Good
175	Excellent	Excellent	Average
176	Excellent	Excellent	Excellent
177	Average	Average	Average
178	Low	Low	Low
179	Excellent	Excellent	Excellent
180	Average	Excellent	Average
181	Average	Excellent	Average
182	Good	Good	Good

6 Conclusion

Predictive modeling of students academic performance using both traditional supervised classifiers and improved decision was implemented using the data driven approach through parameters involved in students' success is determined and predicted. The detection performance of different supervised classifiers such as DT, NB, NN, KNN and Improved DT has been evaluated on student's academic performance dataset. The examined result verifies the planned Improved Decision Tree perform better than available classifiers in term of accuracy, sensitivity, specificity and time taken to build the model than all other classifiers. The proposed model Identifying low performance in students through generating production rules by improved decision tree model and filters the important illustrative rules, to direct students' low performance, to assist poor and average category students to improve their academic performance and results. The problem that we faced during the implementation of the proposed method is the availability of in adequate data and it will be overcome in future by collecting huge volume of data from different universities in order to further improve the detection performance of the prediction model.

References

1. Aziz, A.A., Ismail, N.H., Ahmad, F.: Mining students academic performance. *J. Theor. Appl. Inf. Technol.* **53**(3), 485–495 (2013)
2. Baker, R.S.J.: Data mining for education. In: *Advantages Relative to Traditional Educational Research Paradigms* (2010)
3. Sachin, R.B., Vijay, M.S.: A survey and future vision of data mining in educational field. In: *2012 Second International Conference on Advanced Computer Engineering & Communication Technologies*, pp. 96–100 (2012)
4. Zhang, Y., Oussena, S., Clark, T., Kim, H.: Use data mining to improve student retention in higher education—A case study. In: *ICEIS—12th International Conference on Enterprise Information Systems 2010* (2010)
5. Lye, C.T., Ng, L.N., Hassan, M.D., Goh, W.W., Law, C.Y., Ismail, N.: Predicting pre-university student's mathematics achievement. *Procedia Soc. Behav. Sci.* **8**, 299–306 (2010)
6. Mohsin, M.F.M., Wahab, M.H.A., Zaiyadi, M.F., Hibadullah, C.F.: An investigation into influence factor of student programming grade using association rule mining. *Int. J. Adv. Inf. Sci. Serv. Sci.* **2**(2), 19–27 (2010)
7. Norwawi, N.M., Abdusalam, S.F., Hibadullah, C.F., Shuaibu, B.M.: Classification of students performance in computer programming course according to learning style. In: *2nd Conference on Data Mining and Optimization*, pp. 37–41 (2009)
8. Othman, H., Nopiah, Z.M., Asshaari, I., Razali, N., Osman, M.H., Ramli, N.: A Comparative Study of Engineering Students on Their Pre-University Results with their First Year Performance at Fkab, UKM. *Seminar Pendidikan Kejuruteraan dan Alam Bina (PeKA'09)* (2009)
9. Wook, M., Yahaya, Y.H., Wahab, N., Isa, M.R.M., Awang, N.F., Seong, H.Y.: Predicting NDUM student's academic performance using data mining techniques. In: *2009*

- Second International Conference on Computer and Electrical Engineering, pp. 357–361 (2009)
10. Kabakchieva, D.: Predicting student performance by using data mining methods for classification. *Cybern. Inf. Technol.* **13**(1), 61–72 (2013)
 11. Prakash, S., Ramaswami, K.S., Post, C.A.: Fuzzy K-means cluster validation for institutional quality assessment. In: International Conference on Communication and Computational Intelligence (INCOCCI), pp. 628–635. Prediction of students' academic performance 6425 (2010)
 12. Golding, P., Shaw, F.L., Tennant, V.: Effects of peer tutoring, attitude and personality on academic performance of first year introductory programming students. In: 36th ASEE/IEEE Frontiers in Education Conference, pp. 7–12 (2006)
 13. Huang, S., Fang, N.: Work in progress—prediction of students' academic performance in an introductory engineering course. In: 41st ASEE/IEEE Frontiers in Education Conference, pp. 11–13 (2011)
 14. Sembiring, S., Zarlis, M., Hartama, D., Wani, E.: Prediction of student academic performance by an application of data mining techniques. In: 2011 International Conference on Management and Artificial Intelligence, vol. 6, pp. 110–114 (2011)
 15. Parack, S., Zahid, Z., Merchant, F.: Application of data mining in educational databases for predicting academic trends and patterns. In: 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), pp. 1–4 (2012)
 16. García, E.P.I., Mora, P.M.: Model prediction of academic performance for first year students. In: 2011 10th Mexican International Conference on Artificial Intelligence, pp. 169–174 (2011)
 17. Chamillard, A.T.: Using student performance predictions in a computer science curriculum, *ACM SIGCSE Bull.* **38**(3), 260 (2006)
 18. Larose, D.T.: *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley, Hoboken (2005)
 19. Shana, J., Venkatachalam, T.: Identifying key performance indicators and predicting the result from student data. *Int. J. Comput. Appl.* **25**(9), 45–48 (2011)
 20. Kumar, U., Pal, P.S.: Data mining: a prediction of performer or underperformer using classification. *Int. J. Comput. Sci. Inf. Technol. (IJCSIT)* **2**(2), 686–690 (2011)
 21. Sharma, M.: Development of predictive model in education system: using Naïve Bayes classifier. In: International Conference and Workshop on Emerging Trends in Technology (ICWET 2011) TCET, Mumbai, India, (Icwet), pp. 185–186 (2011)
 22. Pal, S.: Mining educational data using classification to decrease dropout rate of students. *Int. J. Multi. Sci. Eng.* **3**(5), 35–39 (2012)
 23. Frank, E., Witten, I.H.: Generating accurate rule sets without global optimization, In: *Proceedings of the 15th International Conference on Machine Learning*
 24. Romero, C., Ventura, S., Espejo, P.G., Hervás, C.: Data mining algorithms to classify students. In: *The 1st International Conference on Educational Data Mining Montréal, Québec, Canada*, pp. 8–17 (2008)

A New Pulse Width Modulation Technique with Hybrid Carrier Arrangement for Multilevel Inverter Topology

N. Prabaharan, V. Arun, K. Palanisamy and P. Sanjeevikumar

Abstract This article introduces a new reduced switch multilevel inverter topology that operates in asymmetric state. The advantage of proposed multilevel inverter is utilizing the lower device count when compared to traditional configuration. This paper depicts a hybrid pulse width modulation which includes three different carriers and unipolar references. The three different carriers are saw-tooth, triangular, inverted unipolar sine which is arranged in phase disposition technique. The effectiveness of proposed hybrid pulse width modulation (HPWM) technique is tested with a new reduced switch topology using MATLAB/Simulink. The different parameters are calculated with various modulation indices.

Keywords Cascaded H-bridge · PWM · Multilevel inverter · Reduced switch topology · Hybrid carriers

N. Prabaharan (✉) · K. Palanisamy
School of Electrical Engineering, VIT University, Vellore, Tamil Nadu, India
e-mail: prabaharan.nataraj@gmail.com

K. Palanisamy
e-mail: kpalanisamy@vit.ac.in

V. Arun
Department of Electrical and Electronics Engineering, Arunai Engineering College,
Thiruvannamalai, Tamil Nadu, India
e-mail: varunpse@yahoo.com

P. Sanjeevikumar
Department of Electrical and Electronics Engineering, University of Johannesburg,
Auckland Park, Johannesburg, South Africa
e-mail: sanjeevi_12@yahoo.co.in

1 Introduction

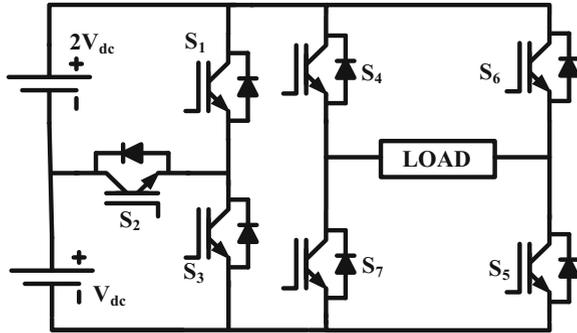
Past decades, the multilevel inverter (MLI) concept has been renowned as an imperative and substantial choice of source in high power application [1]. Utilization of multilevel inverters in high power industries has many advantages such as producing good quality of waveform with minimum distortions, reducing passive filter size, lower switching stress and switching losses [2]. A detailed investigation of MLI, modulation techniques, controls and applications were presented in [1]. Diode Clamped, Flying Capacitor, and Cascaded H-Bridge multilevel inverters are the classical types of MLI and which are utilized in high power application [3]. Increasing the component count is the major concern while the topology is extended to another level [4]. The multilevel inverter topology is possible to interface with DVR by replacing the tradition two-level inverter [5]. Nowadays, many topologies are presented with optimum number of switches [2, 6–10]. Asymmetric type of multilevel inverter produces good quality of waveforms with low distortion when compared to symmetric type of multilevel inverters [11]. In [12], a topology is used for grid connected PV application. In [13], single DC source cascade multilevel inverter is employed. Hybrid topologies are very popular for medium voltage drive application and renewable energy based grid connected application [14].

This article presents a new MLI topology which operates at asymmetric condition for increasing the output voltage levels and reducing the count of utilizing switches. Also, the HPWM is implemented to generate the pulses of switches in proposed MLI. In hybrid pulse width modulation, three different carriers such as saw-tooth, triangular, unipolar inverted sine and unipolar sine reference are utilized. The selection of carriers and reference is based on the control freedom degree principle. The simulation results are taken with hybrid PWM technique at different modulation indices.

2 Proposed Multilevel Inverter Topology

Generally, cascaded type or reduce switch multilevel inverter requires multiple DC sources for producing output voltage. The use of multiple DC sources can demand long cables and the short distance between the sources may lead to voltage unbalance. So, asymmetric DC source multilevel inverter reduces the utilization of DC source count when compared to symmetric DC source multilevel inverter. Also, it requires lesser DC source for large output voltage levels. The new topology becomes more reliable because the lesser count of DC sources are employed. The proposed MLI differs from conventional multilevel inverters by utilizing the reduced count of switches and DC sources. Figure 1 depicts a circuit configuration of new breed of single phase asymmetrical reduced switch multilevel inverter. The proposed topology utilizes two DC sources and seven switches for generating seven

Fig. 1 New type of reduced switch multilevel inverter



level output voltage. The voltage levels are $0V_{dc}$, V_{dc} , $2V_{dc}$, $3V_{dc}$, $-V_{dc}$, $-2V_{dc}$, and $-3V_{dc}$. The two separate DC sources are connected in cascaded manner with appropriate switches. This makes the output voltage level in positive polarity.

A one full bridge inverter is associated with it to develop the output voltage level in both polarities. The generation of $+V_{dc}$ is possible by switching on S_1 , S_2 and S_4 , S_5 switches in full bridge. At that time, the switches S_3 , S_6 and S_7 are turned off. The generation of $+2V_{dc}$ is possible by switching on S_2 , S_3 and S_4 and S_5 switches in full bridge, whereas S_1 , S_6 and S_7 are turned off. The generation of $+3V_{dc}$ is possible by switching on S_4 and S_5 switches whereas the remaining the switches are turned off. Similarly, the negative output voltage levels are also generated by turning on the same switches but the only difference in full bridge inverter switches. Instead of S_4 and S_5 , the S_6 and S_7 switches are turning on to achieve the required output voltage. Table 1 depicts the operating switches of proposed MLI topology. The proposed asymmetric MLI topology utilizes 2^n type of DC source. Here, the n represents the count of DC sources presented in the proposed MLI topology. The general form of output voltage level is given below

$$V_n = 2^{n+1} - 1, \quad n = 1, 2, \dots \tag{1}$$

3 Pulse Width Modulation Technique

From the literature, the single carrier and multi carrier techniques are two major carrier based modulation techniques [1]. Multicarrier PWM techniques are more popular nowadays to generate the switching pulses of MLI switches [3, 15]. Multicarrier pulse width modulation techniques can be categorized into two types such as Level shifting and phase shifting techniques [16]. In level shifting waveforms, the reference waveform is sampled continuously with the carrier waveforms which are displaced by contiguous increments of the waveforms whereas, in the phase shifting techniques the carriers are shifted “ $180^\circ/m$ ” [4]. Here, the ‘ m ’ represents the number of carriers. Generally, many research papers utilizing the triangular waveforms as carriers. In this

Table 1 Switching sequence for proposed seven level multilevel inverter

S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	Level
Off	Off	Off	On	On	Off	Off	3V _{dc}
Off	On	On	On	On	Off	Off	2V _{dc}
On	On	Off	On	On	Off	Off	V _{dc}
On	Off	On	Off	Off	Off	Off	0V _{dc}
On	On	Off	Off	Off	On	On	-V _{dc}
Off	On	On	Off	Off	On	On	-2V _{dc}
Off	Off	Off	Off	Off	On	On	-3V _{dc}

article, the phase disposition techniques are utilized. It is defined as the each carrier has same frequency, amplitude and phases. In this arrangement, the dominant harmonic order is intensified at the carrier frequency. So, the phase disposition carrier technique has lesser harmonic distortion on line voltages. In this paper, the hybrid carriers are utilized. Saw-tooth, triangular and unipolar inverted sine carriers are the three different carriers utilizing in hybrid pulse width modulation techniques. The chosen of carrier arrangement is based on the control freedom degree principle. Also, from the principle, the amplitude and frequency of the each carrier can be decided. The selection of frequency is depends on the application and amplitude of the each carriers are set to be in one. The reference waveform of the PWM is unipolar sinusoidal reference. The unipolar PWM requires half of the carriers when compared to the bipolar or traditional PWM technique. The selecting of particular reference waveform is also possible from the control freedom degree principle. From the principle, the frequency of carrier and reference waveforms, amplitude and phase are to be selected. Figure 2 shows the hybrid carriers with unipolar sine reference pulse width modulation waveform with 0.9 modulation index. The range of amplitude and frequency modulation index can be selected as follows

$$m_a = 2A_m / (m - 1)A_c \quad (2)$$

$$m_f = f_c / f_m \quad (3)$$

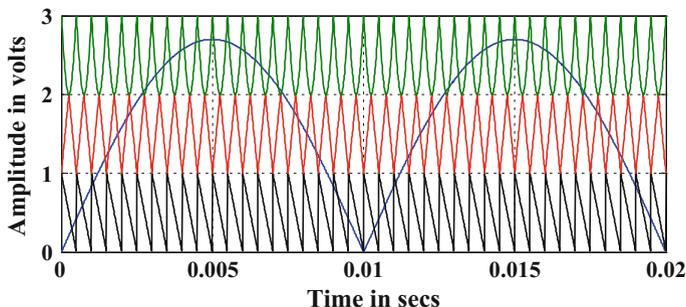


Fig. 2 Proposed hybrid pulse width modulation: hybrid carrier and unipolar sine reference

Here, the m_a and m_f represents the amplitude and frequency modulation index respectively, A_m and f_m represents the amplitude of reference waveform and frequency of reference waveform and the A_c and f_c represents the amplitude of carrier waveform and frequency of carrier waveforms.

4 Simulation Results

The proposed reduced switch multilevel inverter for generating seven level output voltage with two DC sources and seven switches is modeled using MATLAB/Simulink. The simulation parameters are $V_{dc} = 100$ V, resistive load (R) = 100 Ω , resistive and inductive load (RL) = 100 Ω and 20 mH, carrier frequency (f_c) = 2000 Hz and main frequency (f_m) = 50 Hz. Figure 3 shows seven level output voltage for reduced switch multilevel inverter with unipolar sine reference and hybrid carriers with 0.9 modulation index at resistive load. Figure 4 shows the FFT plot for proposed multilevel inverter with unipolar sine reference and hybrid carriers with the modulation index of 0.9 at resistive load. The

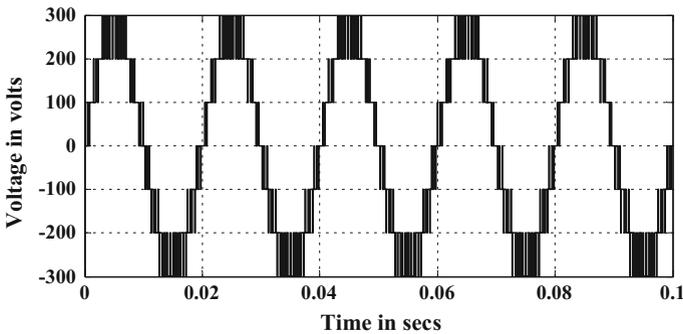


Fig. 3 Output voltage for proposed MLI with hybrid PWM at 0.9 modulation index

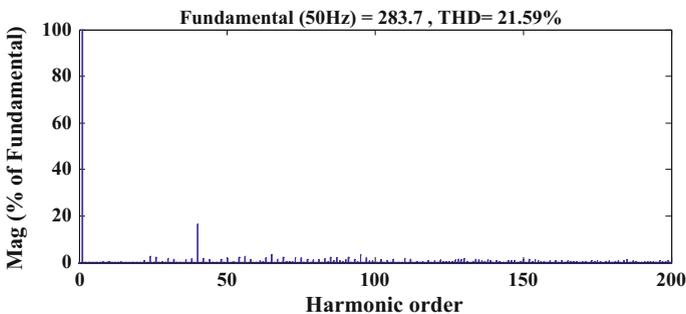


Fig. 4 FFT plot for proposed MLI with hybrid PWM at 0.9 modulation index

simulation results are taken for different modulation indices with the range of 0.8–1. The different parameters such as %THD, fundamental RMS voltage, crest factor, form factor are calculated and tabulated. Table 2 shows that the different parameters value of proposed simulation results. From that table, it is clearly understood that the %THD value increases when the modulation index decreases whereas the fundamental RMS voltage increases. The crest factor value is same for all the range. The harmonic order is dominant in 27th, 29th, 31st, 33rd, 37th and 39th order. The proposed topology is tested with RL load and the waveforms are shown in

Table 2 Simulated results for different parameters with different modulation index

m_a	Sinusoidal PWM with hybrid PD carriers			
	%THD	V_{RMS}	Crest factor	Form factor
1	17	217.7	1.4143	4615.22
0.95	19.73	209.5	1.4143	8218.90
0.9	21.59	200.6	1.4142	72468.48
0.85	22.94	189.9	1.4139	79257.09
0.8	24.48	179.9	1.4140	10984.07

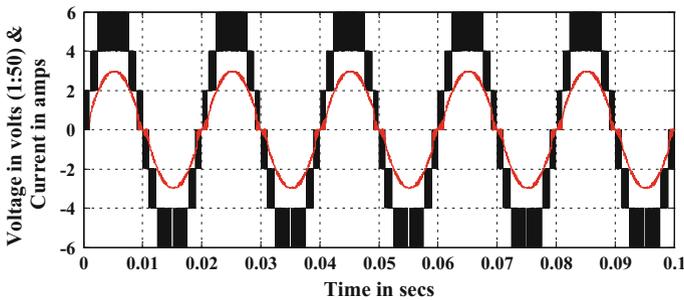


Fig. 5 Output voltage and current waveform for RL load

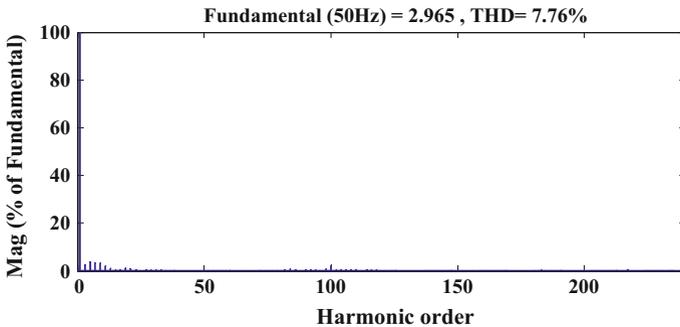


Fig. 6 FFT plot for output current waveform with RL load

Figs. 5 and 6. The output voltage and current waveform for RL load is shown in Fig. 5. The FFT plot for current waveform with RL load is shown in Fig. 6.

5 Conclusions

A new reduced switch MLI topology has been introduced in asymmetric state. The proposed topology needs lower count of DC sources and switches for generating the seven-level output when compared to the traditional CHBMLI. A new HPWM technique has been utilized to generate the switching pulses. The selection of carriers and reference in pulse width modulation is based on the control freedom degree principle. The unipolar sine reference acts as a reference waveform and the combination of saw-tooth, triangular, and unipolar inverter sine carriers has been utilized as carrier waveforms. The proposed multilevel inverter output voltage generates lower %THD based on the hybrid pulse width modulation. Different parameters has been calculated and tabulated.

References

1. Rodriguez, J., Lai, J.-S., Peng, F.Z.: Multilevel inverters: a survey of topologies, controls, and applications. *IEEE Trans. Ind. Electron.* **49**(4), 724–738 (2002)
2. Gupta, K.K., Ranjan, A., Bhatnagar, P., Sahu, L.K., Jain, S.: Multilevel inverter topologies with reduced device count: a review. *IEEE Trans. Power Electron.* **31**(1), 135–151 (2016)
3. Prabaharan, N., Palanisamy, K., Rini Ann Jerin, A.: Asymmetric multilevel inverter structure with hybrid pwm strategy. *Int. J. Appl. Eng. Res.* **10**(55), 2672–2676 (2015)
4. Prabaharan, N., Fathima, A.H., Palanisamy, K.: New hybrid multilevel inverter with reduced switch count using carrier based pulse width modulation technique. In: 2015 IEEE Conference on Energy Conversion, CENCON 2015, pp. 176–180 (2015)
5. Rini Ann Jerin, A., Palanisamy, K., Umashankar, S., Thirumoorthy, A.D.: Power quality improvement of grid connected wind farms through voltage restoration using dynamic voltage restorer. *Int. J. Renew. Energy Res.* **6**(1), 1316–1323 (2016)
6. Babaei, E., Laali, S.: Optimum structures of proposed new cascaded multilevel inverter with reduced number of components. *IEEE Trans. Ind. Electron.* **62**(11), 6887–6895 (2015)
7. Barzegarkhoo, R., Zamiri, E., Vosoughi, N., Kojabadi, H.M., Chang, L.: Cascaded multilevel inverter using series connection of novel capacitor-based units with minimum switch count. *IET Power Electron.* **9**(10), 2060–2075 (2016)
8. Hsieh, C.-H., Liang, T.-J., Chen, S.-M., Tsai, S.-W.: Design and implementation of a novel multilevel dc-ac inverter. *IEEE Trans. Ind. Appl.* **52**(3), 2436–2443 (2016)
9. Odeh, C.I., Obe, E.S., Ojo, O.: Topology for cascaded multilevel inverter. *IET Power Electron.* **9**(5), 921–929 (2016)
10. Zamiri, E., Vosoughi, N., Hosseini, S.H., Barzegarkhoo, R., Sabahi, M.: A new cascaded switched-capacitor multilevel inverter based on improved series-parallel conversion with less number of components. *IEEE Trans. Ind. Electron.* **63**(6), 3582–3594 (2016)
11. Gautam, S.P., Sahu, L.K., Gupta, S.: Reduction in number of devices for symmetrical and asymmetrical multilevel inverters. *IET Power Electron.* **9**(4), 698–709 (2016)

12. Grandi, G., Ostojic, D., Rossi, C., Casadei, D.: A new multilevel conversion structure for grid-connected PV applications. *IEEE Trans. Ind. Electron.* **56**(11), 4416–4426 (2009)
13. Du, Z., Tolbert, L.M., Chias son, J.N.: A cascade multilevel inverter using a single dc source. *IEEE Appl. Power Electron. Conf.* 426–430 (2006)
14. Gautam, S.P., Kumar, L., Gupta, S.: Hybrid topology of symmetrical multilevel inverter using less number of devices. *IET Power Electron.* **8**(11), 2125–2135 (2015)
15. Sanjeevikumar, P., Blaabjerg, F., Wheeler, P., Khanna, R., Mahajan, S.B., Dwivedi, S.: Optimized carrier based five-level generated modified dual three-phase open-winding inverter for medium power application. In: *IEEE International Transportation Electrification Conference and Expo, IEEE-ITEC'16, Korea*, pp. 40–45 (2016)
16. Sanjeevikumar, P., Blaabjerg, F., Wheeler, P., Ojo, O.: Three-phase multilevel inverter configuration for open-winding high power application. In: *6th IEEE International Symposium on Power Electron. For Distributed Generation Systems, IEEE-PEDG'15, Germany* (2015)

Watershed-Matching Algorithm: A New Pathway for Brain Tumor Segmentation

S.M. Kamrul Hasan, Yugoshree Sarkar and Mohiudding Ahmad

Abstract Brain tumor detection through Magnetic Resonance Imaging (MRI) is a very challenging task even in today's modern medical image processing research. To form images of the soft tissue of the human body, surgeons use MRI analysis. They segment the images manually by partitioning into two distinct regions which is erroneous and at the same time, may be time-consuming. So, it is a must be better the MRI images segmentation. This paper outlines a new finding to detect brain tumor for better accuracy than earlier techniques. We segment the tumor area from the MR image and then to find the area of the segmented region, we use another algorithm to match the segmented part with the input image. In addition, the paper concludes with the status checking of the tumor and provides a necessary diagnosis of brain tumor. Lastly, we compare our proposed model with other techniques and get a far better result.

Keywords Artifacts · Brain tumor segmentation · Magnetic resonance imaging Sift algorithm · Topology · Watershed segmentation

S.M.K. Hasan (✉) · M. Ahmad

Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh
e-mail: leon.kuet.eee@gmail.com

M. Ahmad

e-mail: mohiuddin.ahmad@gmail.com

S.M.K. Hasan

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Y. Sarkar

Department of Electrical and Electronic Engineering, National Institute of Technology Warangal, Warangal, India
e-mail: anudka@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_5

1 Introduction

Brain tumor, which is one of the most quotidian neurological brain complications, creates a devastating effect on many lives. It is an uncontrolled mass of tissues found in different parts of the brain like pituitary gland, skull, neurons etc. According to Bangladesh Cancer Organizations and Resources (BCOR), about 122,700 people are newly diagnosed with cancer per year around Bangladesh with 91,300 people dying from cancer. Of the cancer patients approximately, 18.3% are diagnosed with the malignant type brain tumor. It can be the cause of most difficult cancers to treat. So, it is important to detect tumor accurately. Though past researchers have prepared their research, still now it is a vast research field because of the variation of the data of MRI. Brain tumors normally stay in the posterior cranial fossa so; it is difficult to detect it manually by the help of Biopsy. Automatic segmentation of MR images is the only process that partitions an MR image into its constituent regions and identifies ROI (regions of interest). Brain tumors show its symptoms by using two factors: tumor location and tumor volume. The accurate and automated segmentation can reduce the mortality rate and improve the diagnosis. Tumors have high-density blood vessels and so they are highly vascularized tissues. Hence, MRI can only better perform of variation between the viable and non-viable tumor than the CT scanning. It provides the detailed information about the brain tumor anatomy that will allow effective diagnosis of the disease. It also shows the cross sectional view of both horizontal and vertical planes of the brain.

2 Literature Review

Accurate segmentation is the key step in medical image. Different researchers have proposed different segmentation techniques. To understand the efficacy of these techniques, these are classified into two categories named: supervised and unsupervised techniques. Authors in [1] use discrete Markov Random Field (MRF) algorithm for the brain tumor segmentation. Here, the parameter estimation and computing probability for this method is very difficult. Authors in [2] use the Morphological Filtering Technique. But due to having no complexity in the system, it is not perfect segmentation algorithm. Author in [3] uses level set methods as they handle complex geometries very easily. Ironically, they are user-dependent being adjusted several parameters. Apart from that, they need a large number of iterations [4].

This paper presents a completely new finding named WM (Watershed-Matching) to detect tumors. We use a traditional watershed segmentation algorithm for brain tumor segmentation and then use SIFT (Scale-Invariant Feature Transform) algorithm to match the segmented region with the original image and thus, we calculate the area of the tumor region from the matching part by projection and get the better result than any algorithm used earlier.

3 Proposed Methodology

Our proposed method is a new algorithm shown in Fig. 1. We divide our model into three modules: image acquisition, preprocessing, and segmentation.

3.1 Image Database

Before starting our work, we collect brain tumor image database to compare our model. These brain images can be obtained from MRI scan and we present these in a 2D matrix having the pixels. We collect our database from two different sources. We collect 100 T1-weighted brain tumor images of 14 patients along with truth images from Montreal Neurological Institute (MNI BITE) database [5]. All the images were is MINC format. We convert the MINC to MHA format. We collect another 100 MR images from 5 patients (age: 25 ± 6 years, height: 168 ± 4.4 , and weight: 59 ± 9 male). The second database was prepared by ourselves. Then, we scale all the collected images in MATLAB with $256 * 256$ dimensions.

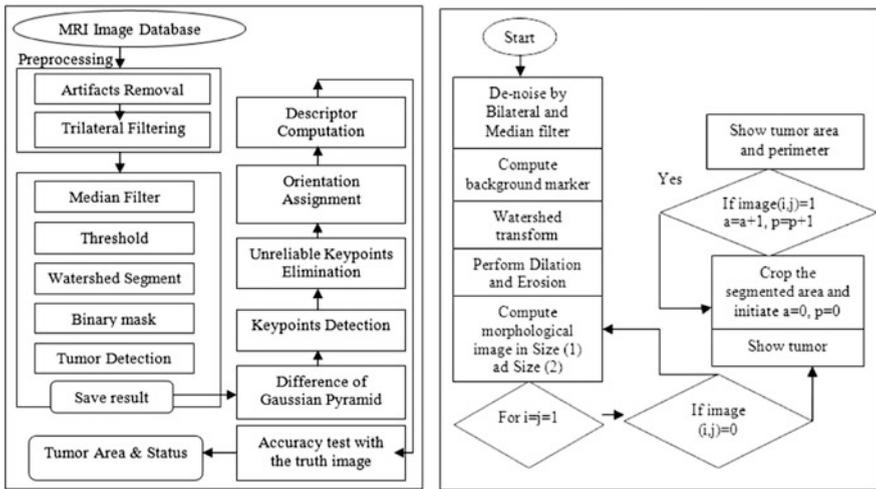


Fig. 1 Block diagram of the overall method (from left to right). The overall process of the watershed segmentation and the SIFT algorithm is depicted here in the block diagram, flow chart of segmentation process using watershed algorithm

3.2 Preprocessing Stage

Preprocessing is the stage where noise is removed and any minute details are ameliorated. To procure segmentation without any noise, this module is a must. Our preprocessing module includes noise reduction from the image using a trilateral filter and artifacts removal that include the patient's name, age, dimension of the scanned image and other information. This refinement will revamp the overall image quality.

4 Watershed Segmentation Algorithm (WSA)

Image segmentation is a technique of extracting the most important characteristics to comprehend the information more easily. In our proposed method, we use Watershed Segmentation Algorithm (WSA) is based on mathematical morphological process. It is used to find the watershed lines in an image for separating the distinct regions in touching objects. To understand the watershed algorithm, we can think of a gray-scale image I , where two objects are touching each other. In watershed transform, this image can be regarded as a topological surface, where the value of $I(x, y)$ corresponds to heights [6].

5 Scale Invariant Feature Transform Method (SIFT)

Scale Invariant Features Transform (SIFT) algorithm normally transforms an image into a collection of local feature vectors. Here, we use segmented tumor region with the original image and to do this we have to find the keypoints to find the matching. We put down the steps in Fig. 1 that we have performed for matching. We have introduced the overall steps of SIFT algorithm to create a new era in the image segmentation accuracy-check process [7]. We implement a 16×16 array and use an 8 bins histogram for computing the keypoints orientation.

6 Results and Discussions

In method 1, we pass the input image through two steps of filtering and then apply watershed segmentation algorithm (WSA). Here, we use the trilateral filter and filtering the improvement factors. We can see from Fig. 2 that, the tumor region is segmented from the earlier over-segmented watershed contour. We apply this transform to get the similar objectives out of the background. As we see from the gradient image, the region with less change in gray has lower gradient and gradient

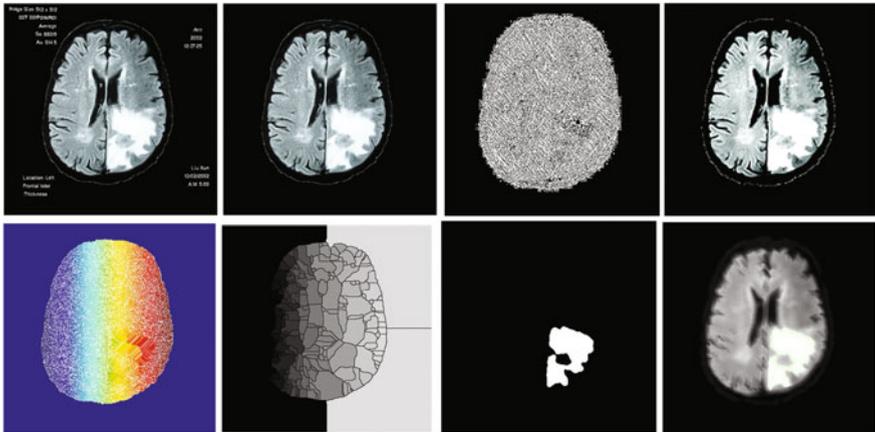


Fig. 2 Experimental result of watershed segmentation (see from *left to right* and *top to bottom*) input MR T1 image, artifacts removal, noise reduction and edge smoothing by trilateral filtering, de-noise image by a bilateral filter and reduction of impulse noise by median filtering, watershed transform in RGB color, gradient watershed transform, *white part* refers to tumor segmentation, *blue contouring* refers to tumor area detection using WSA

is higher in the neighbor boundary than the inside region. That’s why a gradient structural element (size of $3 * 3$) is used as the reference image. Then, we apply a morphological operation that includes opening and closing operation to get back the original one. In the dilation operation, an original image is eroded first and then dilated. If the gradient image can fulfill the requirement of different shape, then opening operation is performed as a union formation and in this way, the whole image when goes through the filters then the main image can be retained and the noise is eliminated. In the same way, closing operation is performed where the main image is first dilated. After this opening-closing operation, the original image is rendered and thus we get the segmented region as in the fourth row.

We analyze our algorithm for 3D MR images. We construct 3D image using 3D slicer and we apply watershed algorithm. The input image is loaded in the software, then we can see the directional images in three different window. The volume model is built with the pre-chosen color. The first model is corpus callosum labeled color: green. The second model is frontal lobe white matter right labeled 17 (threshold 17:17), color: green. The third model is frontal lobe white matter left labeled 17 (threshold 17:17), color: green. Then, we create a surface volume using voxel maker. We can change different parameters like relief, color, luminosity etc. as per our need [8, 9].

In the medical diagnosis field, these tools have wide application for better visualization of segmented image. Two-dimensional images can’t detect the distinction of volume in the tumors system and cannot provide objective data that 3D slicer can do [10].

So, we conclude that our algorithm works well in 3D images also. As we see from Fig. 3 that the tumor exists in the Ventricles of the brain which is shown green color in the image.

6.1 Shift Results

As from our 2D and 3D data, we can segment tumor perfectly and so, to find the tumor area, we apply SIFT algorithm. We can see from Fig. 4 that, the input and output MR image that we use for watershed transform, are also used for SIFT algorithm. According to the algorithm, we find the possible keypoints between input and output and then we discard the unreliable keypoints. For our experiments, we used T1-magnitude images. To create a training set of SIFT keypoints, first we find the features that carry the most relevant information. Then we extract the keypoints to detect the tumor region. From Fig. 4, we can behold that the tumor region, which is our outcome has the keypoints with high-density cluster (green/blue “+” sign). Lastly, the tumor is detected by the selection and matching of the extracted keypoints.

6.2 Tumor Area Calculation

From the last part of Fig. 4, the tumor area is calculated as our following ways in (1)–(3):

$$\text{Image} = \sum_{l=0}^{255} \sum_{w=0}^{255} [x(0, \text{whitepixels}) + x(1, \text{blackpixels})] \quad (1)$$

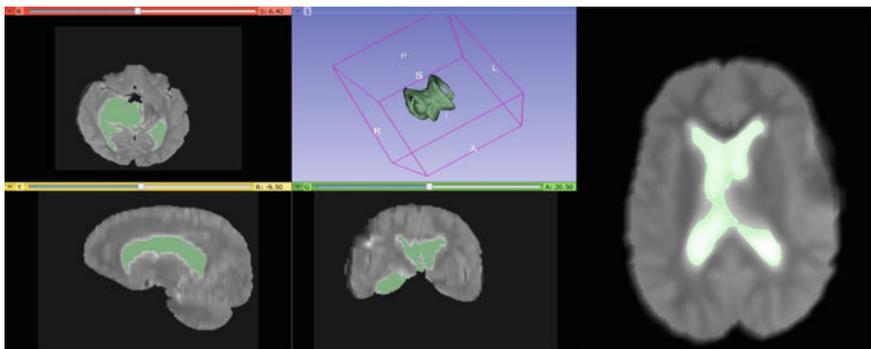


Fig. 3 Experimental result of watershed segmentation (from left to right). All slices with the tumor in 3D, watershed result of brain tumor image (red slice only)

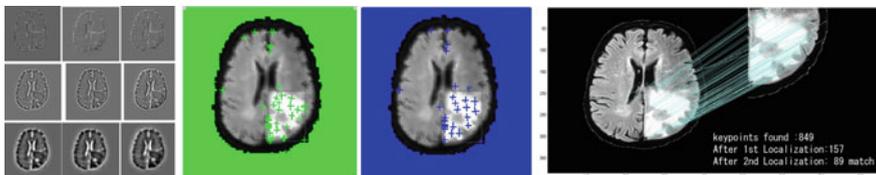


Fig. 4 Experimental result of SIFT segmentation (see from *left to right*) finding features of input MR brain tumor image, finding keypoints; the density of *green signs* and also the *blue signs* on the tumor area indicate the tumor region, keypoints are localized and then finding matching and lastly find 89 matches between the input and the output image and thus confirms the tumor area from the binary image

$$\text{White pixels count, } W = \sum_{l=0}^{255} \sum_{w=0}^{255} [x(0, \text{whitepixels})] \quad (2)$$

$$\text{Tumor size} = [(W)] * 0.264 \text{ mm}^2 \quad (3)$$

where

$$1 \text{ Pixel} = 0.264 \text{ mm} \quad (4)$$

The tumor area is calculated to check the status of the tumor. As in our case, the area calculated was 269. So, it was in the initial stage. If the detected white pixels ≥ 500 then it will be in critical stage. We compare our final result with the truth image and calculate the accuracy.

7 Conclusion

The brain tumor is treatable if it has been identified in the earliest stages of the disease. We proposed and implemented an innovative method for the brain tumor detection and segmentation in this paper. Our main contribution consisted of modeling improved watershed algorithm with three steps of de-noising filtering and designing scale-invariant feature transform algorithm where the optimized features were selected. Traditional over segmentation problem could be minimized by our improved algorithm. As the MR images are highly affected by noise and artifacts. We preprocessed the images using artifacts removal, median filter and trilateral filter for improving the segmentation quality. Due to this improved combination, our proposed method is far better than any single or other combination algorithms. To check the accuracy of our algorithm, we compared the result with the truth images and acquired 98.5% accuracy. Here, we also introduced status checking of the tumor. We calculated the area of the tumor and then set a decision rule to decide whether it is in a critical or initial stage. This status checking made our system more

robust. Our framework can be used in the general application. In future, we would use it not only for brain tumor segmentation but also for other applications like the bone tumor, lung tumor, or other segmentation purposes. We will reduce the number of manual interactions. This will help the physicians to prosecute the further treatment process in advance to treat tumor patients.

Acknowledgements Higher Education Quality Enhancement Project (HEQEP), UGC, Bangladesh Department of EEE, KUET, Bangladesh partly supports this work.

References

1. Besbes, A., Komodakis, N., Lings, G., Paragios, N.: Shape priors and discrete MRFs for knowledge-based segmentation. In: *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1295–1302 (2009)
2. Ananda, R.S., Thomas, T.: Automatic segmentation framework for primary tumors from brain MRIs using morphological filtering techniques. In: *IEEE International Conference on Biomedical Engineering and Informatics*, pp. 238–242 (2012)
3. Dubey, R.: Semi-automatic segmentation of MRI brain tumor. *ICGST-GVIP J.* **9**, 33–40 (2009)
4. Kazerooni, A.F., Ahmadian, A., Serej, N.D., Rad, H.S., Saberi, H., Yousefi, H., Farnia, P.: Segmentation of brain tumors in MRI images using multi-scale gradient vector flow. In: *IEEE International Conference on Engineering in Medicine and Biological Society*, pp. 7973–7976 (2011)
5. Mercier, L., Del Maestro, R.F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.L.: On-line database of clinical MR and ultrasound images of brain tumors. *Int. J. Med. Phys. Res. Pract.* **39**(6), 3253–3261 (2012)
6. Rettmann, M.E., Han, X., Prince, J.L.: Watersheds on the cortical surface for automated sulcal segmentation. In: *Proceeding of IEEE Workshop Mathematical Methods in Biomedical Image Analysis*, pp. 20–27 (2000)
7. Yong, L., Zhengyuan, Y., Yuanzhi, C.: Efficient parallel recursive gaussian SIFT algorithm based on multi-core DSP. In: *IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 402–405 (2015)
8. Gonzalo Dominguez, M. Hernandez, C., Ruisoto, P., Juanes, J.A., Prats, A., Hernandez, T.: Morphological and volumetric assessment of cerebral ventricular system with 3D slicer software. *J. Med. Sys.* **40**(154) (2016)
9. Kim, G., Jung, H.J., Lee, H.J., Lee, J.S., Koo, S., Chang, S.H.: Accuracy and reliability of length measurements on three-dimensional computed tomography using open-source osirix software. *J. Digital Imag.* **25**, 486–491 (2012)
10. Puisoto, P., Juanes, J.A., Contador, I., Mayoral, P., Prats, A.: Experimental evidence for improved neuroimaging interpretation using three-dimensional graphic models. *Anat. Sci. Educ.* **5**(3), 132–137 (2012)

A Compact Slotted Textile Patch Antenna for Ultra-wide Band Application

Nikhil Kumar Singh, Vinod Kumar Singh, Anurag Saxena,
Akash Kumar Bhoi, Amik Garg and K.S. Sherpa

Abstract In this article a compact slotted textile patch antenna for Ultra-wideband applications with enhanced bandwidth is presented. The optimized antenna is intended on textile substrate with dielectric constant of 1.7. The compact size of antenna is $50 \times 50 \times 1$ mm. This presented rectangular-shaped antenna that has been simulated using CST tool. The presented antenna resonates at frequency 3.98 GHz with the reflection coefficient of -34.12 . It covers a bandwidth of 6.27 GHz extending from 2.86 to 9.13 GHz. It has maximum directivity of 2.969 dBi.

Keywords Partial ground · Ultra-wide band · Wireless application
Textile patch · CST tool

N.K. Singh · V.K. Singh (✉) · A. Saxena
Department of Electrical Engineering, Bhagwant University, Rajasthan, Ajmer,
Rajasthan, India
e-mail: singhvinod34@gmail.com

N.K. Singh
e-mail: nitinm18@gmail.com

A. Saxena
e-mail: anurag.saxena.anurag@gmail.com

A.K. Bhoi · K.S. Sherpa
Department of EEE, Research & Development Section, Sikkim Manipal Institute
of Technology, Sikkim Manipal University, Majhitar, India
e-mail: akash730@gmail.com

K.S. Sherpa
e-mail: karmasherpa23@gmail.com

A. Garg
Research & Development Section,
Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Majhitar, India
e-mail: amikkarg@gmail.com

1 Introduction

Currently, the fast improvement of modern communication systems is necessary for transportable devices for some important features which includes easy designing, lightweight, small in size, compatible with microwave, millimeter wave integrated circuits, less production cost, and easy fabrication of microstrip antennas. The microstrip antenna has abundant useful properties which include tiny size, lowcost of the fabrication, lightweight, ease of setting up but the main limitations of printed antennas remain their narrow bandwidth features which limit the range of frequency over which the antenna can work efficiently. Microstrip antenna comprises three most important parts which is substrate, patch, and ground. A dielectric substrate is sandwiched between radiating patch and ground plane. The conducting patch is located on the dielectric substrate which is used as radiating element. On other side of dielectric substrate there is conducting layer used as ground part [1–6]. Microstrip antenna plays major role in wireless communication system. Microstrip antennas are used in high-performance aircrafts, radar, missiles, and other spacecraft. It has many advantages such as its lightweight, simple structure, ease of addition, and less cost. Microstrip antenna requires very less space for installation as these are simple and small in size. The only space these require is the space for the feed line which is placed behind the ground plane. Microstrip antennas are low profile, simple, conformable to planar and nonplanar surfaces, inexpensive to manufacture using modern printed circuit technology [7–14].

There are a variety of methods for enhancing the bandwidth of textile microstrip antenna such as expand the substrate thickness, utilizing low dielectric material, using numerous feeding techniques and impedance matching. But the bandwidth and the thickness of the antenna are both conflicting properties, that is, enhancement of bandwidth increases the size and thickness of presented antenna [15–22].

2 Antenna Design Configuration

The offered antenna works for three resonant frequencies which are shown in Fig. 1. If there is more than one resonant part available with each operating at its own resonant frequency, then the overlapping of multiple resonance leads to broadband applications. The basis of the proposed antenna was an annular ring with outer patch radius 14 mm and inner patch radius of 10 mm. Then rectangular shape was obtained by cutting a rectangular shape slot from the circular patch and finally the proposed design was obtained by cutting a circle in rectangle and the parameters of presented antenna is shown in Table 1.

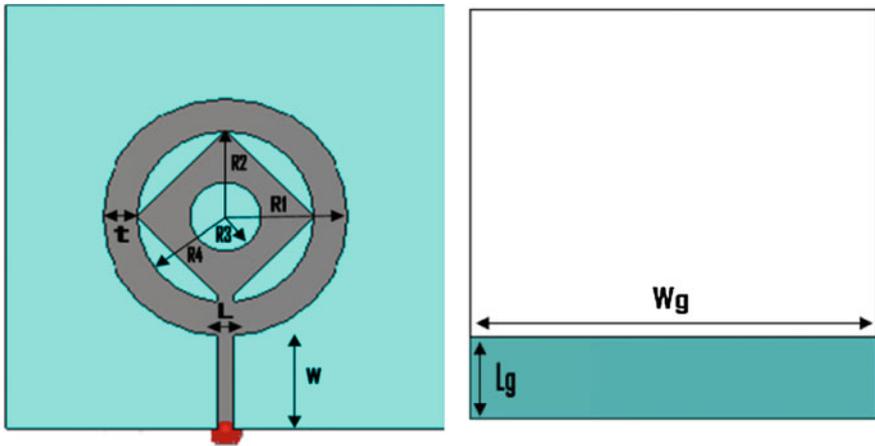


Fig. 1 Configuration of the proposed pentagonal textile microstrip antenna

Table 1 Dimensions of anticipated textile antenna

Design parameters	Value
Relative permittivity (ϵ_r)	1.70
Loss tangent ($\tan \delta$)	0.025
Substrate thickness (h)	1.0 mm
Outer circle patch radius (R_1)	14 mm
Inner circle patch radius (R_2)	10 mm
Inner circle slot radius (R_3)	4 mm
Square patch dimension	5 × 5 mm
Microstrip feed line ($L \times W$)	2 × 15
Substrate dimension ($L_s \times W_s$)	50 × 50
Partial ground plane ($L_g \times W_g$)	50 × 10

3 Optimization of Proposed Microstrip Antenna

The CST software has been used to design and optimize the results produced by the rectangular-shaped antenna. In the beginning circular patch was used to achieve the desired bandwidth, then the circular patch was modified to rectangular-shaped patch and at last the proposed shaped patch was achieved to optimize the results. Figure 2

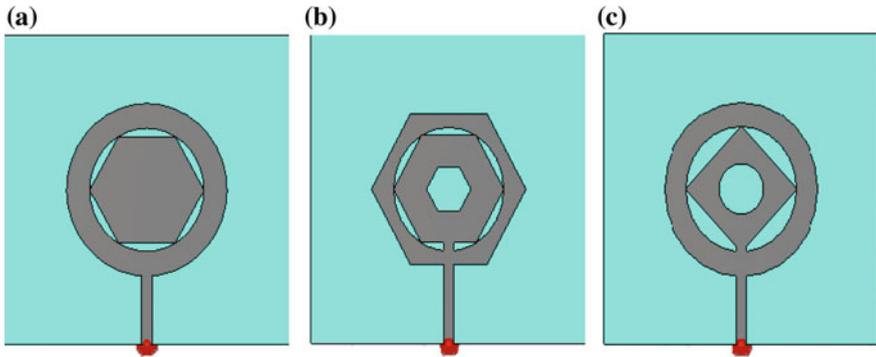


Fig. 2 Geometry of the proposed antennas **a** Antenna 1 **b** Antenna 2 **c** proposed Antenna 3

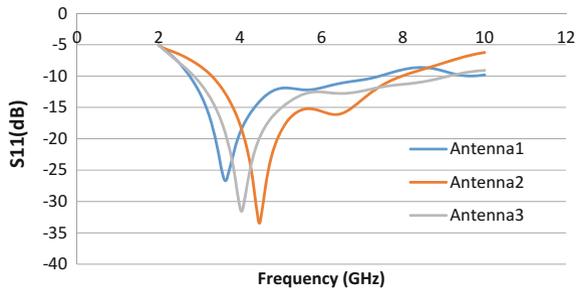


Fig. 3 Simulated return loss versus frequency for the proposed textile microstrip Antenna 1, Antenna 2, and Antenna 3

shows the geometry of antenna1, antenna2, and antenna3. The return loss for antenna1, antenna2, and antenna3 has been denoted by blue, orange, and gray colored curves, respectively, in Fig. 3.

4 Radiation Pattern

Figure 4 depicts simulated 2D radiation pattern of proposed microstrip antenna at 3.98 GHz which describes the main lobe direction = 168° and angular width (3 dB) = 73.3° and main lobe magnitude = 3 dBi at $\varphi = 90$. Also it gives the main

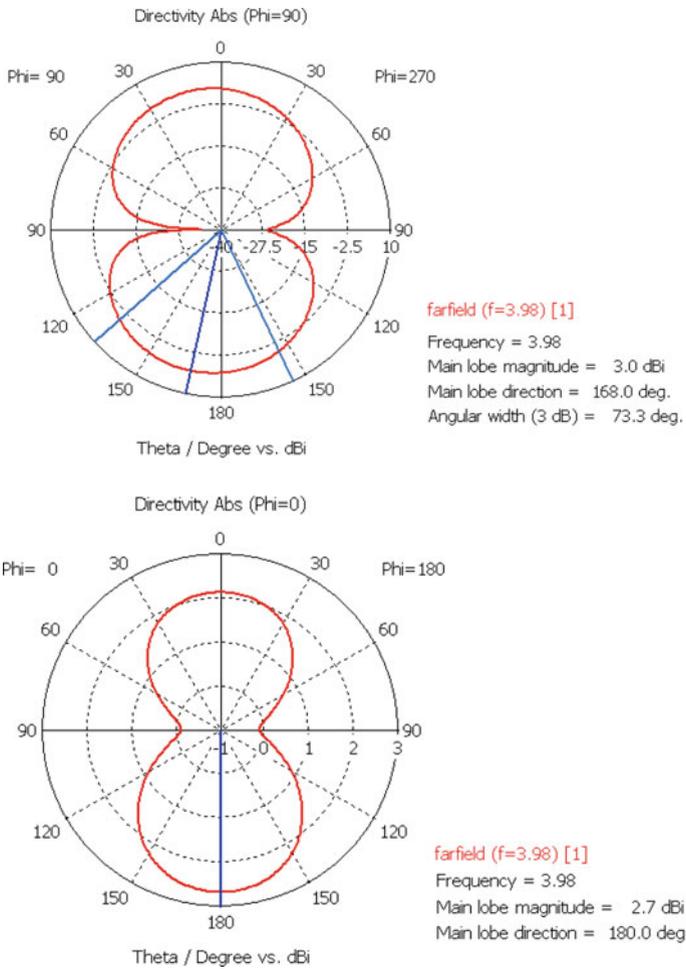


Fig. 4 Simulated 2D radiation pattern of proposed textile microstrip antenna at 2.98 GHz

lobe direction = 180° and main lobe magnitude = 2.7 dBi at $\varphi = 0^\circ$. Figure 5 shows simulated 3D radiation pattern of proposed microstrip antenna at 3.98 GHz which gives good radiation efficiency of about -1.798 dB and directivity of 2.969 dBi.

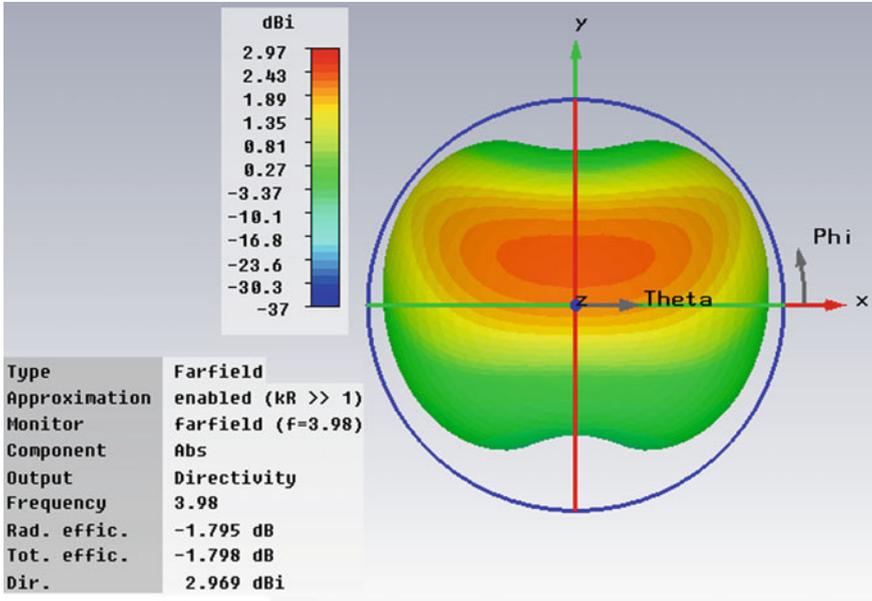


Fig. 5 Simulated 3D radiation pattern of proposed textile microstrip antenna at 2.98 GHz

5 Conclusions

A slotted textile patch antenna with circle in its center is estimated for Ultra-Wide Band application. The overall dimension ($50 \times 50 \times 1$ mm) of presented antenna is very compact, thus suitable for installation on heavy satellites. The results are obtained by using CST software. The antenna has maximum directivity of about 2.97 dBi. The antenna has stable radiation pattern over the entire 6.27 GHz bandwidth that extends from 2.86 to 9.13 GHz, thus the presented antenna is appropriate for Ultra-wide Band application.

References

1. Raghupatruni, V., Krishna, R., Kumar, R.: Design of temple shape slot antenna for ultra wideband applications, progress. Electromagn. Res. B **47**, 405–421 (2013)
2. Balanis, C.A.: Antenna Theory: Analysis and Design. Wiley, New York (2004)
3. Wong, K.-L.: Compact and Broadband Microstrip Antennas. Wiley. ISBNs: 0-471-41717-3 (Hardback) (2002)
4. Singh, V.K., Ali, Z., Ayub, S., Singh, A.: A wide band compact microstrip antenna for GPS/DCS/PCS/WLAN applications. Intell. Comput. Netw. Inf. (Springer) **243**, 1107–1113 (2014). ISBN: 978-81-322-1664-3

5. Srivastava, R., Singh, V.K., Ayub, S.: Comparative analysis and bandwidth enhancement with direct coupled C slotted microstrip antenna for dual wide band applications. *Front. Intell. Comput. Theor. Appl.* (Springer) **328**, 449–455 (2015). ISBN: 978-3-319-12011-9
6. Singh, V.K., Ali, Z., Ayub, S., Singh, A.: Bandwidth optimization of compact microstrip antenna for PCS/DCS/Bluetooth application. *Central Eur. J. Eng.* (Springer) **4**(3), 281–286 (2014)
7. Singh, N.K., Singh, A., Singh, V.K.: Design and performance of wearable ultra wide band textile antenna for medical applications. *Microwave Opt. Technol. Lett.* **57**(7), 1553–1557 (2015)
8. Srivastava, S., Singh, V.K., Ali, Z., Singh, A.: Duo triangle shaped microstrip patch antenna analysis for WiMAX lower band application. In: *International Conference on Computational Intelligence: Modelling Techniques and Applications (CIMTA-2013)*, Procedia Technology Elsevier, vol. 10, pp. 554–563 (2013)
9. Singh, N.K., Singh, V.K., Naresh, B.: Textile antenna for microwave wireless power transmission. In: *International Conference on Computational Modelling and Security (CMS 2016)*, Procedia Computer Science, vol. 85, pp. 856–861 (2016)
10. Gupta, N., Singh, V.K., Ali, Z., Ahirwar, J.: Stacked textile antenna for multi band application using foam substrate. in: *International Conference on Computational Modelling and Security (CMS 2016)*, Procedia Computer Science, vol. 85, pp. 871–877 (2016)
11. Rawat, A., Singh, V.K., Ayub, S.: Compact wide band microstrip antenna for GPS/WLAN/WiMax applications. *Int. J. Emerg. Trends Eng. Dev.* **7**(2), 140–145 (2012)
12. Baudh, R., Kumar, R., Singh, V.K.: Arrow shape microstrip patch antenna for WiMax application. *J. Env. Sci. Comput. Sci. Eng. Technol.* **3**(1), 269–274 (2013)
13. Loni, J., Ayub, S., Singh, V.K.: Performance analysis of microstrip patch antenna by varying slot size for UMTS application. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 01–05 (2014)
14. Srivastava, R., Ayub, S., Singh, V.K.: Dual band rectangular and circular slot loaded microstrip antenna for WLAN/GPS/WiMax applications. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 45–48 (2014)
15. Dhupkariya, S., Singh, V.K.: Textile antenna for c-band satellite communication application. *J. Telecommun. Switching Syst. Netw.* **2**(2), 20–25 (2015)
16. Loni, J., Singh, V.K.: Development of bandwidth enhanced microstrip patch antenna for UMTS application. *J. Microwave Eng. Technol.* **2**(1), 01–07 (2015)
17. Singh, V.K., Naresh, B.: Multi resonant microstrip antenna with partial ground for radar application. *J. Telecommun. Switching Syst. Netw.* **2**(1), 01–05 (2015)
18. Singh, M., Singh, V.K., Naresh, B.: Rectangular slot loaded circular patch antenna for WLAN application. *J. Telecommun. Switching Syst. Netw.* **2**(1), 07–10 (2015)
19. Ali, Z., Singh, V.K., Kumar, A., Shahanaz, A.: E shaped microstrip antenna on rogers substrate for WLAN applications. *Proc. IEEE*, 342–345 (2011)
20. Din, N.M., Chakrabarty, C.K., Bin Ismail, A., Devi, K.K.A., Chen, W.Y.: Design of RF energy harvesting system for energizing low power devices, progress. In *Electromag. Res.* **132**, 49–69 (2012)
21. Osman, M.A.R., Rahim, M.K.A., Azfar, M., Samsuri, M., Zubir, F., Kamardin, K.: Design, implementation and performance of ultra-wideband textile antenna, progress. *Electromag. Res. B* **27**, 307–325 (2011)
22. Singh, V.K., Singh, N.K.: Compact circular slotted microstrip antenna for wireless communication systems. *J. Microw. Eng. Technol.* **1**(1), 07–14 (2015)

Optimization of Hodgkin–Huxley Conductance-Based Model Using Particle Swarm Optimization and Firefly Method

Rashmi Deka, Kuntala Boruah and Jiten Ch. Dutta

Abstract Hodgkin–Huxley have developed a neuron model describing the biophysical nature of a neuron in 1952. The model describes the ionic exchange taking place in the neuron mathematically. In this paper, the values of the parameters of the H-H model are optimized. Particle swarm optimization (PSO) and Firefly algorithm (FA) methods are used here and both the methods are compared for optimal solution. These methods are metaheuristic methods and they follow the nature's law of flocking in PSO and flashing pattern in FA. The optimized parameters are found keeping in reference the signals generated directly from the nonlinear equations given by Hodgkin–Huxley. Estimation of parameters has its own importance in biomedical field for determination of relation between parameters.

Keywords Hodgkin–Huxley · Particle swarm optimization (PSO) · Firefly algorithm (FA) · Conductance · Membrane potential

1 Introduction

Hodgkin–Huxley have developed a neuron model describing the biophysical nature of a neuron [1–3]. It describes the ionic flow inside and outside the cell mathematically. This followed the development of many neuron models but H-H model has its own significance. In this paper, the parameters related to this model are taken

R. Deka (✉) · K. Boruah · J.Ch. Dutta
Department of Electronics and Communication Engineering,
Tezpur University, Tezpur, Assam, India
e-mail: rashmeedeka10@gmail.com

K. Boruah
e-mail: kuntala17@gmail.com

J.Ch. Dutta
e-mail: jitend@tezu.ernet.in

into consideration for optimization purpose using PSO and FA [4–7]. Both the methods are compared and the method which yields result closer to H-H model efficiently is observed. Optimization of neuron model and its signals has been done earlier using stimulated annealing, Genetic algorithm [8–10] for application in neurology. But these advanced algorithms were found to be more efficient. Action potential is a signal which transfers information from brain to the body and vice versa. It is produced from the ionic flow of potassium, sodium, and leakage ions present in the cell. Flow of ions is dependent on the gates opening and closing for sodium and potassium ions. The equations given below (1) and (2) are the parameters related to neuron signal. g_{Na} , g_K , and g_l are the conductances related to sodium, potassium, and leakage ions. m and h are the gates for activation and inactivation of sodium ions. n is the number of gates opening for potassium ions. The parameters g_{Na} , g_K , E_{Na} , E_K are optimized in this paper with PSO and firefly method. Both the methods are compared and their efficiency is observed. Total current is represented by I and it is divided into ionic current and capacitance membrane. V is the membrane potential.

$$I = C_M \frac{dV}{dt} + I_{ion} \quad (1)$$

$$I = C_{membrane} \frac{dV}{dt} + \bar{g}_K n^4 (V - V_K) + \bar{g}_{Na} m^3 h (V - V_{Na}) + \bar{g}_l (V - V_l). \quad (2)$$

2 Materials and Method

Estimation of the parameter is done in MATLAB using PSO and FA algorithm.

Dr. Eberhart and Dr. Kennedy have developed the particle swarm optimization method. It follows the principle of bird flocking. It repeatedly searches for optimum solution regularly. The particle optimizer searches for a best solution and another optimizer searches for the best solution. Then the PSO searches for the global optimum solution. PSO do not use the principle of mutation and crossover. PSO yields better results and in a cheaper way.

Firefly algorithm follows the principle of fireflies flashing behavior. It searches for the brighter ones and when it cannot search for the brighter ones, it moves randomly. The brightness decreases with distance. The brightness is dependent on objective function. The population is divided into subgroups and the subgroups search for the optimum solution simultaneously. FA subdivides into group automatically which is an advantage for large population. The best solution is then evaluated from the groups combined.

These two optimization methods are used for estimation of parameters in conductance-based model in H-H model.

3 Results and Discussions

Figures 1 and 2 show the estimation of sodium conductance using FA and PSO keeping the H-H signal as reference signal. It has been observed that FA generates results faster than PSO. Moreover, the value generated by FA is more optimized and closer to H-H value. Figures 3 and 4 depict the potassium conductance estimation using FA and PSO. Table 1 shows the estimated values and the values given by H-H is shown for comparison.

Fig. 1 Optimization of sodium conductance using PSO

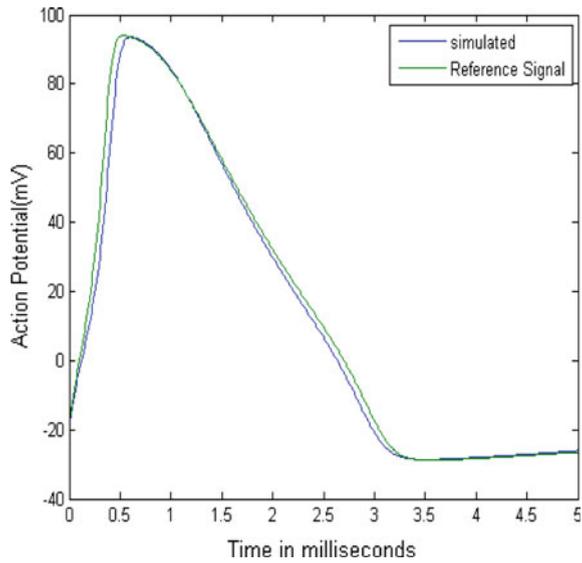


Fig. 2 Parameter estimation of potassium conductance using PSO

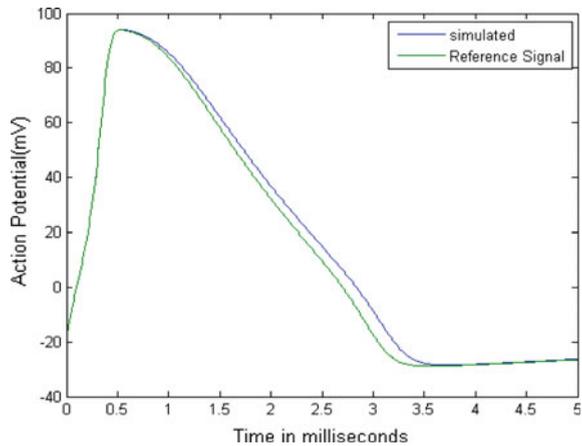


Fig. 3 Estimation of sodium conductance using FA

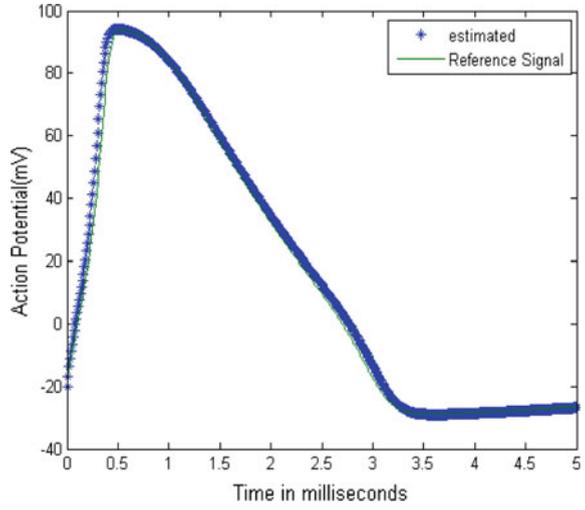


Fig. 4 Potassium conductance optimization using FA

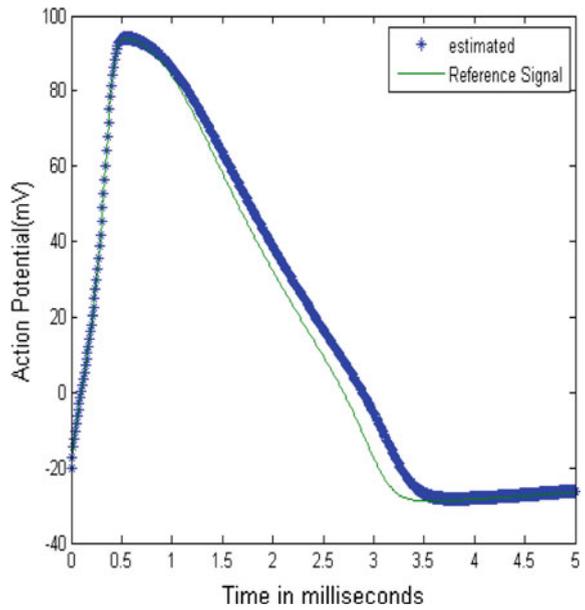


Table 1 Estimated values and the values given by H-H model

Parameters related to neuron signals	Values given by H-H model	Estimated by PSO	Estimated by FA
g_{Na}	120	180	150
g_K	36	22	30
g_l	0.3	0.2	0.3
E_{Na}	-115	-111	-114
E_K	12	11	12
E_l	-10.613	-10	-10.6
I	0.1	0.1	0.09

4 Conclusion

Parameters related to action potential are optimized using Firefly and PSO algorithm. It has been seen that firefly method is more efficient. It takes lesser time in evaluation of the optimized value. Moreover, the value estimated by firefly method is closer to the value given by H-H model. Estimation of parameters is significant in biomedical field since it helps to find the relationship between the parameters.

References

1. Lapique.: Introduction of the integrate-and-fire model of neuron. *J. Physiol. Pathol. Gen.* **9**, 620– 635 (1907)
2. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. physiol.* **117**(4), 500–544 (1952)
3. Hodgkin, A.L., Huxley, A.F.: Ionic movements and electrical activity in giant nerve fibers. *Proceedings of the Royal Society of London, Series B, Biological Science* **148**, 1–38 (1957)
4. Vavoulis, D.V., Straub, V.A., Aston, J.A.D., Feng, J.: A self-organizing state—space—model approach for parameter estimation in Hodgkin-Huxley—Type models of single neurons. *PLOS Comput. Biol.* **8**(3) (2012)
5. Buhry, L., Pace, M., Saighi, S.: Global parameter estimation of an Hodgkin–Huxley formalism using membrane voltage recordings: application to neuro-mimetic analog integrated circuits. *Neurocomput.* **81** (2012)
6. Buhry, L., Saighi, S., Giremus, A., Grivel, E., Renaud, S.: Parameter estimation of the Hodgkin-Huxley model using metaheuristics: application to neuromimetic analog integrated circuits. In: *IEEE Biomedical Circuits and Systems Conference*, pp. 173–176 (2008)
7. Buhry L., Giremus A., Grivel E., Saighi S., and Renaud S.: New variants of the differential evolution algorithm: application for neuroscientists. In: *17th European Signal Processing Conference*, pp. 2352–56 (2009)
8. Crotty, P., Sangrey, T.: Optimization of battery strengths in the Hodgkin-Huxley model. *Neurocomputing* **74**, 3843–3854 (2011)
9. Willms, A.R., Baro, D.J., Harris-Warrick, R.M., Guckenheimer, J.: An improved parameter estimation method for Hodgkin-Huxley models. *J. Comput. Neurosci.* 145–68 (1999)
10. Cedersund, G., Samuelsson, O., Ball, G., Tegnér, J., Gomez-Cabrero, D.: Optimization in biology parameter estimation and the associated optimization problem. *Uncertainty in Biology a Computational Modeling Approach*: Springer Publishing, Berlin (2016)

A Model to Demonstrate the Universality of DNA-NAND Gate

Kuntala Boruah, Rashmi Deka and Jiten Ch. Dutta

Abstract A model is proposed to demonstrate the universality of DNA-NAND gate at molecular level. NAND gate is one of the universal logic gates as any other gates can be obtained from it alone. A theoretical proof of principle simulation is presented in this paper to derive other basic logic gates with the help of DNA-NAND gate. Well established DNA operations are used as tools during the entire simulation.

Keywords DNA · DNA computing · NAND gate · Universal gate

1 Introduction

Deoxyribonucleic acid (DNA) computing is a newly evolved nonconventional computational paradigm which uses the information processing capability of DNA. The first implementation experiment was performed by Adleman in 1994 [1]. He successfully presented a molecular approach to solve an instance of 7-noded Travelling salesman's problem (TSP). In the following years, tides of ideas and models were proposed to solve several computational problems. Initially tremendous efforts were put into solving NP-Class problems but with time several other branches and sub-branches of research emerged. Area of simulating biochemical Boolean circuits and logic gates was stated after the publication of Ogihara and Ray in 1997. They demonstrated an AND-OR circuit with run time proportional to size of the circuit [2]. Since then numbers of models have been proposed to develop a robust, efficient, scalable, and reusable circuits using different biochemical properties as computational tools [3–12].

K. Boruah (✉) · R. Deka · J.Ch. Dutta
Department of Electronics and Communication Engg,
Tezpur University, Tezpur, Assam, India
e-mail: kuntala17@gmail.com

R. Deka
e-mail: rashmeedeka10@gmail.com

In this paper the authors have proposed a simulation process to recreate and verify the universality of DNA-NAND gate at molecular scale. Emphasis is given to reduce the number of biochemical operations involved in the simulation process.

2 DNA-NAND Gate Design

During the gate design process eight unique DNA strands of length 6 nucleotides each has been preassigned to the input variables depending on its associated values, i.e., for $X = 0$, $X = 1$, $Y = 0$, and $Y = 1$ different sequences have been assigned (shown in Table 1). While designing gates, all the rows of the corresponding truth table are scanned and examined sequentially. Depending on the output value, the gate strand is designed. When a row has output = '1' or 'True', the corresponding DNA sequences associated with the input variables present in that row is ligated one after another in linear fashion. Whereas when output value is '0', the ligation process is skipped for that particular row and continues scanning the next row until the end of the table is reached.

In case of NAND gate (shown in Table 2) the first, second, and third rows have output value '1', therefore while designing the DNA-NAND gate, the associated nucleotide sequences corresponding to each variable values are tailored together for three rows excluding the fourth row which has output '0'.

The designed DNA-NAND with $3' \rightarrow 5'$ orientation obtained by the gate design strategy is shown below.

3'-GCTATTGACTTAGCTATTGTAACATCTGACGACTTA-5'

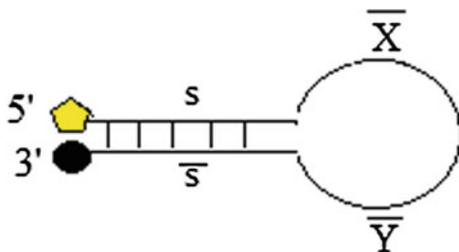
Input design is an important part of the simulation process. During the simulation process single-stranded DNA sequence is designed as inputs which are equivalent to digital input values. For a two-variable truth table there are four digital input combinations, i.e., (0,0), (0,1), (1,0), and (1,1). DNA strand is designed in such a way that apart from reproducing the digital inputs it also helps in final readout of outputs. The input strand consists of two parts; digital information part and self-complementary sequences at its two ends. The digital information part is

Table 1 Preassigned sequence

X	$3' \rightarrow 5'$	\bar{X} $5' \rightarrow 3'$	Y	$3' \rightarrow 5'$	\bar{Y} $5' \rightarrow 3'$
1	TCTGAC	AGACTG	1	GTAACA	CATTGT
0	GCTATT	CGATAA	0	GACTTA	CTGAAT

Table 2 NAND gate truth table with inputs X and Y

X	Y	Output
0	0	1
0	1	1
1	0	1
1	1	0

Fig. 1 Input strand**Table 3** DNA input strands

Inputs	Derived Input
0,0	5' S-CGATAACTGAAT- \bar{S} 3'
0,1	5' S-CGATAA CATTGT- \bar{S} 3'
1,0	5' S-AGACTG CTGAAT- \bar{S} 3'
1,1	5' S-AGACTG CATTGT- \bar{S} 3'

encoded by ligating the complements of associated variable sequences. The self-complementary sequences are represented by 'S' and ' \bar{s} ' and are assigned with unique self-complementary DNA sequences of 6 nucleotides each. The input strands have the tendency to self-hybridize to obtain hairpin formation. A fluorophore and a quencher attached in its two 5' and 3' ends of input strand due to which they can act as molecular beacon (MB) (shown in Fig. 1). The DNA MB equilibrated between ON and OFF conformation depending on the presence or absence of target molecular sequence in the gate strand. Table 3 shows the designed input sequence equivalent to the digital inputs.

During the execution steps, the gate strand and the input strands are allowed to anneal. The loop part of the input strand undergoes hybridization with the gate strand on finding its target sequence and hence opens up. Because of the conformational change fluorescence donor and fluorescence acceptor separate and restore the fluorescence emission. Therefore, the successful hybridization can be detected by emission of fluorescence and the output is readout as '1' or 'True'. In other case when the hybridization is unsuccessful, the input beacon does not open up and hence there is no event of transformation of darkness to brightness and in such case the output is read as '0' or 'False'. In Sect. 3 of this paper the above-mentioned technique is used for evaluation of NAND gate and reading out the output.

3 Universality of DNA-NAND Gate

3.1 Realization of AND Gate with DNA-NAND Gate

AND gate can be derived from NOT of NAND gate as shown in Fig. 2. There are two NAND gates integrated in such a fashion that the output of first gate acts as input to the second gate. The digital property of AND gate can be reproduced at

molecular level with the help of DNA-NAND gate shown in Figs. 3, 4, 5, 6 and 7. The input strand and the gate strands are allowed to undergo hybridization and the output obtained at g_1 again acts as an input to the g_2 . Depending on four input combinations the four cases are shown as Case 1, Case 2, Case 3, and Case 4.

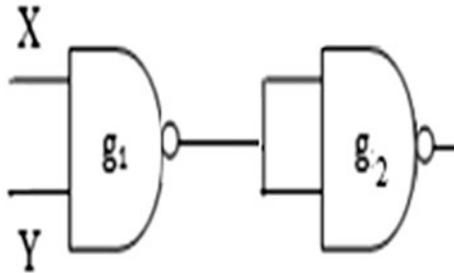


Fig. 2 AND gate from NAND

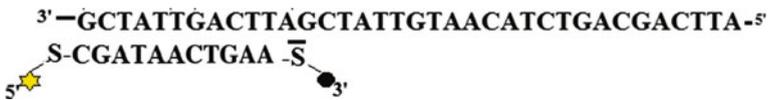


Fig. 3 Reaction at Gate 1

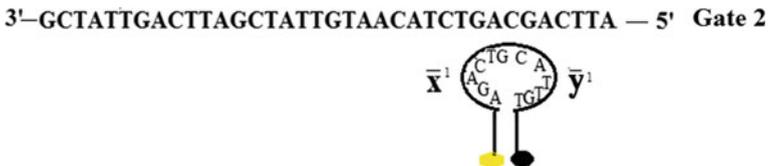


Fig. 4 Reaction at Gate 2

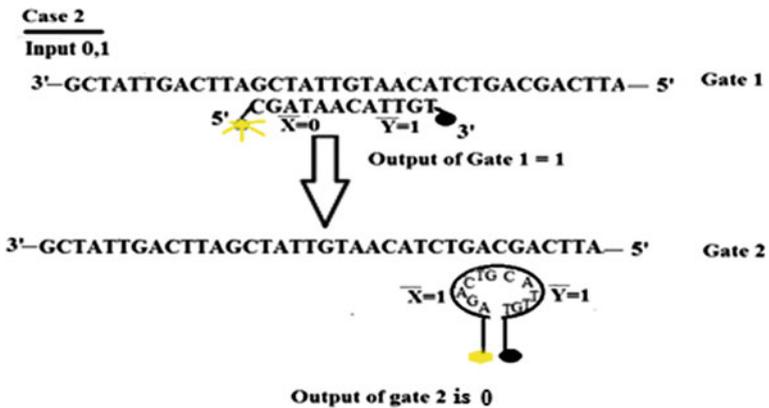


Fig. 5 Simulation of NAND gate as AND gate for input (0,1)

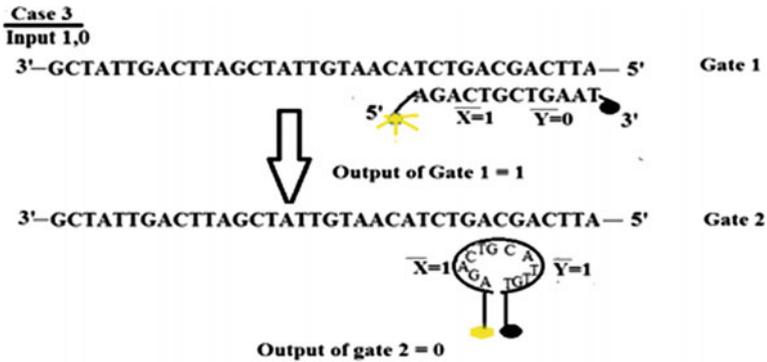


Fig. 6 Simulation of NAND gate as AND gate for input (1,0)

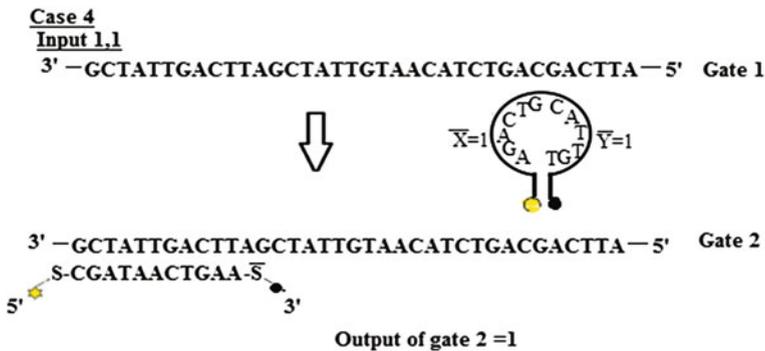


Fig. 7 NAND as AND gate for input (1,1)

Case 1: When $X = 0$ and $Y = 0$, i.e., inputs are (0,0), the equivalent DNA input sequence obtained is 5'-S-CGATAA GACTTA- \bar{S} -3' (shown in Table 3).

Gate 1: The input strands and the gate strand are complementary, hence they successfully hybridize producing the output of g_1 as 'TRUE' or '1'. Inputs to g_2 are obtained from output of g_1 .

Reaction at gate 1(g_1):

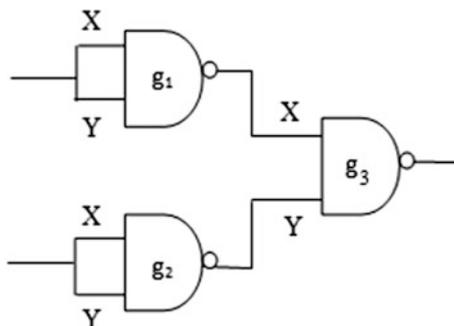
Gate 2: Since the output of g_1 is '1', therefore the input to g_2 is also '1'. While translating digital input to DNA input '1' is considered as (1,1). From Table 3, the DNA equivalent of (1,1) is 5'-S-AGACTGGTAACA- \bar{S} -3' strand.

The reaction at g_2 is shown in Fig. 4.

In gate 2 simulation no segment of gate strand and input strand are complementary, hence no hybridization occurs. Due to none emission, the output is evaluated as 0 or 'false'. Therefore, it can be summed up by saying that whenever the input is (0,0), then the final output is '0' which is in agreement with digital AND gate.

Table 4 Digital and simulated output (AND)

X	Y	Gate 2 output	X and Y
0	0	Hairpin, dark	0
0	1	Hairpin, dark	0
1	0	Hairpin, dark	0
1	1	No hairpin, bright	1

Fig. 8 OR gate obtained from NAND gate

Similarly, simulation and validation processes are carried out for all the digital input cases, i.e., Case 1: (0,1), Case 2: (1,0), and Case 3: (1,1). Table 4 shows the summary of simulation process and the output of digital AND gate (Figs. 5, 6 and 7).

3.2 NAND to OR Gate

In digital electronics NAND gate can be used to obtain the functionality of OR gates. In this model author shows a simulation process at molecular level to reflect the derivability of OR functionality with the help of DNA-NAND gate. Figure 8 shows the digital circuit to obtain OR gate and Figs. 9, 10, 11 and 12 show the molecular simulation. Gate 1 and gate 2 are separately simulated on the basis of the inputs as shown in Case 1, Case 2, Case 3, and Case 4. The output from g_1 to g_2 acts as input to g_3 . Table 5 shows the comparison of simulation output to the digital OR gate output.

3.3 NAND Gate as NOT Gate

Whenever the two inputs of NAND gate are joined together to a same input value, then it acts as NOT gate (shown in Fig. 13). Though it is digitally well established but in molecular level this theoretical validation is shown in Figs. 14 and 15.

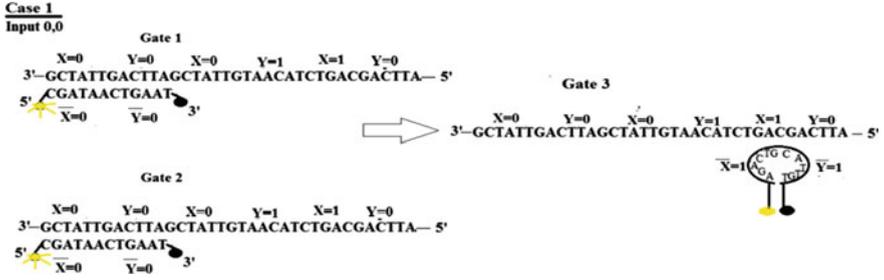


Fig. 9 Simulation of NAND gate as OR gate for input (0,0)

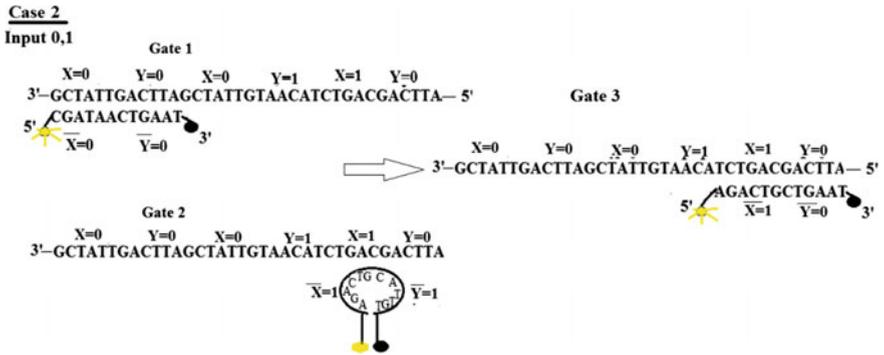


Fig. 10 Simulation of NAND gate as OR gate for input (0,1)

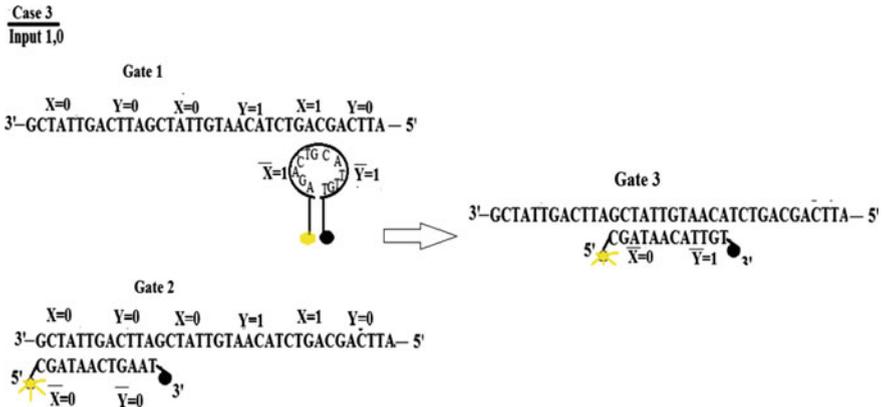


Fig. 11 Simulation of NAND gate as OR gate for input (1,0)

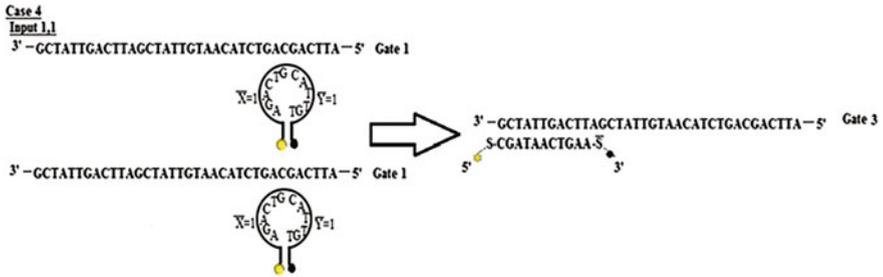


Fig. 12 Simulation of NAND gate as OR gate for input (1,1)

Table 5 Digital and simulation model output of OR gate

X	Y	Gate 3 output	X OR Y
0	0	Hairpin, dark	0
0	1	No hairpin, bright	1
1	0	No hairpin, bright	1
1	1	hairpin, bright	1

Fig. 13 NAND gate as NOT gate

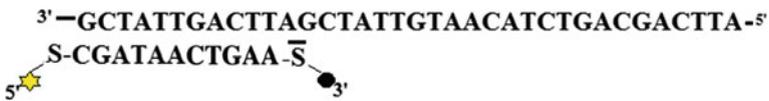
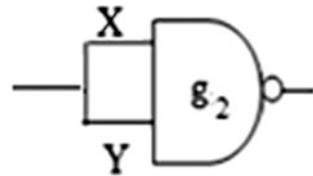


Fig. 14 Simulation of NAND gate as NOT gate for input 0

There are two cases; case 1 and case 2. In case 1, both the inputs are '0' or 'false', therefore the DNA equivalent input strand is 5' S-CGATAACTGAA-T \bar{S} 3' (Table 3). Since the gate strand and the input strand are complementary, hence hybridizes and results in fluorescence signal. Table 6 shows the comparison of simulation output to the digital NOT gate output.

Case 1: Similarly, Case 2 shows the simulation process of input 1. Since the gate strand and the input strand do not hybridize the output is read as 0.

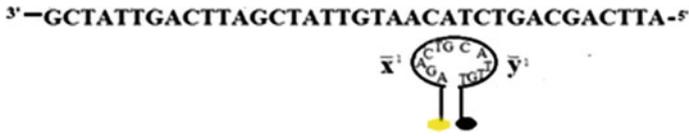


Fig. 15 Simulation of NAND gate as NOT gate for input 1

Table 6 Digital and simulation model output of NOT gate

Input	Gate 2 output	NOT
0	No hairpin, bright	1
1	hairpin, bright	0

Case 2: See Fig. 15 and Table 6.

4 Conclusion

In this paper, a theoretical model has been proposed to design and simulate a two-input DNA-NAND gate at molecular level and to theoretically verify the universality of the proposed gate by conceptual simulation of OR, AND, and NOT gate. Series of inputs are provided in the form of DNA MB probe such that the loop carries the information equivalent to digital input values. The simulation process is based on sequence specificity of hybridization reaction between input MB and gate strands. It is concluded that the proposed molecular gate is universal and retains the digital property but involves substantial amount of human intervention which affects the efficiency of the model. In coming future works it must be carried out to overcome the limitation and increase the degree of automation.

Acknowledgements The authors would like to acknowledge the Tezpur University for providing necessary facilities.

References

1. Adleman, L.: Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994)
2. Ogihara, M., Ray, A.: DNA-based self-propagating algorithm for solving bounded-fan-in Boolean circuits. *Genet. Program.* **98**, 725–730 (1998)
3. Amos, M., Dunne, P.E., Gibbons, A.: DNA simulation of Boolean circuits. In: Proceedings of 3rd Annual Genetic Programming Conference, pp. 679–68 (1997)
4. Erk, K.: Simulating Boolean circuits by finite splicing. In: Proceedings of the Congress on Evolutionary Computation, vol. 2, IEEE (1999)

5. Mulawka, J.J., Piotr, W., Plucienniczak, A.: Another logical molecular NAND gate system. In: Proceedings of the 7th International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems, IEEE, pp. 340 (1999)
6. Ahrabian, H., Ganjtabesh, M., Nowzari-Dalini, A.: DNA algorithm for an unbounded fan-in Boolean circuit. *Biosystems* **82**, 52–60 (2005)
7. Liu, W., Shi, X., Zhang, S., Liu, X., Xu, J.: A new DNA computing model for the NAND gate based on induced hairpin formation. *Biosystems* **77**(1), pp. 87–92 (2004)
8. Kadkhoda, M., Pouyan, A.: A DNA-based simulation model for bounded fan-in Boolean circuits. In: Proceedings of the 10th WSEAS International Conference on Computers. World Scientific and Engineering Academy and Society (WSEAS), pp. 1231–1235 (2006)
9. Shapiro, E., Gil, B.: Biotechnology: logic goes in vitro. *Nat. Nanotechnol.* 84–85 (2007)
10. Gearheart, C.M., Rouchka, E.C., Arazi, B.: DNA-based active logic design and its implications. *J. Emerg. Trends Comput. Inform. Sci.* (2012)
11. Zoraida, B.S.E., Arock, M., Ronaldet, B.S.M., Ponalagusamy, R.: A novel generalized design methodology and realization of Boolean operations using DNA. *Biosystems* **97**, 146–153 (2009)
12. Boruah, K., Dutta, J.: Development of a DNA computing model for Boolean Circuit. In: 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), IEEE (2016)

Micro-Raman and FTIR Analysis of Silicon Carbo-Nitride Thin Films at Different H₂ Flow Rate

Dhruva Kumar, Umesh Rizal, Soham Das, Bhabani S. Swain and Bibhu Prasad Swain

Abstract Silicon carbo-nitride thin films were deposited on Si (100) substrate by thermal chemical vapour deposition using C₂H₂ and Si powder precursors. The thin films were characterized by scanning electron microscope (SEM), Fourier transform infrared spectroscopy and Raman spectroscopy. The FTIR spectra reveals the presence of vibration signature of various bonds at 512, 1135, 1688, 2444, 3032, 3550 cm⁻¹ which correspond to Si-N, SiC-N, C-N, Si-H, C-H and N-H, respectively, in the SiCN thin films. Raman spectra reveal the presence of three prominent stoke shifts at 617, 1141 and 1648 cm⁻¹ corresponding to Si-H, SiC-N and C-C respectively. The vibrational signature of SiC-N shifted from 1126 to 1050 cm⁻¹ with increase in H₂ flow rate indicates formation of nanosized cluster in deposited thin film.

Keywords Thermal CVD · SiCN · Raman · FTIR

D. Kumar (✉) · U. Rizal · S. Das · B.P. Swain
Department of Mechanical Engineering,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: dhruva.ju@gmail.com

S. Das
e-mail: sohamengg10@gmail.com

B.P. Swain
e-mail: bibhuprasad.swain@gmail.com

D. Kumar · S. Das
Centre for Material Science and Nano Technology,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India

B.S. Swain
Advanced Material Processing, Kookmin University, Seoul, Republic of Korea

1 Introduction

The silicon carbo-nitride (SiC_xN_y) films have caught the attention of researchers across the world due its excellent physical, chemical and mechanical properties making it a promising material to be used for number of operations like thermal protective coatings, wear resistive coating, production of electronic devices, high-temperature semiconductors, electromechanical sensors for harsh environments, anode materials for lithium-ion batteries. [1–6]. Moreover, its optical band gap tunes between 1.1 and 5.5 eV concerning carbon and nitrogen alloying. Therefore, the selecting application mainly depends on chemical network of distributed sp^2 and sp^3 bonds carbon atoms and surrounding nitrogen and silicon in the SiCN.

To great extent the quality of thin film depends on the deposition techniques, CVD is one of the most popular methods to prepare silicon carbonitride films in which the deposition conditions determine the phase and chemical composition of the films. [7]. Raman spectroscopy is used for structural analysis of crystalline and amorphous films. Raman analysis has shown that Si–N, Si–Si, Si–H, Si–C–N bonds were found in the films. Hydrogen is diatomic light gas which can diffuse through any material during the synthesis process, so H_2 dilution plays an important role in transformation from amorphous to microcrystalline and polycrystalline silicon carbonitride as it prevents polymerization reactions. The effects of hydrogen dilution have been also studied and reported [8]. The present paper talk over the Raman studies of SiCN thin films deposited at different H_2 flow rates. The change in H_2 flow rate results in change of phase and bonds of the deposited film which are key components in determining the physical, mechanical and microstructural behaviour of the films. A correlation and illustration of the Raman spectral observations has been concluded.

2 Experimental Details

SiCN thin films were synthesized by using SiC and SiN powder as source and N_2 and H_2 mixture as precursor gases in a thermal CVD reactor. H_2 flow rate was varied between 80 and 120 sccm with N_2 flow rate fixed at 120 sccm. Mixture of SiC and SiN powder were taken and placed at the hot zone which is in the middle of horizontal furnace. Si (100) substrate was cleaned by standard RCA cleaning procedure and was placed at downstream to the ceramic boat. Residual gases were removed by pumping the reactor to subatmospheric pressure. Nitrogen was flown into the reactor for 2 h in order to remove other residual gases. The processing temperature of the experiment was 1300 °C. The heating rate of the furnace was 5 °C/min. The duration of deposition was fixed at 3 h. After completion of deposition, furnace was cooled down gradually at the rate of 3 C/min. The microstructures were observed by Field Emission Scanning Electron Microscopy

(FESEM) (Model: JEOL-JEM-3000F). The chemical network was analyzed by Raman and FTIR. Horiba Jobin Yvon was used at room temperature for characterizing the Raman spectrum in the range of 150–1850 cm^{-1} using the 488 nm line of an Ar+ laser as an excitation source. Perkin Elmer spectrometer (Model: Spectrum 2) was used for collecting FTIR spectra which was operating in transmission mode between 450 and 3000 cm^{-1} having a resolution of 1 cm^{-1} .

3 Results and Discussion

3.1 Microstructure and Morphology

See Fig. 1

3.2 Raman Analysis

Figure 2 shows the broad Raman spectra of SiCN film deposited at 80 and 120 sccm H_2 flow rate. Gaussian curve fitting of the Raman bands was performed. Raman spectra reveal the presence of three prominent stoke shifts. The signature appearing at 520 cm^{-1} corresponds to Si–Si network. The signature observed at 620 cm^{-1} corresponding to Si–Hn rocking which is due to (TO + TA) vibration of Si–Si network, moreover it also represents the amount of bonding H_2 present in the SiCN films. The vibrational signature appeared at 1150 cm^{-1} corresponding to SiCN asymmetric stretching. The signature from 1625 to 1675 cm^{-1} assign for

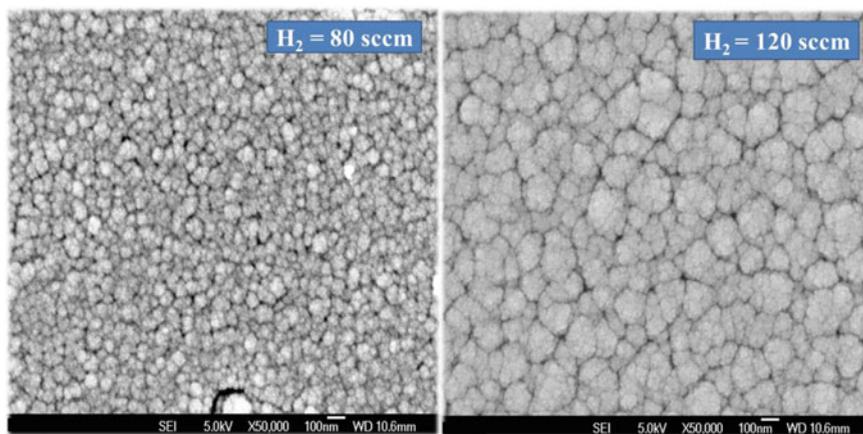


Fig. 1 SEM figures of silicon carbo-nitride thin films deposited at $\text{H}_2 = 80$ sccm and $\text{H}_2 = 120$ sccm flow rate respectively

Fig. 2 Raman spectra of SiCN thin films with 80 and 120 sccm H_2 flow rates

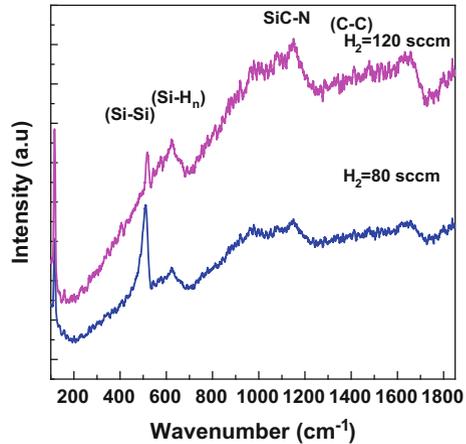
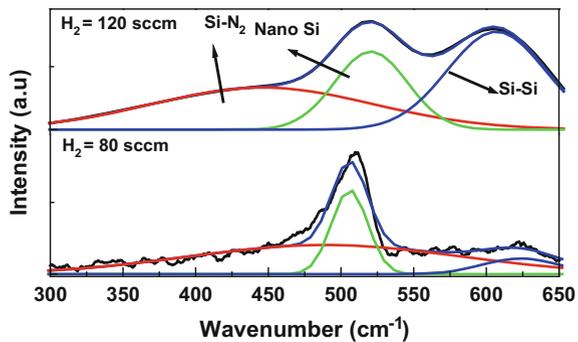


Fig. 3 Deconvolution of Raman spectra of SiCN thin films at the range of 300–650 cm^{-1} with different of H_2 flow rate



C=C–Si and C=C–N bonds [7]. The shift of vibrational signature of SiC–N from 1126 to 1050 cm^{-1} with increase in H_2 flow rate indicates formation of nanosized cluster deposited in thin film. The signature appearing at 810, 880, 940, 1000, 1065 cm^{-1} are from Si–C stretching, Si–H₂ rocking, Si–N asymmetric stretching [7], SiC–H₃ wagging, SiC–N stretching and NH₂ stretching respectively.

Figure 3 shows the deconvolution of Raman spectra of SiCN thin films at the range of 300–650 cm^{-1} with different of H_2 flow rates. The vibration signatures of Si–Si network and Si–N are generally found in the region 300–650 cm^{-1} [7]. Hence we have deconvoluted the region into three individual Gaussian peaks for the 80 and 120 sccm H_2 flow rate SiCN thin films. The Raman signature of Si–N is found out in the range from 489.14 to 494.5 for increased of H_2 flow rate from 80 to 120 sccm. The Raman signature for crystalline Si increased monotonically from 505.95 to 520.6 cm^{-1} for increasing of H_2 flow rate from 80 to 120 sccm. The Raman signature for Si–Si bond appeared from 606.91 to 623.7 cm^{-1} from 80 to 120 sccm H_2 flow rate respectively [9]. It is observed that, with increase in H_2 flow rate the peaks shifted towards right hand side which represents the growth of crystallinity in the thin film.

Fig. 4 Deconvolution of Raman spectra of SiCN thin films at the range of 700–1275 cm^{-1} with different H_2 flow rates

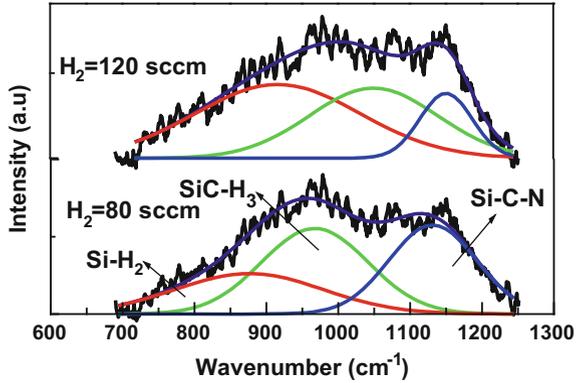


Fig. 5 Deconvolution of Raman spectra of SiCN thin films at the range of 1250–1750 cm^{-1} with different H_2 flow rates

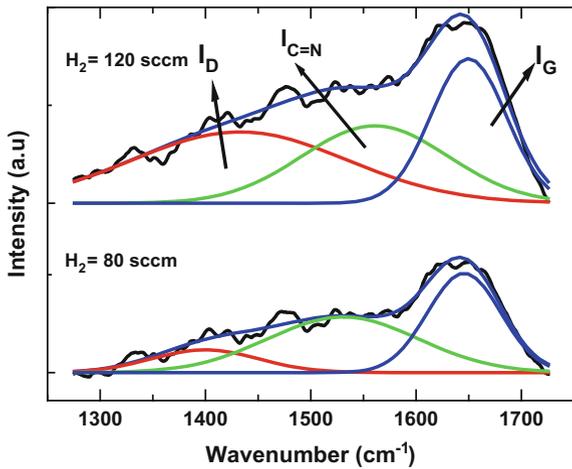


Figure 4 shows the deconvolution of Raman spectra of SiCN thin films at the range of 700–1275 cm^{-1} with different H_2 flow rates. The vibration signatures of Si-H₂, Si-N and SiC-N signatures are generally found in the region 700–1275 cm^{-1} [7]. Hence we have deconvoluted the region into three individual Gaussian peaks for the 80 and 120 sccm H_2 flow rate SiCN thin films. The Raman signature of Si-H₂ is found out in the range from 916.7 to 967.4 for increased of H_2 flow rate from 80 to 120 sccm. The Raman signature of SiC-H₃ appeared from 875.95 to 1049.9 cm^{-1} from 80 to 120 sccm H_2 flow rate respectively. [7]. The Raman signature for SiC-N increased monotonically from 1131.3 to 1150.8 cm^{-1} for increasing H_2 flow rate from 80 to 120 sccm. [7]

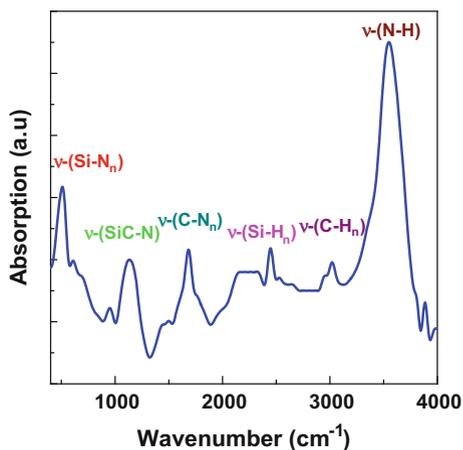
Figure 5 shows the deconvolution of Raman spectra of SiCN thin films at the range of 1250–1750 cm^{-1} with different H_2 flow rates. The vibration signatures of

DLC, C=N and GLC signatures are generally found in the region $1250\text{--}1750\text{ cm}^{-1}$ [7]. Hence we have deconvoluted the region into three individual Gaussian peaks for the 80 and 120 sccm H_2 flow rate SiCN thin films. The Raman signature of DLC is found out in the range from 1398 to 1403 for increased of H_2 flow rate from 80 to 120 sccm. The Raman signature of C=N increased monotonically from 1547.3 to 1553.5 cm^{-1} for increasing of H_2 flow rate from 80 to 120 sccm. The Raman signature for GLC was observed from 1647.2 to 1650.1 cm^{-1} from 80 to 120 sccm H_2 flow rate respectively. [7]. The absence of $\text{C}\equiv\text{N}$ represents good interatomic linking which will result in good mechanical properties [10].

4 FTIR Analysis

Figure 6 shows the absorption FTIR spectra of SiCN film deposited at 80 sccm H_2 flow rate. Study of effect of the deposition conditions on the film's composition is generally performed by FTIR spectroscopy measurements. A large band at about 512 cm^{-1} was observed associated to the Si-N stretching mode. The vibration signature of N-Si-H, Si-N-H and Si-N-Si-N was found at 468.45, 505.52 and 532.2 cm^{-1} respectively. [11]. Vibration signature of SiC-N was revealed at 1135 cm^{-1} . A shoulder is clearly observed at about 1688 cm^{-1} which can be due to the C-N stretching mode. Si-H bond can be observed at 2440 cm^{-1} . The stretching modes of C-H and N-H can be observed at 3015 and 3547 cm^{-1} respectively [12]. Important information like stress, degree of disorder present in the film can be derived from deviation in peak position and full width half maxima (fwhm) values.

Fig. 6 FTIR spectra of SiCN thin films grown at 80 sccm H_2 flow rate



5 Conclusion

Figure 2 shows the broad SiCN thin film has been successfully prepared via thermal chemical vapour deposition technique. H₂ flow rate strongly influenced the morphology of SiCN. However, higher H₂ flow rate produces larger SiCN grain in the thin films. The chemical network study confirmed formation of nanostructure material.

References

1. Ryu, H., Wang, Q., Raj, R.: Ultrahigh-temperature semiconductors made from polymer-derived ceramics. *J. Am. Ceram. Soc.* **93**, 1668–1676 (2010)
2. Bhattacharyya, A., Mishra, S., Mukherjee, S., Das, G.: A comparative study of Si–C–N films on different substrates grown by RF magnetron sputtering. *J. Alloys Compd.* **478**, 474–478 (2009)
3. Bai, X., Zhang, C., Tao, A.: Superhard multilayers and nanocomposite coatings. *Prog. Nat. Sci.* **105**, 97–107 (2005)
4. Hoche, H., Pusch, C., Riedel, R., Fasel, C., Klein, A.: Properties of SiCN coatings for high temperature applications—comparison of RF-, DC- and HPPMS-sputtering. *Surf. Coat. Technol.* **205**, S21–S27 (2010)
5. Leo, A., Andronenko, S., Stiharu, I., Bhat, R.: Characterization of thick and thin film SiCN for pressure sensing at high temperatures. *Sensors* **10**, 1338–1354 (2010)
6. Chen, Y., Li, C., Wang, Y., Zhang, Q., Xu, C., Weid, B., An, L.: Self-assembled carbon–silicon carbonitride nanocomposites: high-performance anode materials for lithium-ion batteries. *J. Mater. Chem.* **21**, 18186–18190 (2011)
7. Mera, G., Riedel, R., Poli, F., Müller, K.: Carbon-rich SiCN ceramics derived from phenyl-containing poly (silylcarbodiimides). *J. Eur. Ceram. Soc.* **29**, 2873–2883 (2009)
8. Fainer, N., Golubenko, A., Rummyantsev, Y., Kesler, V., Ayupov, B., Rakhlin, V., Voronkove, M.: Tris (diethylamino) silane—a new precursor compound for obtaining layers of silicon carbonitride. *Glass Phys. Chem.* **38**, 15–26 (2012)
9. Wang, Y., Ding, J., Feng, W., An, L.: Effect of pyrolysis temperature on the piezoresistivity of polymer-derived ceramics. *J. Am. Ceram. Soc.* **94**, 359–362 (2011)
10. Mera, G., Tamayo, A., Nguyen, H., Sen, S., Riedel, R.: Nanodomain structure of carbon-rich silicon carbonitride polymer-derived ceramics. *J. Am. Ceram. Soc.* **93**, 1169–1175 (2010)
11. Chen, Y., Yang, X., Cao, Y., Gan, Z., An, L.: Quantitative study on structural evolutions and associated energetics in polysilazane-derived amorphous silicon carbonitride ceramics. *Acta Mater.* **72**, 22–31 (2014)
12. Prasad, R., Mera, G., Morita, K., Müller, M., Kleebe, H., Gurlo, A., Fasel, C., Riedel, R.: Thermal decomposition of carbon-rich polymer-derived silicon carbonitrides leading to ceramics with high specific surface area and tunable micro- and mesoporosity. *J. Eur. Ceram. Soc.* **32**, 477–484 (2012)

Applying Classification Methods for Spectrum Sensing in Cognitive Radio Networks: An Empirical Study

Nayan Basumatary, Nityananda Sarma and Bhabesh Nath

Abstract Spectrum sensing is the paramount aspect of cognitive radio network where a secondary user is able to utilize the idle channels of the licensed spectrum band in an opportunistic manner without interfering the primary (license) users. The channel (band) is considered to be idle (free) when primary signal is absent. The channel accessibility (free) and non-accessibility (occupied) can be modeled as a classification problem where classification techniques can determine the status of the channel. In this work supervised learning techniques is employed for classification on the real-time spectrum sensing data collected in test bed. The power and signal-to-noise ratio (SNR) levels measured at the independent CR device in our test bed are treated as the features. The classifiers construct its learning model and give a channel decision to be free or occupied for unlabelled test instances. The different classification technique's performances are evaluated in terms of average training time, classification time, and F1 measure. Our empirical study clearly reveals that supervised learning gives a high classification accuracy by detecting low-amplitude signal in a noisy environment.

Keywords Cognitive radio · Spectrum sensing · Primary user detection
Supervised learning techniques

N. Basumatary (✉) · N. Sarma · B. Nath
Department of Computer Science and Engineering, Tezpur University,
Napam, Assam, India
e-mail: nayanb@tezu.ernet.in

N. Sarma
e-mail: nsarma@tezu.ernet.in

B. Nath
e-mail: bnath@tezu.ernet.in

1 Introduction

Cognitive Radio (CR) is the emerging technology in the domain of new age wireless communication. It can dynamically change its transmission parameters based on changes in environmental factors [1]. In cognitive radio network a secondary or unlicensed user can sense the licensed channels for any opportunity to transmit, which results to efficiently utilize the available channel of primary licensed users. For performing this spectrum access in an opportunistic manner the CR devices need to sense the radio spectrum licensed to primary users. So, efficient spectrum sensing is very important for opportunistic spectrum access. In Cognitive Radio [2, 3] the spectrum sensing is carried out in a co-operative and independent manner. In co-operative sensing all the CR devices co-operate with each other to take a collective decision which results into get high sensing reliability. While, in case of independent sensing each CR device performs the sensing individually and make its own sensing decision to use unoccupied spectrum portion. Here, in this work analysis of the prominent supervised learning techniques [4–6] was done for noncooperating spectrum sensing framework to decide the presence or absence of primary user in a channel.

In low SNR environment (fading channels) where there is high noise level and regardless of the fact that there is a signal present (low amplitude) it cannot be distinguished. This work exploits the signal power and the SNR feature to take a decision in such condition. The conventional energy detection method may cause misdetection of the signal as it fails in a low SNR environment. The motivation of using supervised learning [7] is that in supervised learning the classifier learns from some objects which are having some class marks and when unknown object's class is to be predicted this class marks are assigned to unknown objects based on learning done previously which actually gives high detection accuracy. Here all supervised learning models are built not only based on just the power received of the signal, but also the SNR feature so that even if there is a low power signal in a highly noisy environment the classifier can still give a decision to detect the signal with a priori knowledge.

Contribution:

1. The customized dataset was created by capturing both the power and SNR features in USRP-based test bed for performing the classification task.
2. The SNR or the Signal-to-Noise ratio value used for classification for detection of signal in low SNR environment which is not explored in literature. The classifier decision is based on SNR parameter value which helps in detection of low amplitude signal.
3. Due to high prediction accuracy analysis was carried out using known supervised learning techniques such as the SVM (Linear, Poly, and RBF kernels), Logistic regression, K-Nearest Neighbor, Gaussian Naives Bayes, and Decision Trees.
4. The performance of the classification techniques was evaluated in terms of training time, classification delay, and F1 measure.

The remaining sections of the paper are as follows. In Sect. 2 the system model and the assumptions are presented. Then the supervised learning techniques in a noncooperative framework are discussed in Sect. 3. Section 4 describes the experimental test bed setup to amass the sensing data and prepare the data set. The numerical results and plots are given in Sect. 5. Finally, Sect. 6 gives the conclusion and future work to be done.

2 Framework Model and Presumptions

Here, a homogeneous Cognitive Radio network is assumed which consist of both the Primary and Secondary Network which is formed by ' N ' number of primary users denoted as $\{PU_1, PU_2 \dots PU_N\}$. The primary users have a set of channels $C = \{1 \dots L\}$ that are used for its transmission. The channel occupancy states are assumed to be independent and any channel can be used by the primary user. The secondary network consist of a set of ' K ' users called secondary users (SU) denoted by $\{SU_1, SU_2 \dots SU_K\}$ are also present in the CR network which are sensing these primary channels to access the channels in an opportunistic way. These users are equipped with some learning techniques to detect whether the channels are free and then to access these primary channels for its own transmission. In this secondary network each user independently searches for a spectrum opportunity. Also, it was assumed that each secondary user can sense only one channel at a time and they do not co-operate with each other. Both the primary and the secondary network are assumed to be synchronized with a global clock, and transmission is done in a time-slotted manner.

The spectrum sensing in Cognitive Radio essentially decides between two states of the channel whether there is signal present or absent. The features extracted from the samples received of the signal can be analyzed to decide between two hypotheses:

$$\begin{aligned} H_0 : x(i) &= n(i) && \text{(Signal is absent.)} \\ H_1 : x(i) &= s(i) + n(i) && \text{(Signal is present.)} \end{aligned}$$

The received samples of the signal represented by $x(i)$, the primary user signal $s(i)$ and the noise $n(i)$, where i denotes the i th sample. The signal transmission and the signal received are done in a continuous way. In this paper the signal is treated as discrete since the receiver or the sensor takes discrete samples of the signal.

Here, in case of indoor environment the primary transmission is in Line-of-sight propagation (LOS) between the transmitter antenna and the receiver antenna, hence the probability density function of the fast varying amplitude of the received instantaneous signal can be described by Rician distribution. The environmental noise is assumed to be AWGN (Additive white Gaussian Noise). The average path loss $PL(d)$ for a transmitter and receiver with separation distance d is

$$PL(d) = PL(d_0) + 10n \log(d/d_0),$$

where ‘ n ’ is path loss component and indicates the rate at which path loss increases with distance d , with close in reference distance d_0 .

In a Machine Learning framework the features extracted from the signal can be used for the feature vector and then build the model for both the training and testing phases of different classifiers. The performance of the classifier can then be separated in two hypotheses:

$$H_0 = \text{Incorrect Signal Classification.}$$

$$H_1 = \text{Correct Signal Classification.}$$

3 Supervised Learning for Spectrum Sensing

Here, some of known supervised learning techniques are applied for noncooperative spectrum sensing data in cognitive radio network. As known that CR devices have cognition capability to learn from the environment, supervised learning can be an effective way to extract information from the signals present or absent in the environment at the physical level and utilize this cognizance to make some decision in the upper hierarchy levels. The Support Vector machines (SVM), Logistic Regression (LR), K-nearest Neighbor (KNN), Decision Trees are the main classifiers used in a Nonco-operating Spectrum Sensing environment in CR Networks to detect the presence of the Primary User signal based on features of the collected data. Here, the two features power and SNR value forms an object $y^{(i)}$ and which is labeled with the corresponding channel availability CA_i .

In case of Supervised learning there are two phases, training and testing where the goal is to construct a classifier to map the samples received by the secondary users to that of the labeled samples.

1. *Training*: Let the training object is of the form $y^{(i)} = (P_i, SNR_i, CA_i)$ where P_i denotes the power value, SNR_i is the Signal to Noise Ratio and CA_i is corresponding labeled channel availability status in different time slots T_i . In the training phase this objects are fed to the classifier to build its model.
2. *Testing*: After the classifier is trained it is then ready to classify the test object of the form $y^{(i)} = (P_i, SNR_i)$ and then give the corresponding class label based on its model.

Therefore, to implement supervised learning in a Non Co-operating sensing environment, the classifier should be informed about the channel status information for some object values for the purpose of training.

4 Experimental Setup

In a cognitive radio network (CRN) in which SU can access free channel/band when unutilized by Primary users (PU). The CR devices in the network are the secondary users who are sensing the primary channels in a noncooperative manner and making its own individual decision. The secondary users are equipped with a classification framework. The sensing data obtained by the secondary users are fed to the classification framework which helps the secondary user to determine whether the channel is free or not. The experimental setup for real time dataset generation is implemented with the help of GNU Radio and USRP devices [8–11]. In the setup, three PC designated as PC1, PC2 and PC3 connected with USRP1 with a RFX2400 daughter board by a USB cable. PC1 acts the transmitter and PC2 and PC3 are the receiver/sensing nodes (Fig. 1).

In GNU Radio the existing sample python scripts *usrp_spectrum_sense.py* and *benchmark_tx.py* was modified as *sensing.py* and *transmission.py* for sensing and transmission respectively.

Other program parameters in the transmission.py script are

The Sampling rate	1 Mega Samples
Modulation used	GMSK
Sub Channel Bandwidth	6.25 kHz
So, the no of fft bins collected in a particular Channel	160 (1MS/6.25e3)

The receiver which is tuned to the center frequency of the channel can sweep only 8 MHz channel Bandwidth due to the USRP1 daughter board constraint. Out of the total 160 bins 75% is taken and 25% is discarded from both the lower and upper cut frequency (12.5% each) of the channel. The program senses the power level from bin 20 to bin 140.

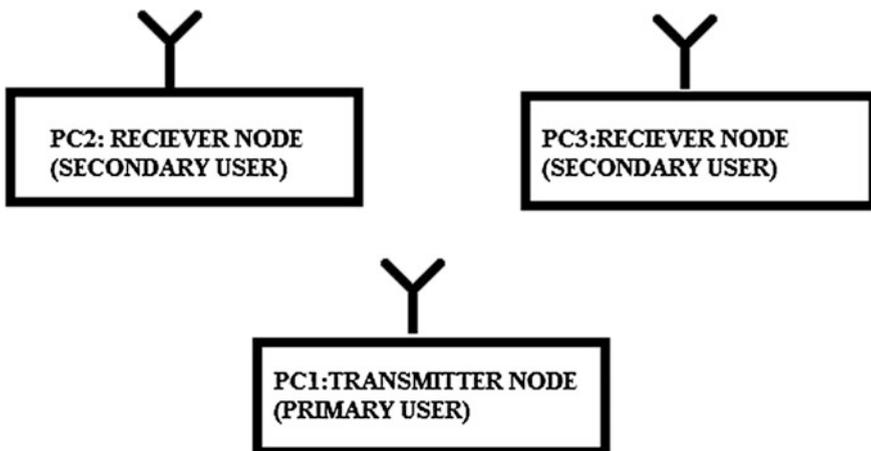


Fig. 1 Block diagram of the experimental setup

The sensing data was captured with the power and SNR features with active transmission and another with no transmission. All the sensing data are labeled as “**Free**” and “**Occupied**” class with respect to the known occupied and free channels respectively.

5 Numerical Results and Discussion

For carrying out the analysis the machine learning tool in python scikit learn [11] was used for performing the classification of the sensing data. Different Supervised Classification Algorithms that were used are Support Vector Machines with different kernels like Linear, Polynomial and RBF kernel, K-Nearest Neighbor (KNN), Logistic Regression (LR), Gaussian Naïve Baye’s (GNB) and Decision Trees (DT).

The total of samples collected from the test bed experiments was 600 after preprocessing. Cross validation random split into train and test subsets. Some percentage of the data is considered for training the classifier and the rest for testing purpose in an offline mode.

Table 1 shows the comparison of the average training time taken by different classifiers. The training time measured is high in case of SVM classifier with radial basis function (RBF) with the increase in percentage of the training samples. The Logistic Regression classifier also takes more time compared with the others.

Table 2 shows the comparison of the average classification time taken by the supervised classifiers, where the SVM classifier with radial basis function

Table 1 Average training duration of the supervised classifiers (seconds)

Training samples (%)	SVM (linear)	SVM (RBF)	SVM (polynomial-degree 3)	KNN	LR	Gaussia naïve bayes	Decision trees
50	0.001	0.004	0.001	0.001	0.001	0.001	0.001
60	0.001	0.005	0.001	0.001	0.001	0.001	0.001
70	0.001	0.006	0.001	0.001	0.001	0.001	0.001
80	0.001	0.008	0.001	0.001	0.003	0.001	0.001

Table 2 Average classification duration of the supervised classifiers (seconds)

Test samples (%)	SVM (linear)	SVM (RBF)	SVM (polynomial-degree 3)	KNN	LR	Gaussia naïve bayes	Decision trees
20	0.001	0.001	0.001	0.001	0.001	0.001	0.001
30	0.001	0.001	0.001	0.001	0.003	0.001	0.001
40	0.001	0.002	0.001	0.001	0.002	0.001	0.001
50	0.001	0.002	0.001	0.001	0.001	0.001	0.001

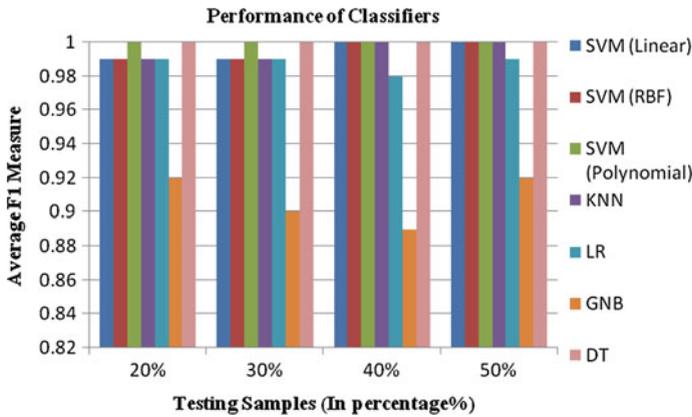


Fig. 2 Performance of classifiers with different number of testing samples and average F1 measure

(RBF) relatively takes more classification time than the other two kernel function. With the increase in percentage of the test samples the RBF classification time also increases than the others. It shows the comparison of the average classification time taken by the rest of the classifiers taken into consideration. The Logistic Regression classifier takes more time among others which are more or less having the same average time.

Figure 2 show performance of all classifiers used based on average F1 measure versus on the different percentage of test samples used for the classification. From the plot it is very clear that for all the different percentage of test samples used the SVM with polynomial kernel and decision tree performs very well than the rest of the classifiers which gives a higher primary user detection rate. The results also reveal that with the increase of percentage of the test samples the other SVM kernel functions along with KNN classifier performs well.

6 Conclusion and Future Work

In this work emphasis was on detection of primary user signal in a low SNR environment. An analysis was conducted using supervised learning techniques in a Noncooperating sensing manner based on real-time sensing data. The classifiers were trained in an offline manner to build its model and then decision for signal present or absent was done with different percentages of test samples. Compared to all the classifiers the SVM classifier with Polynomial and Decision Tree classifiers outperforms other techniques. The performance of all the classifiers was analyzed in terms of high detection rate like training duration and classification delay by using real-time dataset. The other classifiers may perform well with more number of samples in the dataset which may increase the sensing time of the secondary users.

But in a CR network if the sensing time is less to give a near-optimal detection rate of the primary user signal which is an additional advantage for secondary user communication (transmission time).

Future work could be to apply these learning techniques for cooperative spectrum sensing in cognitive radio network. This work is offline training and testing. If this can be extended to be implemented in an online fashion in real time it will be beneficial in a dynamic Cognitive Radio Network.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Select. Areas Commun.* **23**, 201–220 (2005)
2. Urkowitz, H.: Energy detection of unknown deterministic signals. In: *Proceedings of IEEE*, vol. 55, pp. 523–231 April 1967
3. Cabric, S.D., Mishra, S.M., Brodersen, R.W.: Implementation issues in spectrum sensing for cognitive radios. In: *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 772–776, 7–10 Nov (2004)
4. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn. J.* **20**(3), 273–297 (1995)
5. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley, New York (2001)
6. Thilina, K.M., Choi, K.W., Saquib, N., Hossain, E.: Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE J. Sel. areas commun.* **31**(11) 2013
7. Kassiny, M.B., Li, Y., Jayaweera, S.K., A survey on machine Learning techniques in cognitive radios. *IEEE Commun. Surv. Tutorials* **15**(3), 1136–1159 2013
8. Software Defined Radio Forum.: www.sdrforum.org
9. Ettus Research LLC.: <http://www.ettus.com/>
10. <http://www.gnu.org/software/gnuradio>
11. <http://scikit-learn.org/>

Highly Negative Dispersive, Low Loss Single-Mode Photonic Crystal Fiber

Shahiruddin, Dharmendra K. Singh and Sneha Singh

Abstract In this paper a new five-layer circular air holes photonic crystal fiber (PCF) having low dispersion and confinement loss is analyzed. A full-vector finite-element method approach is used. Due to proper selection of air-fill fraction, it has been made possible to obtain the two required properties of PCF for wide range of wavelengths, that is, low (nearly zero and high negative) dispersion and ultralow confinement loss of the order of 10^{-7} dB/km. It has been found that the proposed fiber is single mode. The confinement loss and dispersion is demonstrated for wavelength range which are 1–2 μm . Thus highly negative dispersion of PCFs useful for application in optical communication links as dispersion compensating fibers.

Keywords Confinement loss · Dispersion · Finite-element method
Photonic crystal fiber · Single-mode fiber · Single-polarization single-mode (SPSM) fibers

1 Introduction

Photonic crystal fibers (PCFs) are used as an alternative fiber technology. PCF is a new category of fiber optics that contains an arrangement of roughly wavelength-sized holes organized along the fiber optics immensely extends the possibilities of

Shahiruddin (✉) · S. Singh
Department of Electronics and Communication Engineering,
Birla Institute of Technology, Patna 800014, India
e-mail: shahir@bitmesra.ac.in

S. Singh
e-mail: snehasinghcit@gmail.com

D.K. Singh
Department of Electronics and Communication Engineering,
National Institute of Technology, Patna 800005, India
e-mail: dksingh@nitp.ac.in

optical fiber technology. The cladding was formed by using photonic crystal around the core. To achieve better waveguide in PCF the light is confined inside the core compared to traditional fiber optics. Single-polarization single-mode (SPSM) fibers were planned by using an elliptical core or using absorbent substance along the core. The bandwidth is around 180 nm is reported under single polarization [1], which may not be enough for several applications.

An extremely large refractive index variation is required between to make this class of fiber. The ability to carrying the information in optical communication system played an important role by dispersion. The PCF provide the liberty to controlled and tailored dispersion values. In a PCF with an infinite air hole in the cladding, propagation is theoretically lossless. However, in the fabricated fibers due to finite number of air holes the guided modes are leaky. As yet various types of low-dispersion PCF with low confinement loss have been proposed [2–4]. The negative dispersion of -203 ps/(km nm) of Ge-doped core PCF is reported [2]. The main drawback of such kind of PCF is fabrication difficulties due to doped core. At 1.55 μm the projected value of dispersion is -491.5 ps/(km nm) and confinement loss is 0.035 dB/km [3]. A PCF design is reported using an equiangular spiral form, which covers 1.36 to 1.675 μm , with negative dispersion of -393 ps/(km nm) [4]. The main impediment of such structure which is not realizable using established fabrication method.

In this work explore to investigate the potential of circular lattice PCF to provide a new technique for flexible organize of dispersion and confinement loss characteristics of PCF. An elaborated discussion has also been reported about influence circular rings and air fill fraction on dispersion and confinement loss.

2 Design Methodology of Single-Mode PCF

A five-ring single-mode PCF structure is shown in Fig. 1. The proposed PCF design is based on the circular structure. The whole span of fiber stand-in as cladding arranged with air holes across the core. The core refractive index is higher compared to cladding refractive index in this type of PCF. The designed fiber has five rings in cladding with solid core. The refractive index of pure silica is 1.45 chosen as a wafer and the air hole refractive index is 1 . The spacing between nearest air holes from center to center is representing as pitch (Λ) is taken as 2.3 μm . The designed PCF have five rings with the different parameters d (diameter) and constant Λ (pitch). The values of this parameter in the design are given in Table 1.

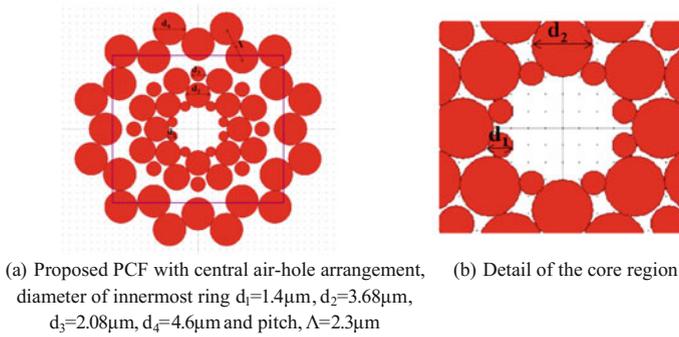


Fig. 1 Five-layer proposed structure of different diameter air holes PCF

Table 1 Design parameter of proposed structure

Design parameter	Magnitude (μm)	Applicable for
d_1	1.4	First ring
d_2	3.68	Second ring
d_3	2.08	Third ring
d_4	4.6	Fourth and fifth ring
Λ	2.3	Adjacent holes spacing from center to center

3 Mode Field Pattern of Proposed PCF

The field intensity distribution is shown in Fig. 2a–d of the single mode PCF where light is confined in core.

4 Simulation Results and Discussion

The boundary condition of PCF, perfectly matched, finite-element method has been used to look into the modal properties. The boundary condition is chosen effectively to framework leakage-free fiber and no reflection at the cladding portion. Dispersion represents a phenomena related to the velocity of electromagnetic wave depends on the wavelength.

Chromatic dispersion restricts the data transmission rates due to light pulses is spread and also affects the length of fiber links. The dispersion (D) represents as second derivative of the real part of n_{eff} , with respect to the wavelength (λ) is shown in Eq. 1:

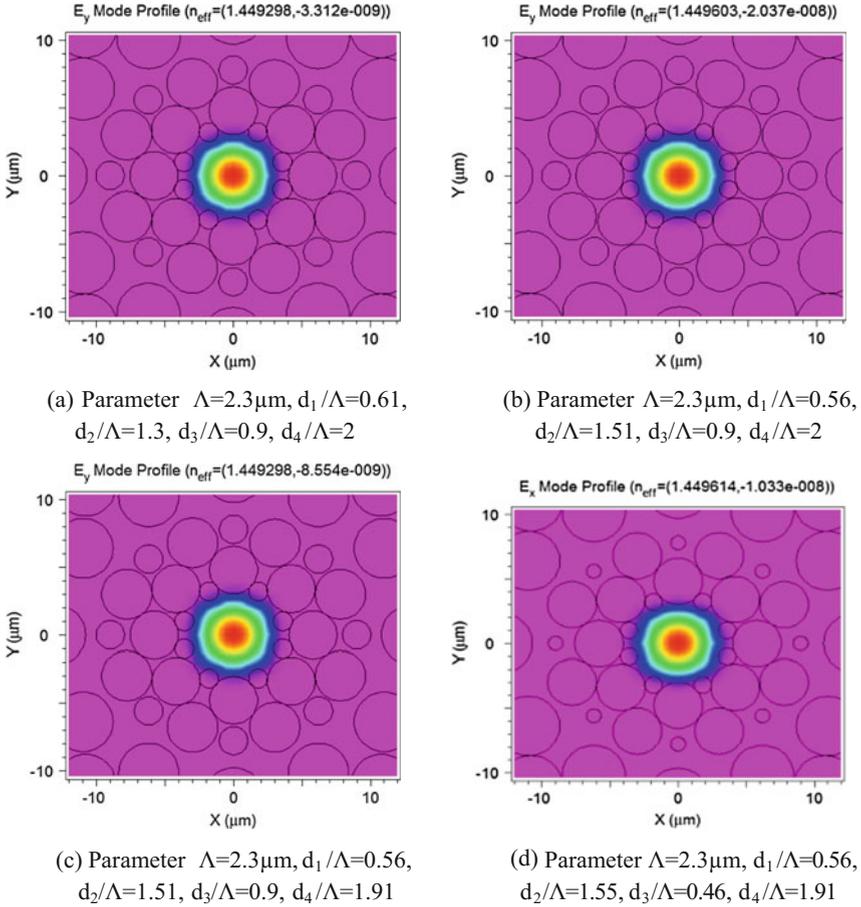


Fig. 2 Field intensity distribution of the single-mode PCF at 1.55 μm

$$D = -\frac{\lambda}{C} \frac{d^2 \text{Re}[n_{\text{eff}}]}{d\lambda^2}, \quad (1)$$

where λ is the wavelength, c is the velocity of light in vacuum, and $\text{Re}[n_{\text{eff}}]$ is the real part of n_{eff} [5–10].

The leakage losses or confinement losses associated with the PCFs depend on pattern of structure. Available options are wide effectively single-mode operation wavelength range [11] and non-polarization-degenerate effectively single-mode operation [12, 13]. The imaginary part of n_{eff} as in Eq. 2 is required to calculate confinement loss [14, 15]:

$$\text{Conf. Loss} = \frac{(20 \times 10^6)}{\ln(10)} K_0 \text{Im}[n_{\text{eff}}], \tag{2}$$

where $K_0 = 2\pi/\lambda$ is the wave number in free space, and $\text{Im}[n_{\text{eff}}]$ is the imaginary part of the refractive index.

In this work, a PCF design has been proposed with low confinement losses and ultralow chromatic dispersion at a wavelength 1–2 μm . PCF performance has been analyzed at wavelength of 1.55 μm for single-polarization window. The effect of PCF parameters on dispersion and confinement loss are computed and presented for different air filling fractions and ring diameters.

The refractive index of proposed PCF for different air fill fractions has been shown in Fig. 3. It has been observed from Fig. 3 the outline of the refractive index will change at change in air fill fraction. The figure comprises refractive index versus wavelength curve of PCF, with four different air holes diameter and constant pitch magnitude and it has been observed that refractive index decreases with enhance in wavelength.

- Case 1 In this case following air fill fraction has been used with constant pitch value of $\Lambda = 2.3 \mu\text{m}$: $d_1/\Lambda = 0.61$, $d_2/\Lambda = 1.3$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$. Its impact has been analyzed on dispersion and confinement loss. Figures 4 and 5 represent the dispersion and confinement losses for wavelength from 1 to 2 μm . At 1.55 μm , dispersion is approximately $-900 \text{ ps}/(\text{km nm})$ and confinement loss is approximately $-1.35 \times 10^{-7} \text{ dB}/\text{km}$.
- Case 2 The air fill fraction of the third and fourth layer is same as in Case 1 with $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$. The air fill fraction of first and second layer is

Fig. 3 Graph of refractive index versus wavelength at change in diameter of air holes of proposed PCF

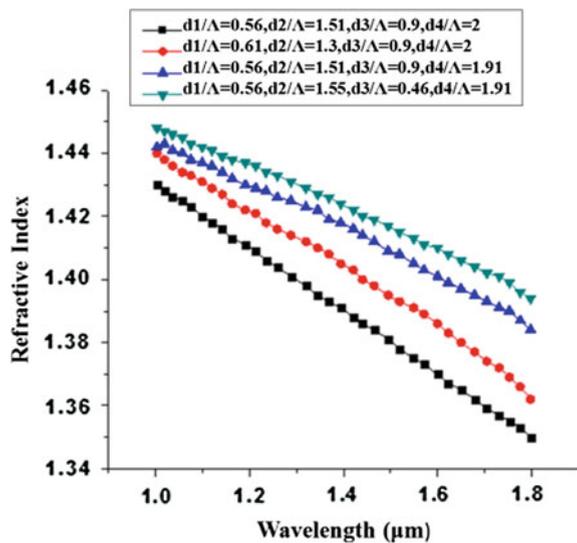


Fig. 4 Chromatic dispersion of SPSM PCF with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.61$, $d_2/\Lambda = 1.3$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$

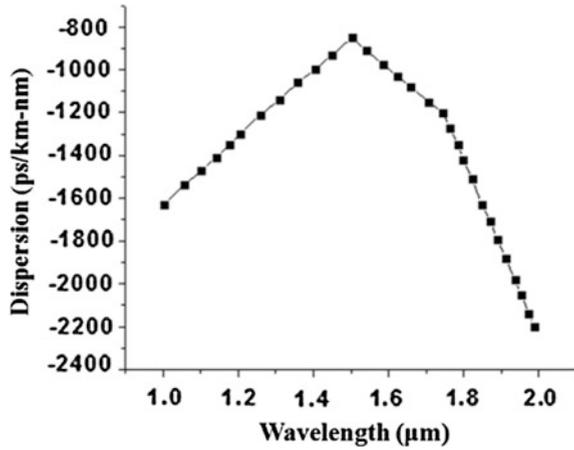
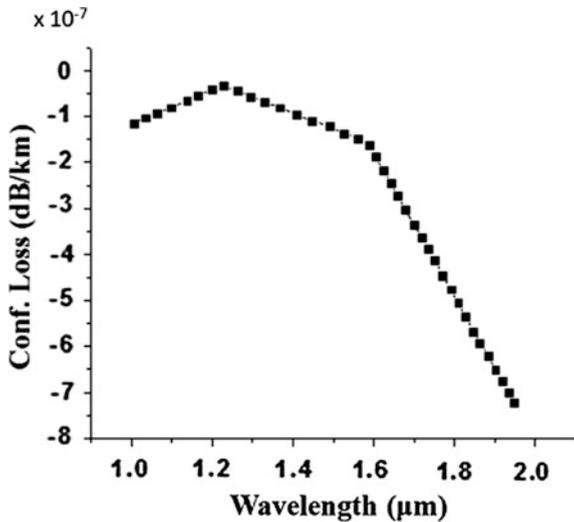


Fig. 5 Confinement loss of SPSM PCF with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.61$, $d_2/\Lambda = 1.3$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$



$d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.51$ to minimize the dispersion and confinement loss. Figure 6 shows the negative chromatic dispersion variation with wavelength range 1–2 μm . The approximate value of dispersion at 1.55 μm is $-1100 \text{ ps}/(\text{km nm})$. Figure 7 shows the value of confinement loss is $-1.35 \times 10^{-7} \text{ dB/km}$ at 1.55 μm which is the same as we get in Case 1.

Case 3 The air fill fraction of all the layers are same except the fourth layer which is $d_4/\Lambda = 1.91$ (outermost layer). From Fig. 8, at 1.55 μm dispersion is $1000 \text{ ps}/(\text{km nm})$ and from Fig. 9 confinement loss is approximately $-1.35 \times 10^{-7} \text{ dB/km}$. The dispersion value is too high but the confinement loss is same as in Case 1 and Case 2.

Fig. 6 Chromatic dispersion with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.51$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$

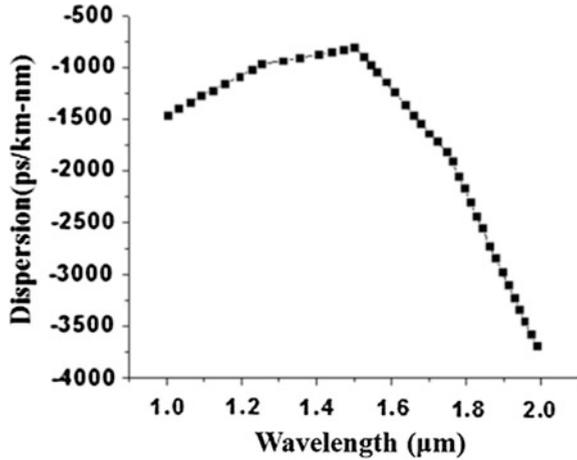
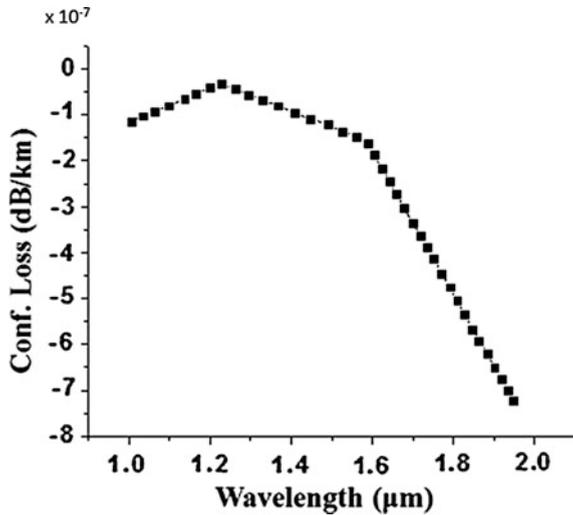


Fig. 7 Confinement loss with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.51$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 2$



Case 4 The air fill fraction values of the layers are as follows: $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.55$, $d_3/\Lambda = 0.46$, $d_4/\Lambda = 1.91$. In Figs. 10 and 11 at $1.55 \mu\text{m}$, we analyzed chromatic dispersion value of $-1000 \text{ ps}/(\text{km}\cdot\text{nm})$ and confinement loss is $1.35 \times 10^{-7} \text{ dB}/\text{km}$.

After analyzing the Case 1–4, the projected design in Case 2 has been used. The negative chromatic dispersion played a vital role in PCF application. It has been used in dispersion compensation, optical communication systems, and linear/nonlinear

Fig. 8 Chromatic dispersion with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.51$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 1.91$

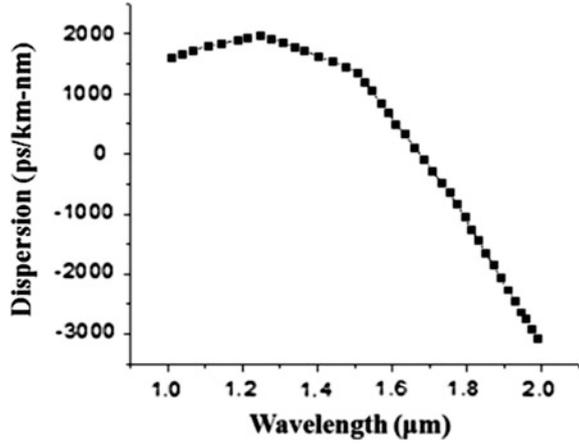


Fig. 9 Confinement loss with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.51$, $d_3/\Lambda = 0.9$, $d_4/\Lambda = 1.91$

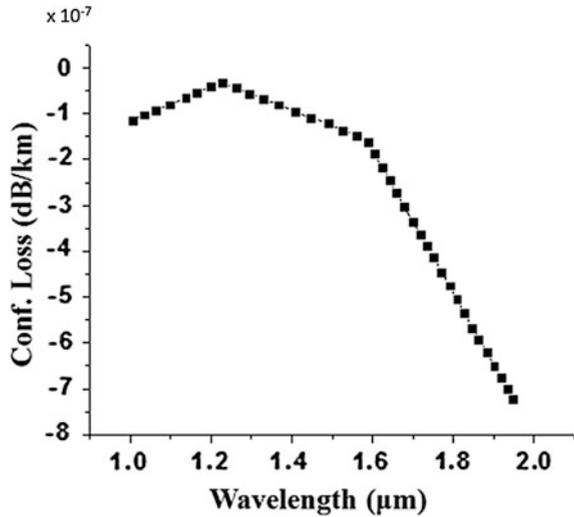


Fig. 10 Chromatic dispersion with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.55$, $d_3/\Lambda = 0.46$, $d_4/\Lambda = 1.91$

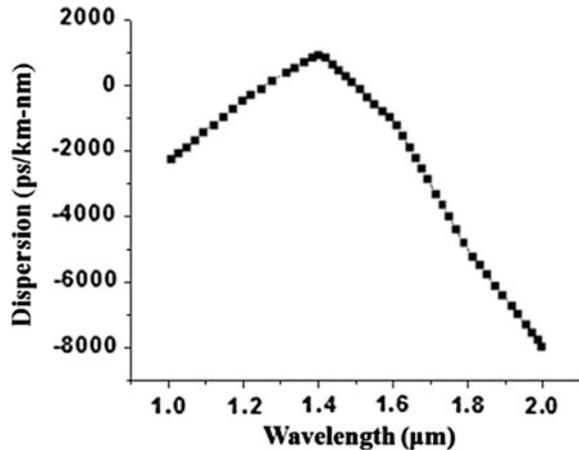
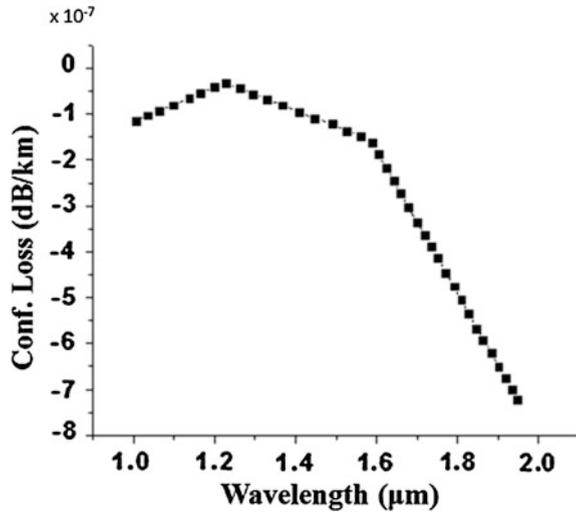


Fig. 11 Confinement loss with wavelength for $\Lambda = 2.3 \mu\text{m}$, $d_1/\Lambda = 0.56$, $d_2/\Lambda = 1.55$, $d_3/\Lambda = 0.46$, $d_4/\Lambda = 1.91$



optics. Low or negative dispersion and confinement loss in PCFs can be obtained for a broad range of wavelengths.

5 Conclusion

In this work, a noble design of PCF having negative chromatic dispersion and low confinement loss in all optical bands are presented. It has been found that the SPSM PCF can be optimized for 1.55 μm wavelength by selecting proper design parameters. Proposed design offers negative dispersion over a wide range of wavelengths. It has been used as dispersion compensating fibers due to highly negative dispersion values are achieved. It has been also shown that the confinement loss sharply falls down to a very low magnitude. Looking toward the near future, fabrication of proposed PCF is believed to be enabled with better feasibility and efficiency.

References

1. Messerly, M.J., Onstott, J.R., Mikkelsen, R.C.: A broad-band single polarization optical fiber. *J. Light w. Technol.* **9**, 817–820 (1991)
2. Silva, J.P., Bezerra, D.S., Fonseca, I.E.: Photonic crystal fiber design with Ge-doped core for residual chromatic dispersion compensation. In: *Proceedings of the Microwave and Optoelectronic Conference, Belem Para, Brazil, 3–6 Nov 2009*

3. Samiul, H.M., Ahmad, R., Selim, H.M., Hasan, M.: Residual dispersion compensation over the S + C+L + U wavelength bands using highly birefringent octagonal photonic crystal fiber. *Appl. Opt.* **53**, 3057–3062 (2014)
4. Islam, M.A., Alam, M.S.: Design optimization of equiangular spiral photonic crystal fiber for large negative flat dispersion and high birefringence. *J. Light w. Technol.* **30**, 3545–3551 (2012)
5. Areed, N.: Ultra-flattened dispersion honeycomb lattice photonic crystal fiber. In: *Proceeding of 5th International conference of Sciences of Electronic, Technologies of Information and Telecommunications*, Tunisia, 26–29 March 2009
6. Saitoh, K., Koshiba, M., Hasegawa, T., Sasaoka, E.: Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion. *Opt. Express* **11**, 843–852 (2003)
7. Poli, F., Cucinotta, A., Fucchi, M., Selleri, S., Vincetti, L.: Characterization of microstructured optical fibers for wideband dispersion compensation. *J. Opt. Soc. Am. A* **20**, 1958–1962 (2003)
8. Begum, F., Namihira, Y., Kinjo, T., Kaijage, S.: Supercontinuum generation in square photonic crystal fiber with nearly zero ultra-flattened chromatic dispersion and fabrication tolerance analysis. *Opt. Commun.* **284**, 965–970 (2011)
9. Begum, F., Namihira, Y., Razzak, S.M.A., Kaijage, S.F., Hai, N.H., Miyagi, K., Higa, H., Zou, N.: Flattened chromatic dispersion in square photonic crystal fibers with low confinement losses. *Opt. Rev.* **16**, 54–58 (2007)
10. Andres, P., Ferrando, A., Silvestre, E., Miret, J.J., Andres, M.V.: Dispersion and polarization properties in photonic crystal fibers. In: *Proceedings of the 4th International Conference on Transparent Optical Networks*, 21–25 Apr 2002
11. Birks, T.A., Knight, J.C., Russell, P.S.J.: Endlessly single-mode photonic crystal fiber. *Opt. Lett.* **22**, 961–963 (1997)
12. Basset, I.M., Argyros, A.: Elimination of polarization degeneracy in round waveguides. *Opt. Express* **10**, 1342–1346 (2002)
13. Argyros, A., Issa, N., Basset, I., van Eijkelenborg, M.: Microstructured optical fiber for single-polarization air guidance. *Opt. Lett.* **29**, 20–22 (2004)
14. Nejad, S.M., Ahramezani, M., Pourmahyabadi, M.: Design and simulation of a dual core photonic crystal fiber for dispersion compensation over E to L wavelength band. In: *International Symposium on Telecommunications*, Kyoto, 27–28 Aug 2008
15. Nejad S.M. Ehteshami, N.: Novel design to compensate dispersion for index guiding photonic crystal fiber with defected core. In: *Proceeding on 2nd International Conference on Mechanical and Electronics Engineering*, Kyoto, 1–3 Aug 2010

Triple Gate SOI MOSFET

Amit Agarwal, P.C. Pradhan and Bibhu Prasad Swain

Abstract Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) had been one of the best devices designed for integrated circuits over the decades. Due to continuous downscaling of the transistor the short-channel effects comes into play and further scaling becomes difficult. So, Multi-gate devices have been proposed so as to reduce these effects. Analytical modeling of Tri-gate MOSFET by solving Poisson's equation and necessary boundary condition is proposed in this paper. Surface potential for Tri-gate SOI MOSFET has been obtained and the effects of the device parameters like oxide thickness, different oxide material, channel length, gate voltage and drain voltage are plotted using MATLAB simulator.

Keywords Silicon on insulator • Short-channel effects • Tri-gate MOSFET

1 Introduction

MOSFET have delivered high performance for the reduction of cost in chips used for communication system and high-end processors [1, 2]. Scaling down [3, 4] of the MOSFET is reaching its physical limits due to short-channel effects (SCEs) [5] such as hot carrier effects [6, 7], gate oxide tunneling, subthreshold swing [8, 9], gate-induced drain lowering, drain-induced barrier lowering [10], voltage threshold

A. Agarwal (✉) · P.C. Pradhan
Department of Electronic and Communication Engineering,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: amiteng2007@gmail.com

P.C. Pradhan
e-mail: prashant.c.pradhan@gmail.com

B.P. Swain
Centre for Material Science and Nano Technology,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: bibhuprasad.swain@gmail.com

shift [11]. Different structures previously have been implemented like double gate MOSFET so as to reduce short-channel effects. But this device is not able to reduce these effects. In triple gate MOSFET [12, 13] there are three gates and this channel is controlled by three sides so as to better control the electrostatic potential in the channel and reduce the short-channel effects.

2 Triple Gate SOI MOSFET

Figure 1 shows the schematic of Tri-gate MOSFET [14, 15] and from the structure we can see that there are three gates along the sides of the channels. Figures 2 and 3 show the front gate and two side gate of the transistor. Due to presence of more than one gate there is better electrostatic potential control over the channel and performs better transportation of the electron through the channel [16, 17]. Due to this better control the short-channel effects such as gate oxide tunneling, hot carrier effects, drain induced barrier lowering, gate induced drain lowering, voltage threshold shift, subthreshold swing can be minimized.

2.1 Analytical Modeling

In semiconductor body the potential distribution [18, 19] $\Psi(x, y)$ is given by Poisson's equation as

$$\frac{d^2\Psi(x, y)}{dx^2} + \frac{d^2\Psi(x, y)}{dy^2} = \frac{qN_A}{\epsilon_{si}} \quad 0 \leq x \leq L \text{ and } 0 \leq y \leq t_{si}, \quad (1)$$

Fig. 1 Schematic of Tri-gate SOI MOSFET [12]

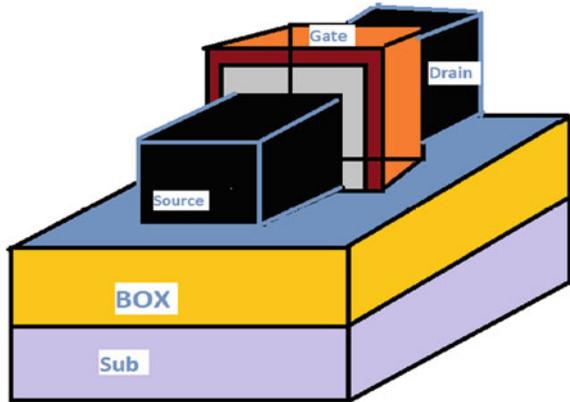


Fig. 2 Schematic of front gate of Tri-gate MOSFET [12]

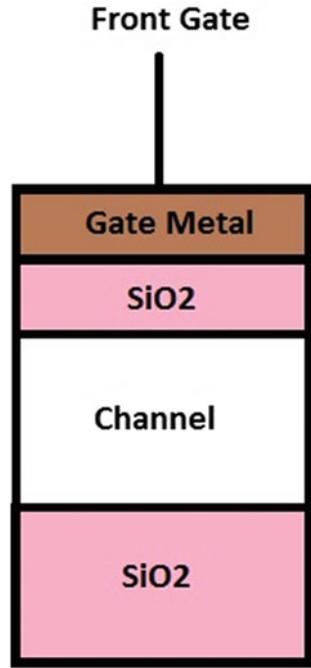
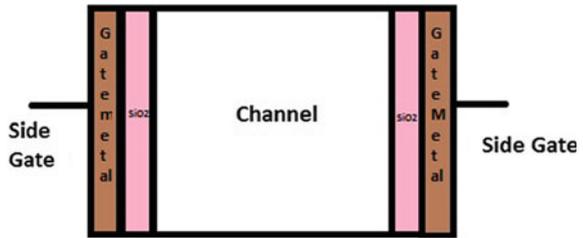


Fig. 3 Schematic of side gates of Tri-gate of MOSFET [12]



where $N_A, \epsilon_{si}, t_{si}$ denotes the acceptor doping concentration, permittivity of silicon and silicon film thickness. L is device channel length.

For Triple gate SOI MOSFET structure [20] the electrostatic potential Ψ can be obtained by sum of $\Psi_a(x, y)$ and $\Psi_s(y, z)$, where Ψ_a and Ψ_s are the potential distribution of the asymmetric and the symmetric DG MOSFET respectively and considering the boundary conditions [21] we obtain:

$$\Psi(x, 0) = \Psi_f(x), \Psi(x, t_{si}) = \Psi_b(x) \tag{2}$$

$$\frac{d\Psi(x, t_{si})}{dy} = \frac{\epsilon_{ox}}{\epsilon_{si}} \frac{\Psi_{Sub} - \Psi_s(x)}{t_b} \tag{3}$$

$$\Psi_{a,s}(x, y) = \frac{1}{\sinh \frac{L}{\lambda_{a,s}}} \left[(V_{bi} + V_d - A_{a,s}) \sinh \frac{x}{\lambda_{a,s}} (V_{bi} - A_{a,s}) \sinh \frac{L-x}{\lambda_{a,s}} + A_{a,s} \sinh \frac{L}{\lambda_{a,s}} \right], \quad (4)$$

where

$$A_a = - \frac{2C_{si} (c_{ox} t_b \Phi_{gs} + \epsilon_{ox} \Phi_{Sub} - qN_A t_{si} t_b) + 2\epsilon_{ox}^2 \Phi_{gs} t_{ox} - qN_A t_{si}}{2c_{ox} (c_{si} (t_f + t_b) + \epsilon_{ox})} \quad (5)$$

$$\lambda_a = \sqrt{\frac{2\epsilon_{si}^2 t_b t_{si} + \epsilon_{si} t_{si}^2}{2c_{ox} \epsilon_{si} (t_f + t_b) + \epsilon_{ox}}} \quad (6)$$

$$A_s = \Phi_{gs} - qN_A \frac{\epsilon_{si} t_{ox} W + \epsilon_{ox} (W - z) z}{2\epsilon_{ox} \epsilon_{si}} \quad (7)$$

$$\lambda_s = \sqrt{\frac{\epsilon_{si} t_{ox} W}{2\epsilon_{ox}}} - \frac{1}{2} W_z - \frac{1}{2} z^2 \quad (8)$$

With the factor $R = \frac{W}{W + t_{si}}$ is used to obtain the sum of two potential distributions

$$\Psi(x, y, z) = m\Psi_a(x, y) + \frac{\Psi_s(x, z)}{m}. \quad (9)$$

3 Results and Discussion

Figures 4, 5, 6, 7, 8, 9, 10, 11 and 12 show the surface potential along the channel for Tri-gate SOI MOSFETs. For Tri-gate SOI MOSFET device we set $N_A = 10^{16} \text{ cm}^{-3}$ and $N_D = 10^{20} \text{ cm}^{-3}$ and dimensions of $t_f = 1 \text{ nm}$, $t_b = 100 \text{ nm}$, $t_{si} = 5 \text{ nm}$, $W = 5 \text{ nm}$ and $L = 10 \text{ nm}$.

Figure 4 shows that surface potential varying with channel length of a triple gate MOSFET. The surface potential curve, as shown in, is parabolic in nature. With the increase of channel length from 1 nm the surface potential decreases and with the increase of channel length above 5 nm the surface potential again starts increasing. Also as the thickness of the back gate is varied from 6 to 10 nm the surface potential starts increasing from around 0.1 to 0.45 V and deviation is about 75% of the lower value of surface potential. From this figure we can also analyse that surface potential when back gate thickness is 6 nm, the curve flattens than that of 10 nm. Figure 5 shows that surface potential varied with channel length from 1 to 10 nm of Tri-gate MOSFET with different widths of the channel. With the increase of channel length from 1 to 10 nm, the surface potential starts decreasing and when the channel length is above 5 nm the surface potential increases which depicts a

Fig. 4 Surface potential versus channel length with different back gate thickness in triple gate MOSFET

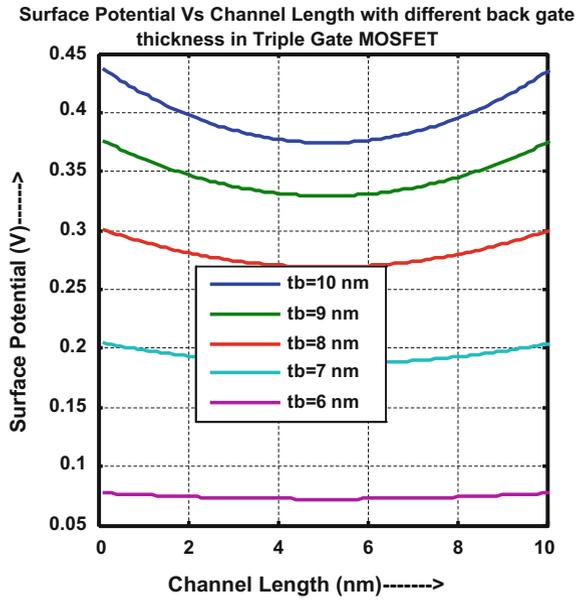


Fig. 5 Surface potential versus channel length with different widths in triple gate MOSFET

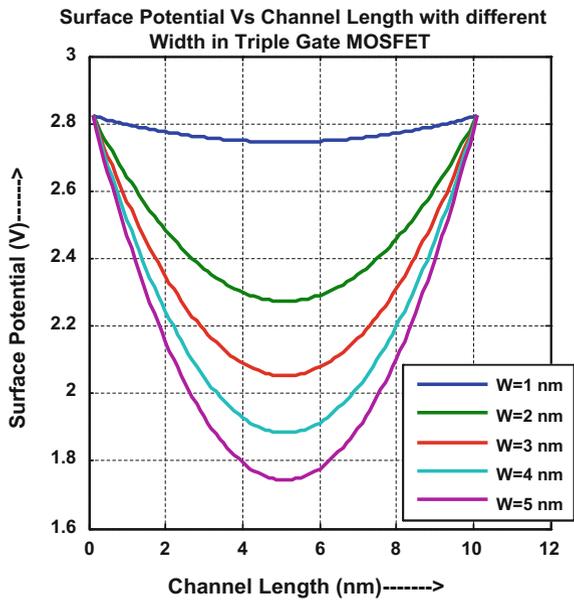


Fig. 6 Surface potential versus channel length with different gates to source voltage in triple gate MOSFET

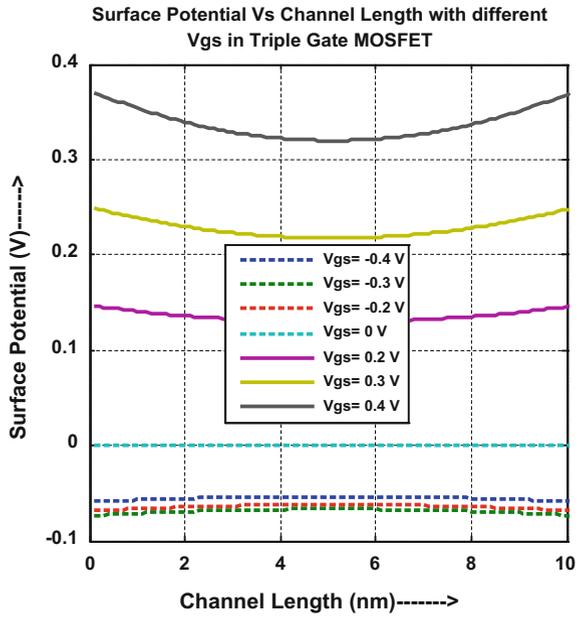


Fig. 7 Surface potential versus channel length with different drains to source voltage in triple gate MOSFET

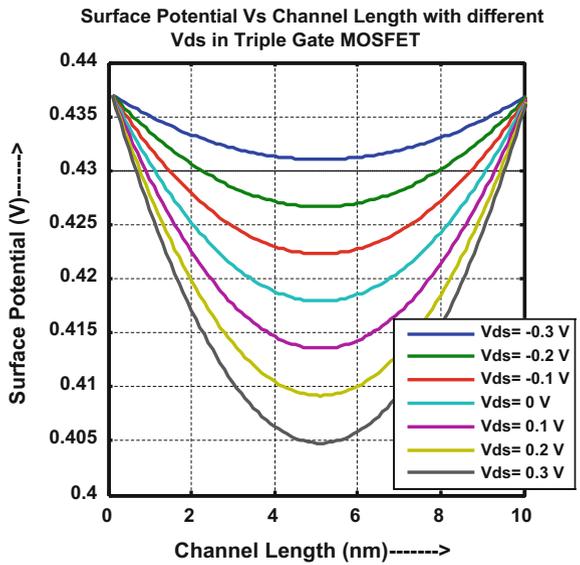


Fig. 8 Surface potential versus channel length with different thickness of t_{si} in triple gate MOSFET

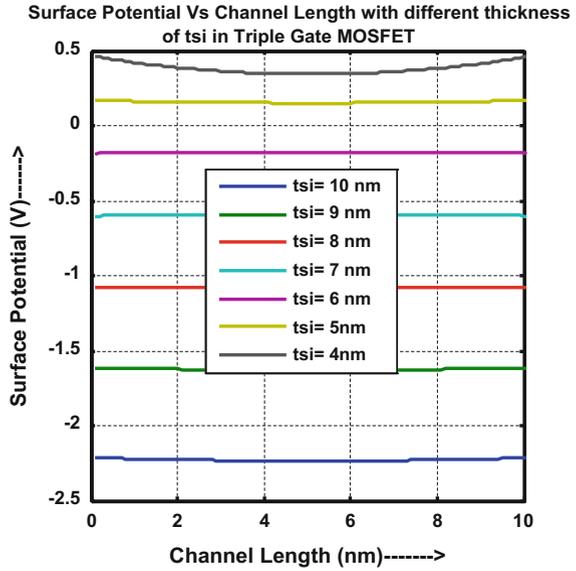


Fig. 9 Surface potential versus channel length with different gate oxide thicknesses in triple gate MOSFET

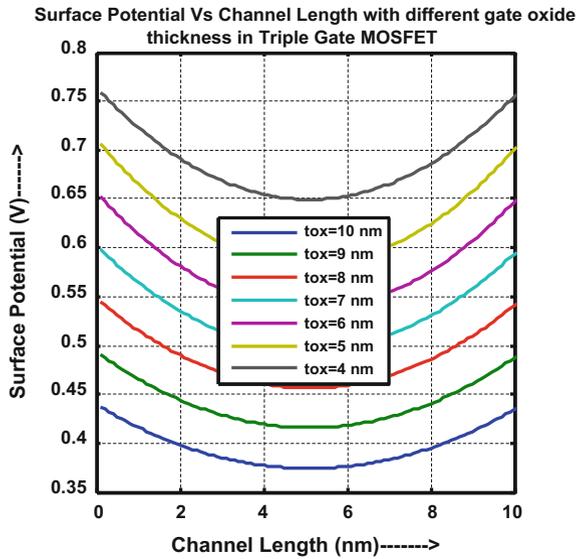


Fig. 10 Surface potential versus channel length with different front gate thickness in triple gate MOSFET

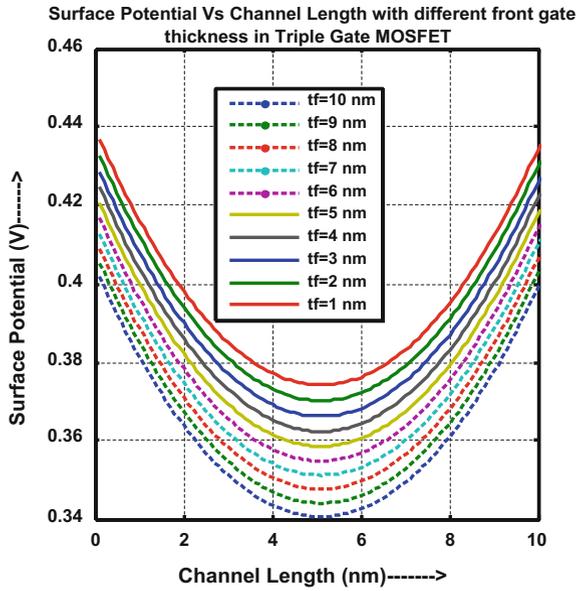


Fig. 11 Surface potential versus channel length with different gate oxide material in triple gate MOSFET

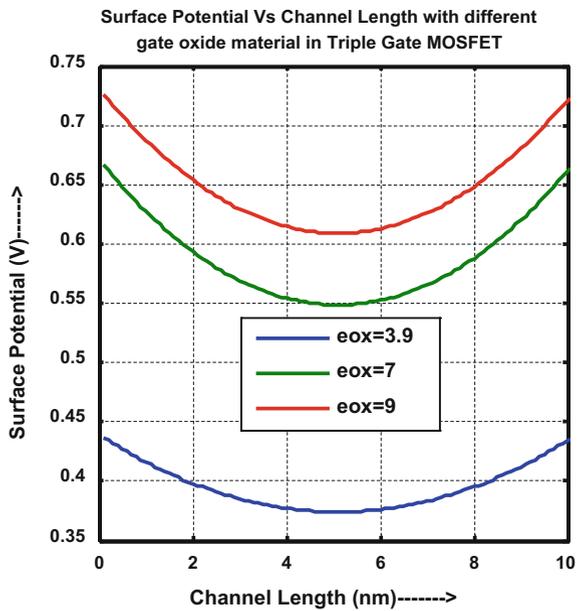
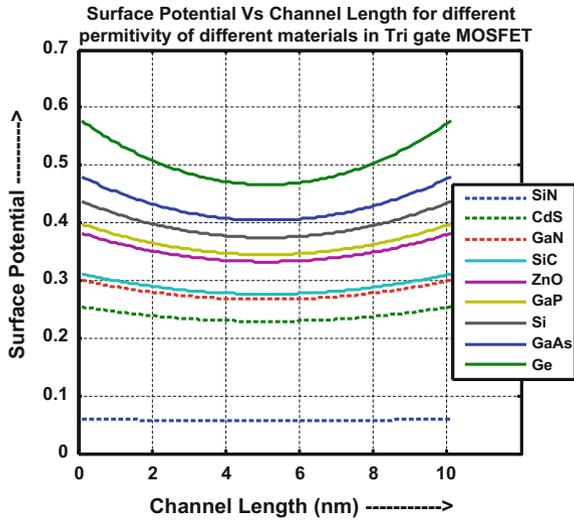


Fig. 12 Surface potential versus channel length with different permittivities of different materials in Tri-Gate MOSFET



parabolic curve. In this figure, we observe that, for a width of 5 nm along the channel length of 1–10 nm has less surface potential and more steeper as compared to channel width of 1 nm. This change is about 35% of the lower value of surface potential.

Figure 6 shows that surface potential varying with channel length of a triple gate MOSFET. In this case gate to source voltage is varied from -0.4 to $+0.4$ V. The increasing of +ve gate to source voltage indicates parabolic nature where the surface potential initially decreased and increased after the channel length of 5 nm. However, for the negative gate to source voltage the surface potential is negative and minimum with an independent of applied voltage. Figure 7 shows that surface potential varied with channel length varies 1 to 10 nm of a Tri-gate MOSFET with different drain to source voltages. In the present modeling the drain to source voltage varied from -0.3 V to $+0.3$ V. The potential distribution curve of a Tri-gate MOSFET shows parabolic nature and shows minimum value for 5 nm channel length. The minimum variation of surface potential, i.e. 1.5 is for -0.3 V while maximum potential changes, i.e. about 8% for $+0.3$ V drain to source.

Figure 8 shows that surface potential varied with channel length varies 1–10 nm of a Tri-gate MOSFET with different thicknesses of the film. As we increase the thickness of the film from 4 to 10 nm, the surface potential starts decreasing from 0.5 to about -2.2 V. When the thickness of the film is 4 nm and the channel length is varied from 1 to 10 nm, the surface first starts decreasing from 0.5 V and then nearly about 5 nm channel length the surface potential starts increasing from about 0.2 V till 0.5 V. But when the thickness of the film is about 10 nm, the surface potential is almost flattened as compared to film of 4 nm thickness, along the channel length which is varied from 1 to 10 nm. Figure 9 shows that surface potential varying with channel length of triple gate MOSFET. From the figure, we

can see that surface potential starts decreasing from 0.75 to 0.65 V when the channel length is varied from 1 to 5 nm and again when the channel length is increased from 5 to 10 nm the surface potential starts increasing from 0.65 to 0.75 V. In this figure we also observe that with the increase of the thickness of the oxide from 4 to 10 nm the surface potential starts decreasing from 0.75 to 0.45 V, which deviates by 40% of lower value of surface potential.

Figure 10 shows that surface potential varying with channel length of a triple gate MOSFET. From the figure, we can analyse that with the increases of channel length from 1 to 5 nm the surface potential starts decreasing from 0.44 to 0.37 V and when the channel length is further increased from 5 to 10 nm the surface potential starts increasing from 0.37 to 0.44 V. This curve is parabolic in nature. Also when front gate thickness is varied from 1 to 10 nm, the surface potential starts decreasing from 0.44 to 0.40 V. Figure 11 shows that the surface potential varied with channel length varies 1–10 nm of a Tri-gate MOSFET with different oxide materials such as silicon dioxide, aluminium oxide. Figure shows a parabolic curve in nature, as the channel length is varied from 1 to 5 nm, the surface potential starts decreasing from 0.75 to about 0.61 V and again when the channel length is varied from 5 to 10 nm, the surface potential starts increasing from 0.61 to 0.75 V. In this paper, we have used a silicon dioxide material which has a permittivity of 3.9 and shows higher surface potential, i.e. 0.75 V as compared to aluminium oxide material which is having a higher permittivity, i.e. 9 with surface potential of 0.45 V. This shows a deviation of about 40% in surface potential.

Figure 12 shows that surface potential varied with channel length varies 1–10 nm of a Tri-gate MOSFET with different oxide materials such as SiN, CdS, GaN, SiC, ZnO, Si, GaAs, Ge. From the figure, we can see that when we use different films with different permittivities then we obtain different surface potentials. For SiN film which is having less permittivity has surface potential of 0.05 V which is lower than that of Ge film with 0.60 V surface potential. As we increase the permittivity of the material, i.e. SiN, CdS, GaN, SiC, ZnO, Si, GaAs and Ge, the surface potential starts increasing from 0.05 to 0.60 V. Also as we vary the channel length from 1 to 5 nm for Ge, the surface potential starts decreasing from 0.6 to 0.45 V and when the channel length is varied from 5 to 10 nm, the surface potential starts increasing from 0.45 to 0.60 V. But for SiN, when channel length is varied from 1 to 10 nm the surface potential curve almost flattens for the entire span with a value of 0.05 V.

4 Conclusion

In this paper we have done analytical modeling for Tri-gate SOI MOSFET and plotted the graphs between surface potential and channel length from 1 to 10 nm using MATLAB. We have investigated that the Tri-gate MOSFET below 10 nm is

showing better performance than previously existing MOSFET and reduces short-channel effects. We have also analysed that with the change in the different parameters of the MOSFET like drain to source voltage, thickness of the film, thickness of the oxide, gate to source voltage, permittivity of oxide the surface potential also changes. Due to these changes in surface potential, short-channel effects can be reduced by setting the proper physical parameter of the MOSFET.

References

1. Moore, G.E.: Cramming more components onto the integrated circuits. (Reprinted from Electronics, pp. 114–117, April 19, 1965). Proc. IEEE **86**, 82–85 (1998)
2. Semiconductor Industry Association. International technology roadmap for semiconductors (2000 update)
3. Bohr, M.: A 30 year retrospective on Dennard's MOSFET scaling paper. IEEE Solid-State Circ. Soci. **9**, 11–15 (2007)
4. Taur, Y., Buchanan, D.A., Chen, W., Frank, D.J., Ismail, K.E., Lo, S.H., SaiHalasz, G.A., Viswanathan, R.G., Wann, H.J.C., Wind, S.J., Wong, H.S.: CMOS scaling into the nanometer regime. Proc. IEEE **85**, 486–504 (1997)
5. Young, K.K.: Short-channel effects in fully depleted SOI MOSFET's. IEEE Trans. Electron Devices **36**, 399–402 (1989)
6. Fischetti, M.V., Laux, S.E., Crabbe, E.: Understanding hot-electron transport in silicon devices—is there a shortcut. J. Appl. Phys. **78**, 1058–1087 (1995)
7. Ning, T.H., Osburn, C.M., Yu, H.N.: Emission probability of hot electrons from silicon into silicon dioxide. J. Appl. Phys. **48**, 286–290 (1997)
8. Zhang, Q., Zhao, W.: Low subthreshold swing tunnel transistors. IEEE Trans. Device Lett. **27**, 297–300 (2006)
9. Banna, S.R., Chan, P.C.H., Ko, P.K., Nguyen, C.T., Chan, M.: Threshold voltage model for deep-submicrometer fully depleted SOI MOSFET's. IEEE Trans. Electron. Devices **42**, 1949–1955 (1995)
10. Garduño, S.I., Cerdeira, A., Estrada, M., Alvarado, J., Kilchystka, C., Flandre, D.: Contribution of the carrier tunnelling and gate induced drain leakage effects to the gate and drain currents of fin-shaped field effect transistor. J. Appl. Phys. **109**, 1–7 (2011)
11. Liu, Z.H.: Threshold voltage model for deep submicrometer MOSFETs. IEEE Trans. Electron. Devices **40**, 86–90 (1993)
12. Colinge, J.: The new generation of SOI MOSFETs. Rom. J. Inf. Sci. Techno. **111**, 3–15 (2008)
13. Park, J.T., Colinge, J.P.: Multiple-gate SOI MOSFETs: device design guidelines. IEEE Trans. Electron. Devices **49**, 2222–2229 (2002)
14. LuD, D., Dunga, M.V., Lin, C.H., Niknejad, A.M., HuC.: A multi-gate MOSFET compact model featuring independent-gate operation. IEDM (2007)
15. Chaudhry, A., Kumar, M.J.: Controlling short-channel effects in deep submicron SOI MOSFET's for improved reliability: a review. IEEE Trans. Device Mater. Rel. **4**, 99–109 (2004)
16. Lemme, M., Mollenhauer, T., Henschel, W., Wahlbrink, T., Gottlob, H., Favi, J., Baus, M., Winkler, O., Spangenberg, B., Kurz, H.: Subthreshold behavior of triple-gate MOSFETs on SOI Material. Solid State Electron. Lett. **48**, 529–534 (2003)
17. Doyle, B., Boyanov, B.: Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout. In: Digest of Technical Papers, Symposium on VLSI Technology, USA (2003)

18. Suzuki, K., Pidin, S.: Short-channel single-gate SOI MOSFET model. *IEEE Trans. Electron. Devices* **50**, 1297–1305 (2003)
19. Mohammadi, H., Abdullah, H., Fu, Dee C.: A review on modeling the channel potential in multi-gate MOSFETs. *Sains Malaysiana Publ.* **43**, 861–866 (2014)
20. Frank, D.J., Taur, Y.: Generalized scale length for two-dimensional effects in MOSFETs. *Electron. Device Lett. IEEE* **19**, 385–387 (1998)
21. Auth, C.P., Plummer, J.D.: A simple model for threshold voltage of surrounding-gate MOSFET's. *IEEE Trans. Electron. Devices* **45**, 2381–2383 (1998)

Wearable Antennas for Medical Application: A Review

Shreema Manna, Tanushree Bose and Rabindranath Bera

Abstract Wearable devices for monitoring different biological signals are now gaining huge interest now a day. Use of textile material as substrate material for antenna design is growing rapidly. In order to reduce the effect of human body the performance of antenna several different antenna structures have been proposed. This article reveals research on different designs and development of wearable antennas for biomedical application at different frequency bands.

Keywords Wearable antenna · ISM band · Microstrip antenna
Textile antenna · Radiation pattern

1 Introduction

Wearable electronic devices for monitoring and telemetering of biological signals now a day have become an integral part of everyday human life. Technology for developing wearable antenna is emerging day by day. One key requirement for wearable devices is small size. Also reliability and flexibility to operate in all environments are other major issues of wearable devices. These devices must be inexpensive, robust and low maintenance required. Antenna is the main constituent of wearable devices. Many researches are going on the development of wearable antenna of miniaturized size, which is having less weight and at the same time conformal to devices and body shape. In this review, primary focus is on how wearable antennas are developed for monitoring biological signals. The performance

S. Manna (✉) · T. Bose · R. Bera
Department of Electronics and Communication Engineering, Sikkim Manipal
Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India
e-mail: shreema123@gmail.com

T. Bose
e-mail: contact_tanushree@gmail.com

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

characteristics, advantages, disadvantages of those antennas are discussed. The transmitter of wearable device must be compact and have minimum power consumption. Also signal transmission must be optimized on human figure. Design of an antenna which can satisfy all these measures is very much difficult. These antennas must produce an omnidirectional radiation shape in order to provide transmission in all direction regardless of the location of the receiver and orientation of the antenna.

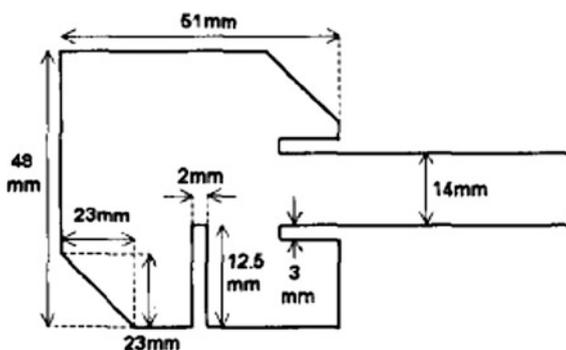
2 Review of Wearable Antenna

In 2003 author of [1] have presented smart cloth made of fabric material. The substrate of the antenna is made of low cost fleece fabric. The radiating element and ground plane is made up of knitted copper. The length and width of the antenna is 56 and 51 mm respectively. Dielectric constant of the material at 2.45 GHz is 1.04. The ground plane dimensions are 76 mm \times 71 mm. Substrate thickness is 3 mm. The antenna is flexible to operate in different frequency bands.

In 2004 authors of [2] have presented a circularly polarized textile antenna for wearable applications. The proposed antenna Fig. 1 uses single inset microstrip feed line. Circular polarization is achieved through truncated corners and single slit in patch. Polyimide spacer of 6 mm thickness is used as substrate material. Its permittivity is 1.15 at 2.4 GHz. Nickel plated woven textile is used for patch and the ground plane. Nickel has high resistance in contradiction of oxidation as well as corrosion. The antenna structure including the patch and the ground plane were mounted to the spacer fabric by means of textile based on ammonia. The SMA jack was electrically connected to two-component glue which is conductive.

In 2005, authors of [3] presented a study on the design, simulation and evaluation of wearable antennas. The antenna operates with a centre frequency of 868 MHz. Substrate material used here is glass epoxy having relative permittivity 4.8, thickness 1.6 mm. The conducting element used is cooper. The ground plane limits the backward wave towards the wearer. In 2006, authors of [4] have

Fig. 1 Textile antenna having circular polarization



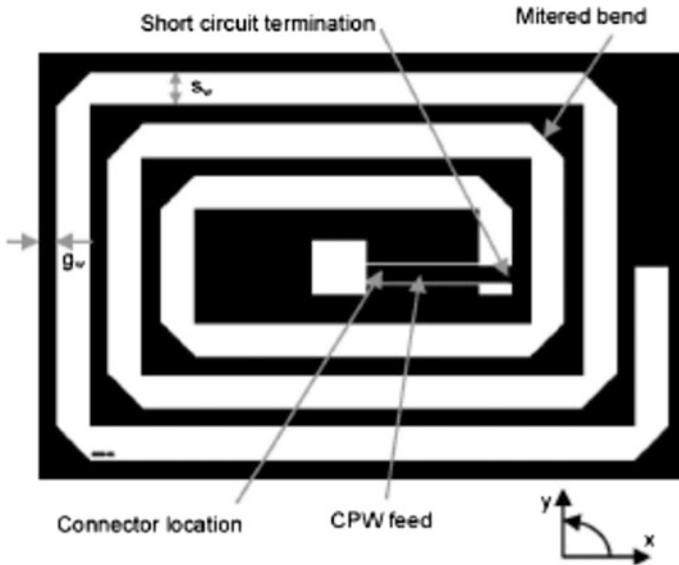


Fig. 2 Meander line slot with short circuit

presented small low cost printed antenna operating at frequency less than 1 GHz. The proposed antenna is tested to operate at 430 MHz. It is co planner waveguide fed to provide uni-planner solution. Proposed antenna, shown in Fig. 2 provides near omnidirectional pattern with 5% band width. The proposed structure is shown in Fig. 2. The overall size of the antenna is 4.5" × 3". The total length of the slot given is 29", width is 6 mm and gap between adjacent slots is 3 mm. Material used for antenna is 1.6 mm thick FR4.

In the year 2006 [5] author Sanz-Izquierdo et al. presented wearable button antenna (Figs. 3 and 4) for WLAN application. The antenna operates at 2.4 and 5 GHz with omni-directional radiation pattern.



Fig. 3 Textile layer small size button antenna with (left)

Fig. 4 Microstrip feed small size dual band button antenna

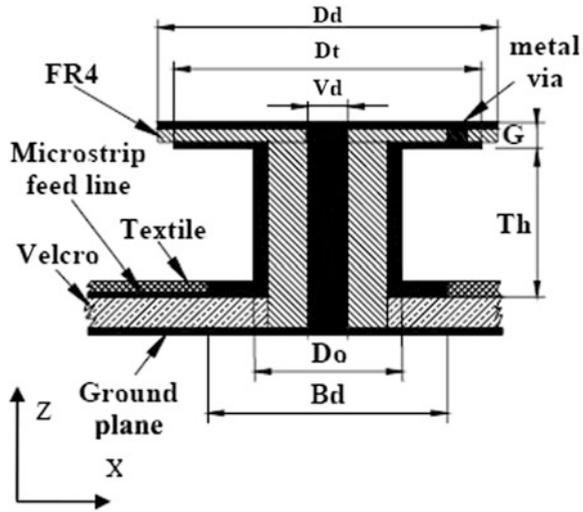


Table 1 Dimensions of wearable button antenna

Parameters of the proposed antenna	Values
Disc diameter, D_d	17
Top disc diameter, D_t	15.4
Base diameter, B_d	11
Cylinder outer diameter, D_o	7
Centered via diameter, V_d	1.6
Track button height, T_h	5.6
Gap between disc and button, G	1.5

Table 1 depicts the different dimensions of proposed small size wearable button antenna.

In the year 2007 [6] author Kim et al. presented embedded UHF RFID tag antenna. The antenna has been designed using flexible thread or textile material. T matching method is used for the impedance matching between tag antenna and RFID chip. The size of the proposed antenna is about 120×9 mm.

In 2009 author of [7] presented flexible wearable antennas operating between 100 MHz and 1 GHz. Three antenna designs are considered here spiral (Fig. 5), bowtie (Fig. 6), and broadband wire dipole. The broadband wire dipole can cover the frequencies, from 100 to 250 MHz. This element consists of a wire dipole with lumped loads placed in each arm to produce a broadband antenna. The spiral and bowtie both are capable of covering the band from 250 to 800 MHz. Scaled versions of the spiral and bowtie can cover higher bands as required. These antennas were designed using finite difference time domain (FDTD) tool.

Fig. 5 Conducting thread spiral

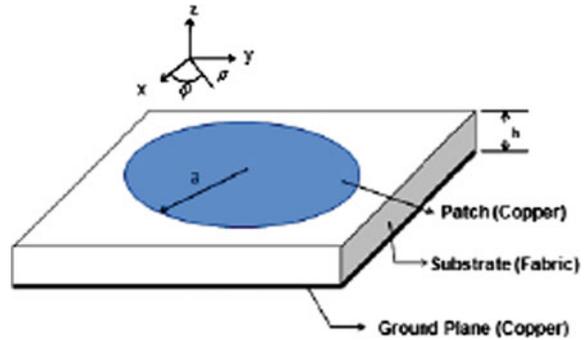


Fig. 6 Copper coated fabric bow tie antenna



In authors of [8] have presented coplanar waveguide fed ultra-wideband antenna. In the proposed structure ultra-thin liquid crystal polymer of thickness 0.05 mm is used as substrate (relative permittivity 2.9). The dimension of the proposed antenna is $6 \times 16 \times 0.068 \text{ mm}^3$ and it is one of the lightest UWB antennas. The proposed

Fig. 7 Circular patch antenna



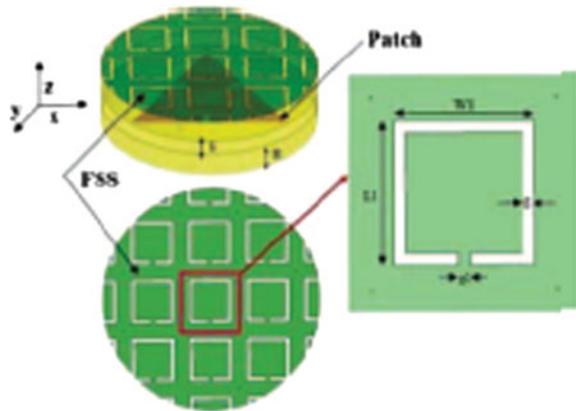
antenna was tested in five different bending angles 20° , 30° , 40° , 45° , 60° . Simulation result shows that -10 dB bandwidth spans from less than 3.1 to 10.6 GHz and hence covers whole UWB frequency band for any bending positions.

In [9] authors have proposed smart cloths. Antenna is designed using textile material. The proposed antenna is a circular patch of radius 26.3 mm (shown in Fig. 7) with a rectangular ground plane. Substrate material used here is indigo jeans. Permittivity is calculated by experimental method and the value is 1.67 and loss tangent 0.03. Thickness of the substrate used here is 2.84 mm. The proposed antenna resonates at frequency of 2.45 GHz.

Lee et al. of [10] presented electrically coupled LC (ELC) resonator resonating at 2.45 GHz. The designed antenna used two layered substrate (FR4, with relative permittivity 4.4). These substrates have thickness of 1.6 and 0.8 mm respectively. The area of the patch is 27.8 mm^2 . The resonant frequency achieved here is 2.45 GHz. The designed antenna has been tested on a body model of size $200 \times 200 \times 50 \text{ mm}^3$. The body model is having electrical properties like: permittivity of 52.7, conductivity 1.95 S/m and loss tangent 0.27. The SAR value of the designed antenna when placed on flat body model is obtained as 0.005 W/kg. Joshi et al. of [11] presented polyester substrate based microstrip patch antenna using inter digital capacitor and rectangular stub for wireless body area network applications. The antenna is fabricated using copper tape and polyester substrate. The designed antenna resonates at 2.45 GHz with bandwidth 40 MHz and gain of 5 dBi. The dimension of the rectangular patch is $30 \text{ mm} \times 4 \text{ mm}$, substrate thickness 3.14 mm, permittivity of 1.39 and loss tangent 0.01. The proposed antenna has been simulated using IE3D.

Sabban of [12] in his paper presented compact dual polarized microstrip antenna with high gain (0–2 dBi) at 434 MHz. The proposed antenna has dimension of $5 \times 5 \times 0.05 \text{ cm}^3$. Authors of [13] have presented CPW fed rectangular monopole patch antenna for wearable applications operating at 2.45 GHz (ISM band). The ground plane and patch of the antenna is made up of graphene and carbon

Fig. 8 3D view of the button antenna



nanotubes. The substrate material used is poly-dimethylsiloxane having dielectric constant 2.65, loss tangent 0.02 (substrate thickness 2 mm). The simulation result is compared with a reference antenna made of copper. The comparison result shows 18% bandwidth improvement in the proposed design. Since graphene is one of the top conductive materials, the proposed antenna can be used to study the effect of improvement in the ground structure.

In 2014, [14] authors Lim et al. presented their paper Wearable Textile Substrate Patch Antenna. Size of antenna is $12'' \times 9''$ and antenna impedance is 50Ω . Here textile material having low dielectric constant is used for designing the antenna. The proposed antenna is grounded on microstrip technology, which is a type of electrical transmission line used to transmit microwave frequency signals. It consists of three layers, ground plane, substrate and patch. They are fabricated on printed circuit board. Here the patch is rectangular having thickness 0.1 mm. Dimension of ground plane is $120 \times 120 \text{ mm}^2$ and thickness is 0.5 mm. Copper is used for patch and ground plane. The performance of antenna with different textile substrate (e.g. wash cotton, curtain cotton, polyester and poly-cotton) has been examined. Simulation result shows that at resonant frequency is 2.4 GHz, the return losses of these antennas are around -30 to -40 dB. Reducing the patch dimension increases resonant frequency which can be compensated by inserting sorting pin between top of the patch and ground plane.

Mondol et al. [15] in their paper presented wearable button antenna (Fig. 8) operating in WLAN frequency band (5.25–5.80 GHz). Here substrate material used is transparent acrylic fiber sheet and copper sheet is used as conducting material. A frequency selective surface is used, which is split ring shaped slot, to enhance the gain of the proposed antenna.

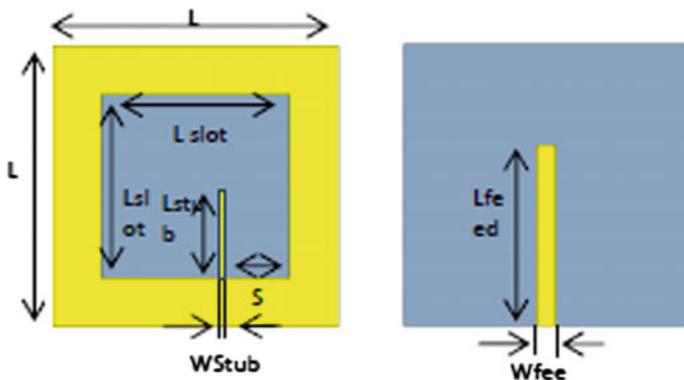


Fig. 9 The dual band square slot antenna with stub

Table 2 Proposed antenna parameters

Parameter	Value (mm)	Parameter	Value (mm)
L	50	L_{feed}	34
L_{slot}	31	W_{feed}	13
S	8	W_{stub}	0.6
L_{stub}	14		

The proposed antenna consists of triangular microstrip patch antenna.

In 2015, authors of [16] Hammoodi et al. have presented dual band square slot antenna, shown in Fig. 9. The antenna operates at ISM and WiMAX frequency band i.e. 2.4 and 3.5 GHz respectively. The gain of antenna is around 2.5 dB at both frequency bands. Vinyl substrate material issued here with relative permittivity of 2.5 and thickness angle of 0.47 mm. A square slot is removed from the ground plane and stub of length 14 mm and width 0.6 mm are introduced. Dimension of the antenna is 50 mm².

Performance of the antenna under bending condition has also been analyzed in both the bands. Gain achieved here is around 2.4–2.9 dB. This shows that the antenna is good in ISM and WiAX band. The antenna parameters are listed in Table 2.

In [17] authors have proposed two different miniaturized wearable microstrip antennas operating at 2.4–2.48 GHz. Using stacked triangular guiding patches and slotting of the patches author have achieved 60% size reduction with respect to the original volume. The proposed antennas have dimension of 39 × 39 × 2.1 mm³. Two substrate materials are used here Taconic CER-10 (dielectric constant 10.1) and FR4 (dielectric constant 4.4), which causes desired shift in the resonating

Fig. 10 3D view of the co-designed CP filtering antenna using SOLRs with a strip line feed (tilted view)

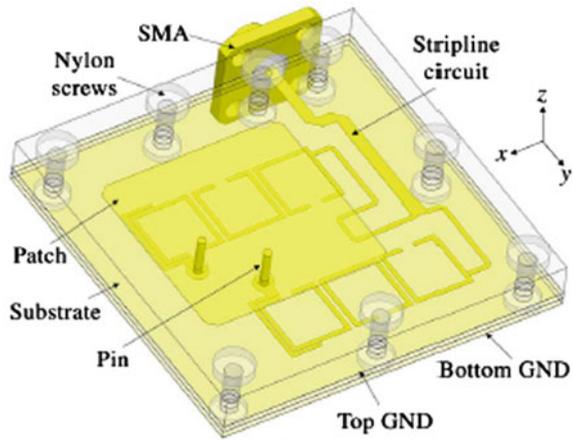
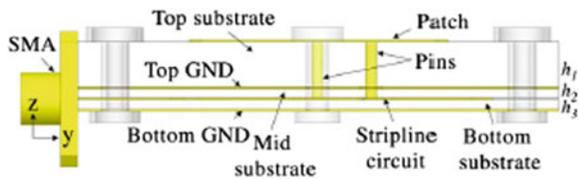
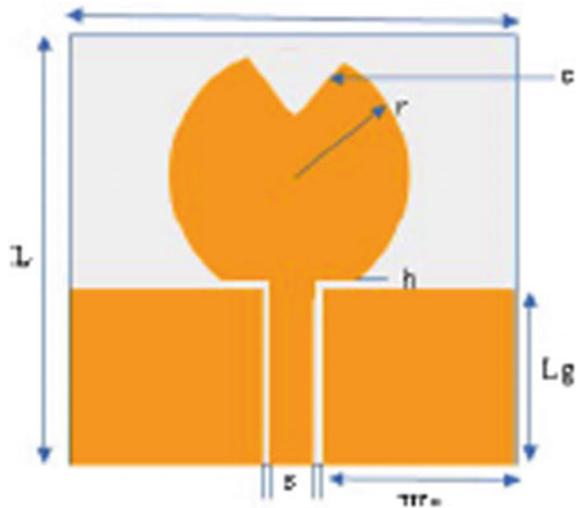


Fig. 11 Antenna structure (the side view)



frequency. Simulations are carried out in HFSS platform. Results show a return loss of -16.69 and -15.53 dB for the designed antennas respectively. Vatankhah et al. of [18] proposed in their report a compact antenna which is capacitively fed using inverted L shape feed for wearable wireless sensors. The antenna operates at 2.4–6.4 GHz band. The total size of the sensor and the elevated antenna is 35 mm \times 22 mm \times 9 mm. The proposed antenna is designed using FR4 substrate with full ground plane at the bottom and microstrip feed line at the top. A large via is created in the middle to connect the feed line with SMA connector for antenna parameter measurement. Jiang of [19] presented co-designed filtering antenna which is circularly polarized. The proposed antenna composed of patch radiator and bandpass filter. The patch in this case functions as both the radiator and last stage resonator of the filter. The filtering circuit not only provides frequency discrimination but also performances as impedance matching circuit. Figures 10 and 11 depicts the proposed antenna. The antenna operates in the frequency range of 3.77–4.26 GHz. The proposed antenna is designed as a square patch truncated at corners on the top and planner strip line microwave circuit at the bottom. As the antenna operates in close vicinity to human body, the strip line shape isolates the band pass resonators from the loading effect of human tissue.

Fig. 12 The proposed UWB antenna



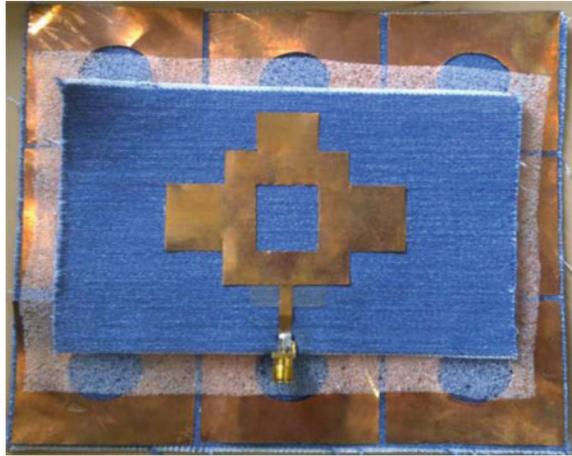
Jose and Kappan of [20] in their paper presented UWB wearable textile antenna with coplanar waveguide feed shown in Fig. 12. Coplanar feed is used to provide the antenna a uniplanar structure.

The antenna operates in the band of 3.1–10.6 GHz. The antenna is made-up on jeans substrate having relative permittivity 1.9. Conductive patch is designed using copper. Artificial magnetic conductor (AMC) is used at the back side of the antenna to give high gain over the entire range of operation. It provides high gain and reflection of the back lobes of back lobes. It acts as high impedance ground surface.

Velan et al. of [21] presented in their report a dual band fractal monopole antenna. The patch antenna is combined with electromagnetic band gap arrangement. The prototype antenna covers the GSM band of 1800 MHz and ISM band of 2.45 GHz. The EBG substrate creates a barrier for radiation into human body, also reduces the effect of frequency shifting due to human body.

Figure 13 represents the proposed fractal design of the monopole antenna. Here also the antenna is fabricated on jeans substrate having dielectric constant 1.7 and loss tangent 0.085. The antenna is backed with a EBG substrate of $150 \times 150 \text{ mm}^2$ dimension. For the whole frequency band the EBG substrate acts as band reject filter. The result is also analyzed under bending condition of the proposed antenna. SAR of the prototype is measured using three layer model of human body. The SAR value at 1.8 GHz with EBG is 0.024 and without EBG is 5.77 W/kg. At 2.45 GHz, the corresponding SAR values are 0.016 and 6.62 W/kg. Using EBG structure a significant reduction in SAR has been obtained in both the bands.

Fig. 13 Prototype fractal monopole antenna



Agarwal et al. of [22] have presented wearable antenna which is backed by artificial magnetic conductor operating in 2.4 GHz ISM band. The permittivity of latex substrate has been characterized to realize a flexible planar Yagi–Uda antenna printed on it using a large area screen printing process. The 0° reflection phase of single-layered and double layered (AMC and D-AMC) surfaces are proposed to minimize the maximum specific absorption rate (SAR) level at 2.4-GHz frequency band.

Jiang et al. of [23] presented in their paper a compact circularly polarized antenna using truncated anisotropic meta surface. In the proposed antenna structure loop monopole is at the top and a metallic sheet backed by meta surface layer is at the bottom. These two are detached by a foam spacer. The proposed antenna operates at ISM band of 2.45 GHz. Polydimethylsiloxane and silver nanowire clusters are used for building the antenna, which are tremendously flexible.

3 Conclusion

A brief review of wearable antenna is presented. Several different aspects have to be taken into account when designing wearable antenna. Several different materials can be taken into account. Wearable antennas have great future mainly due to the rapid development of wireless communication technologies. Table 3 lists the study on wearable antenna, their performance and different materials used as substrate.

Table 3 Summary of wearable antennas and their performance

Ref. No.	Author and year	Method used for size reduction	Size	Frequency of operation	Dielectric material used	Dielectric constant
[1]	Salonen and Hurme (2003)	High frequency operation and use of textile material	Patch is 56 mm × 51 mm, Ground plane is 76 mm × 71 mm	Flexible to operate at different frequency bands	Textile	1.04 at 2.45 GHz
[2]	Klemm et al. (2004)	Truncated corner along with single slit in patch and single inset microstrip feed line	Patch width 48 mm, length 5 mm, feed width 14 mm	2.29–2.36 GHz, centre frequency 2.32 GHz	Spacer fabric textile	1.15 at 2.4 GHz
[3]	Noury et al. (2005)	Introducing input stub	Antenna length and width both are 86.4 mm, thickness 1.6 mm	Centre frequency 868 MHz	Glass epoxy	4.8
[4]	Waterhouse et al. (2006)	Coplaner waveguide fed meander line slot	4.5 in. × 3 in.	Centre frequency 430 MHz	FR4	4.4
[5]	Sanz et al. (2006)	Cylindrical structure	Standard denim button size	Dual band of 2.4 and 5 GHz	Velero substrate of 1.8 mm thick	1.37
[7]	Mathews et al. (2009)	Bowtie structure, Spiral structure	Standard size	Band of 250–800 MHz	Nylon, liquid crystal polymer	2.9
[8]	Sagor et al. (2009)	Coplaner waveguide fed UWB antenna	6 mm × 16 mm × 0.068 mm	3.1–10.6 GHz	Liquid crystal polymer	2.9
[9]	Sankaralingam et al. (2009)	Circular disk structure	Radius 26.3 mm	2.45 GHz	Indigo jeans cotton fabric	1.67 at 2.45 GHz
[24]	Fang et al. (2011)	Rectangular spiral patch	Patch width 4 mm	402–406 MHz	Patch-Cu, Diele.-Roger RO 3210	10.2

(continued)

Table 3 (continued)

Ref. No.	Author and year	Method used for size reduction	Size	Frequency of operation	Dielectric material used	Dielectric constant
[25]	Duan et al. (2012)	Rectangular Spiral patch	27 mm × 14 mm × 1.27 mm	433.9–542.4 MHz	Roger 6010	10.2
[26]	Kiourti et al. (2012)	vertically stacked circular structure	6 mm radius ground plane, two 5 mm radius meandered patch	402–406 MHz, 902–928 MHz	Roger 3210	10.2
[27]	Wnuk et al. (2012)	Truncated corners	51.262 mm × 51.262 mm	1.8 GHz	Roger RT	2.6
[11]	Joshi et al. (2013)	Inter digital capacitor and rectangular stub	0.213 λ × 0.246 λ	2.45 GHz	Polyster cloth	1.39
[28]	Sagor et al. (2013)	Cu radiating patch	6 mm × 16 mm × 0.068 mm	3.1–10.6 GHz	Liquid crystal polymer	2.9
[29]	Dwivedi (2013)	Symmetric L shaped arms are joined together	90 mm × 90 mm	1.31–3.21 GHz	Bed sheet fabric	1.8
[15]	Mondol et al. (2014)	Split ring shaped slot is designed	18.11 mm × 18.11 mm	Centre frequency 5.25 GHz	Transparent acrylic fiber sheet	2.8
[17]	Shubair et al. (2015)	Stacked triangular patches and slotting of the patches	39 mm × 39 mm × 2.1 mm	2.4–2.48 GHz	FR4	4.4
[18]	Varmoosfaderani et al. (2015)	Shorted patch and printed wall made from ACHESON ELECTRODAG PM 496	35 mm × 22 mm × 9 mm	2.2–6.44 GHz	FR4	4.4
[19]	Jiang et al. (2015)	Truncated corner on the top and planner stripline microwave circuit at the bottom	40 mm × 40 mm × 5.3 mm	3.77–4.26 GHz	FR4	4.4

(continued)

Table 3 (continued)

Ref. No.	Author and year	Method used for size reduction	Size	Frequency of operation	Dielectric material used	Dielectric constant
[21]	Velan et al. (2015)	Electromagnetic band gap structure	150 mm × 150 mm	1800 MHz and 2.45 GHz	Jean fabric	1.7
[22]	Agarwal et al. (2016)	Use of Artificial Magnetic Conductor	50 mm × 50 mm	2.4 GHz	Latex	3.31
[23]	Jiang et al. (2016)	Truncated anisotropic metasurface	50 mm × 50 mm	2.4 GHz	PDMS	2.8

References

1. Salonen, P., Hurme, H.: A novel fabric WLAN antenna for wearable applications. In: Antenna and Propagation Society International Symposium, IEEE (2003)
2. Klemm, M., Locher, I., Troster, G.: A novel circularly polarized textile antenna for wearable applications. In: European Conference on wireless technology, Amsterdam, (2004)
3. Noury, N., Barralon, P., Flammarion, D.: Preliminary results on the study of Smart Wearable Antenna. In: Proceedings of the IEEE (2005)
4. Waterhouse, R., Novak, D.: Small uniplanar antenna suitable for body wearable applications. In: IEEE Military Communication Conference (2006)
5. Sanz-Izquierdo, B., Huang, F., Batchelor, J.C.: Small size wearable button antenna. EuCAP (2006)
6. Kim, Y., Lee, K., Kim, Y., Chung, Y.C.: Wearable UHF RFID tag antenna design using flexible electro thread and textile. In: IEEE Antennas and Propagation Society Symposium (2007)
7. Matthews, J.G.G., Pettitt, G.: Development of flexible, wearable antennas. In: 3rd European Conference on Antenna and Propagation (2009)
8. Sagor, M.H., Abbasi, Q.H., Alomainy, A., Hao, Y.: Compact conformal ultra-wideband antenna for wearable applications. In: 5th European Conference on Antennas and Propagation (2009)
9. Sankaralingam, S., Gupta, B.: A circular disk microstrip WLAN antenna for wearable applications. In: IEEE India Conference (2009)
10. Lee, S., Kim, U., Kwon, K., Seo, W., Choi, J.: Design of on-body antenna for wireless body area network. In: Advance Communication Technology (ICACT), pp. 300–303 (2012)
11. Joshi, J.G., Pattnaik, S.S.: Polyester based wearable microstrip patch antenna. In: IEEE conference (2013)
12. Sabban, A.: New wideband printed antennas for medical application. *IEEE Trans. Antennas Propag.* **61**(1), 84–91 (2013)
13. Mansor, M.M., Rahim, S.K.A., Hashim, U.: A CPW-fed 2.45 GHz wearable antenna using conductive nanomaterials for on-body applications. In: IEEE Region 10 Symposium (2014)
14. Lim, E.G., Wang, Z., Wang, J.C., Leach, M., Zhou, R., Lei, C.-U., Man, K.L.: Wearable Textile Substrate patch antennas. *Eng. Lett.* **22**:2, EL_22_2_08 (2014)
15. Mondol, B., Chatterjee, A., Parui, S.K.: A wearable button antenna with FSS superstrate for WLAN health care applications. In: RF and Wireless Technologies for Biomedical and Healthcare Applications (IMWS-Bio), IEEE (2014)
16. Hammoodi, A.I., Al-Rizzo, H.M., Isaac, A.A.: A wearable dual-band square slot antenna with stub for ISM and WiMAX applications. In: IEEE Conference (2015)
17. Shubair, R.M., AlShamsi, A.M., Khalaf, K., Kiourti, A.: Novel miniature wearable microstrip antennas for ISM-band biomedical telemetry. In: Loughborough Antennas & Propagation Conference (LAPC) (2015)
18. Varnoosfaderani, M.V., Thiel, D.V., Lu, J.: A wideband capacitively fed suspended plate antenna for wearable wireless sensors. In: International Symposium on Antenna and Propagation, IEEE (2015)
19. Jiang, Z.H., Werner, D.H.: A compact, wideband circularly polarized co-designed filtering antenna and its application for wearable devices with low SAR. *IEEE Trans. Antenna Propag.* **63**, 3808–3818 (2015)
20. Jose, A., Kappan, S.J.: High gain coplanar feed ultra-wide band wearable antenna using artificial magnetic conductor. In: Fifth International Conference on Advances in Computing and Communications, IEEE (2015)
21. Velan, S., Sundarsingh, E.F., Kanagasabai, M., Sarma, A.K., Raviteja, C., Sivasamy, R., Pakkathillam, J.K.: Dual band EBG integrated monopole antenna deploying fractal geometry for wearable applications. *IEEE Antennas Wireless Propag. Lett.* **14**, 249–252 (2015)
22. Agarwal, K., Guo, Y.-X., Salam, B.: Wearable AMC backed near end-fire antenna for on body communications on latex substrate. In: IEEE Transactions on Components, Packaging and Manufacturing Technology (2016)

23. Jiang, Z.H., Werner, D.H.: A compact and robust circularly polarized wearable antenna using an anisotropic metasurface. In: IEEE/ACES International Conference on Wireless Information Technology and Systems (ICWITS) and Applied Computational Electromagnetics (ACES) (2016)
24. Fang, Q., Lee, S.Y., Parmana, H., Ghorbani, K., Cosic, I.: Developing a wireless implantable body sensor network in MICS band. *IEEE Trans. Inf. Technol. Biomed.* **15**(4), 567–576 (2011)
25. Duan, Z., Guo, Y.X., Xue, R.F., Je, M., Kwong, D.L.: Differentially fed dual band implantable antenna for biomedical applications. *IEEE Trans. Antennas Propag.* **60**(12), 5587–5595 (2012)
26. Kiourti, A., Nikita, K.S.: Miniature scalp implantable antennas for telemetry in MICS and ISM bands: design, safety considerations and link budget analysis. *IEEE Trans. Antennas Propag.* **60**(8), 3568–3575 (2012)
27. Wnuk, M., Bugaj, M., Przesmycki, R., Nowosielski, L., Piwowarczyk, K.: Wearable antenna constructed in microstrip technology. In: Progress in Electromagnetics Research Symposium Proceedings (2012)
28. Sagor, M.H., Abbasi, Q.H., Alomayni, A., Hao, Y.: Compact and conformal ultra wideband antenna for wearable applications. In: EuCAP (2013)
29. Dwibedi, K., Patel, P., Poonkuzhali, R., Thiripurasundari, D., Alex, Z.C.: Dual band CPW fed wearable monopole antenna. In: ICT (2013)
30. Chisca, S., Sava, I., Musteata, V.E., Bruma, M.: Dielectric and conduction properties of polyimide films. In: CAS 2011 Proceedings (2011 International Semiconductor Conference), pp. 253–256, IEEE (2011)

Characterization of Silicon Carbo-Nitride Thin Films

Dhruva Kumar and Bibhu Prasad Swain

Abstract This review paper intends to through a light on different characterization techniques used for silicon carbonitride (SiCN). During the recent years, the interest in silicon carbo-nitride developed remarkably. The structural and chemical network of SiCN thin films are characterized by XRD, Raman and FTIR characterization respectively. Nanoindentation is used for analysing mechanical properties of thin film.

Keywords SiCN · XRD · XPS · FTIR · Raman · Nanoindentation

1 Introduction

Design and synthesis of new materials having exceptional electrical, mechanical or corrosion resistance properties is a big challenge for materials scientists across world. The production of the designed material possessing these qualities is far greater challenge. The excellent properties of silicon carbo-nitride (SiCN) make them a very desirable material to be used in various industrial applications. SiCN coating can be used as wear resisting coatings material, super abrasives tools to be used in precision grinding due to its extra ordinary high hardness. High thermal conductivity of SiCN makes it a promising material to be used as heat sinks material. Its property to withstand high temperature makes it a suitable for sensor material to be used in high temperature environment. SiCN is anticipated to behave like a semiconductor of varying band gap which depends on its atomic arrangement

D. Kumar (✉)

Department of Mechanical Engineering,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: dhruva.ju@gmail.com

B.P. Swain

Centre for Material Science and Nano Technology,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: bibhuprasad.swain@gmail.com

and composition, or in the production of nanotubes [1–6]. The above applications can be fulfilled by selective mechanical properties such as hardness, coefficient of friction, young modulus and suitable adhesion with coating elements.

2 Characterization of Silicon Carbon Nitride Thin Films

This section reviews the different characterisation processes used for silicon carbon nitride thin films.

2.1 X-Ray Diffraction (XRD)

Information like phase identification of a crystalline material, physical property and chemical composition of materials and the films can be obtained by structural analysis of material, thus making it a very important parameter. Observation of scattering radiation from sample is used to form the basis of this technique. XRD and Raman spectroscopy are used for structural analysis of materials. The basis of whole process is elastic scattering of X-rays from the electron clouds for each atom in the system (Fig. 1).

Fainer et al. investigated composite of SiC_xN_y and graphite (SiCN -graphite) at a temperature of 1000 °C for 1 h in argon and found strong diffraction peak at $2\theta = 26^\circ$, indicating a lattice of typical graphite [8]. Chen et al. investigated matrix of amorphous SiC_xN_y and nanocomposites of carbon nanoclusters by decomposing polymeric precursors with thermally induced phase separation and with the help of XRD analysis they concluded that ratio of starting material does not have any effect and the nanocomposites were amorphous [9]. Wang et al. investigated polysilazane-synthesized amorphous SiC_xN_y ceramics at different temperatures

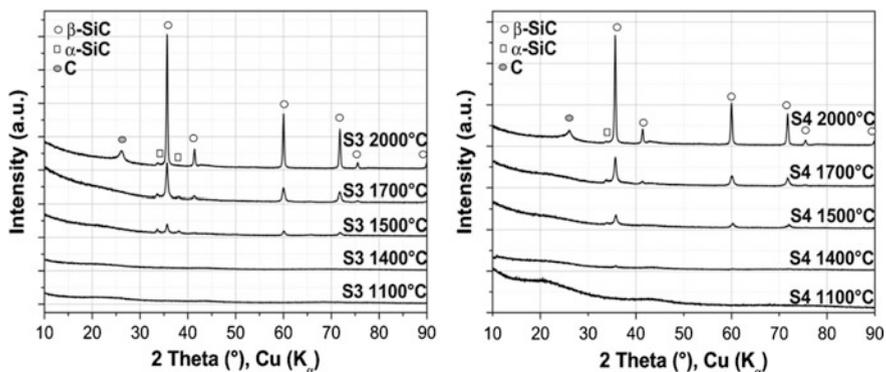


Fig. 1 Typical X-ray diffractograms of SiCN [7]

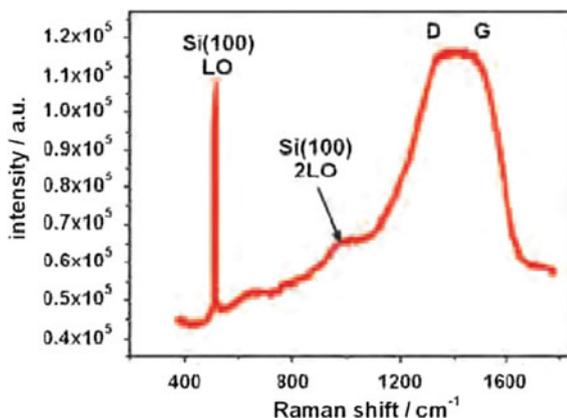
between 900 and 1400 °C and observed no diffraction peaks in XRD patterns of all materials implying that the ceramics were amorphous for the tested range of temperature [10]. Mera et al. investigated carbon-rich SiC_xN_y ceramics at different temperatures. They observed that upto 1500 °C, the ceramic S_2 found out to be amorphous but as soon as temperature was increased, some fragment of $\alpha\text{-Si}_3\text{N}_4$ converted into crystalline SiC and its composition comprised of $\alpha\text{-Si}_3\text{N}_4$, $\alpha\text{-C}$, and $\alpha\text{-}$ and $\beta\text{-SiC}$. At 1700 °C $\beta\text{-SiC}$ was separated [11].

2.2 Raman Spectroscopy

Raman spectroscopy technique is used for studying rotational, molecular vibration of a system. The process is based on inelastically scattered light, or Raman scattering of monochromatic light, generally from a laser in the visible, close to infrared or ultraviolet range (Fig. 2).

Chen et al. studied SiC_xN_y by thermally decomposing commercially available polysilazane and observed D peak at $\sim 1350\text{ cm}^{-1}$ and a G peak at $\sim 1600\text{ cm}^{-1}$ and because of presence of amorphous carbon, several minor peaks were also observed such as a D'' peak at 1500 cm^{-1} and a D' peak at 1620 cm^{-1} [12]. Prasad et al. reported thermal decomposition of SiC_xN_y and observed G band in the range of $2500\text{--}2800\text{ cm}^{-1}$ at 1700 °C [13]. Chen et al. studied amorphous SiC_xN_y and found peaks at 1600 and 1356 cm^{-1} corresponding to the G and D band of graphite [9]. Bhattacharyya et al. investigated nanocomposite SiC_xN_y and studied Raman spectroscopy of film deposited at 500 °C and found G peak at 1522 cm^{-1} and D disorder peak at 1390 cm^{-1} [14]. Mera et al. investigated carbon-rich SiC_xN_y ceramics and found D and D' bands at 1350 and 1620 cm^{-1} and G band at $\sim 1582\text{ cm}^{-1}$ [11]. Kaspar et al. studied carbon-rich polymer-derived SiC_xN_y . They observed D band at 1350 cm^{-1} and G band at 1590 cm^{-1} and G' and D' band at 2680 and 2920 cm^{-1} , respectively [15].

Fig. 2 Typical Raman spectrum of SiCN



2.3 Fourier Transforms Infrared Spectroscopy (FTIR)

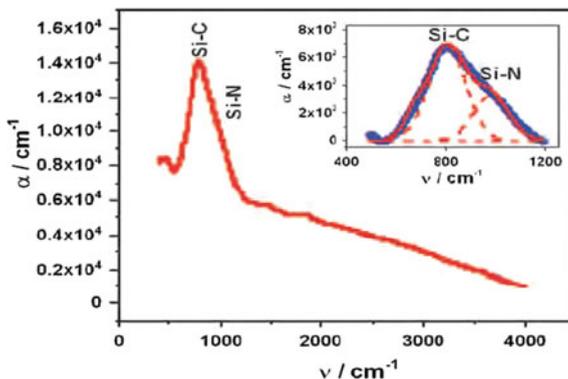
Bonds or a group of bonds always vibrate at characteristic frequencies which form the basic principle of FTIR. Molecule absorbs infrared energy characteristics to their frequencies when exposed to infrared rays (Fig. 3).

Fainer et al. investigated PECVD deposited films of SiC_xN_y and observed vibrations of the Si–C at 800 cm^{-1} , Si–N at 950 cm^{-1} , and Si–O–Si 1030 cm^{-1} bonds. Furthermore, they observed vibrations of the bond Si–H at $\sim 2200\text{ cm}^{-1}$ and Si–CH₃ at $\sim 1250\text{ cm}^{-1}$ [16]. Gao et al. reported nanostructure evaluation of carbon-rich SiC_xN_y and SiBCN polymer-derived ceramics and they found Si–N stretching vibration at 970 cm^{-1} , Si–C stretching at 812 cm^{-1} , C=N vibration at 1720 cm^{-1} and C–N at 1253 and 1188 cm^{-1} [17]. Demin et al. applied a new method of the plasma chemical vapour deposition for the synthesis of SiCN films on a substrate of stainless steel and instrumental alloys using hexamethyldisilazane vapours. The vibrational modes of Si–C and Si–N were observed at 750 cm^{-1} and 960 cm^{-1} , respectively, peaks observed in the region of $1600\text{--}3400\text{ cm}^{-1}$ referred to the vibration modes of the carbon containing HMDS fragments: C–H, C–C ($1400\text{--}1600\text{ cm}^{-1}$); $\text{Csp}_2\text{--H}$, $\text{Csp}_3\text{--H}$ ($2890\text{--}3000\text{ cm}^{-1}$). The band at 2100 cm^{-1} corresponded to the vibrations of Si–H bonds [18].

2.4 X-Ray Photoelectron Spectroscopy (XPS)

X-ray photoelectron spectroscopy (XPS) measures the elemental composition, electronic state, empirical formula and chemical state of the elements that exist within a material. When a material is exposed by a beam of X-rays, XPS spectra is obtained with the help of the kinetic energy and number of electrons that escape from the top 0 to 10 nm of the material (Fig. 4).

Fig. 3 FTIR transmission spectrum of SiCN [7]



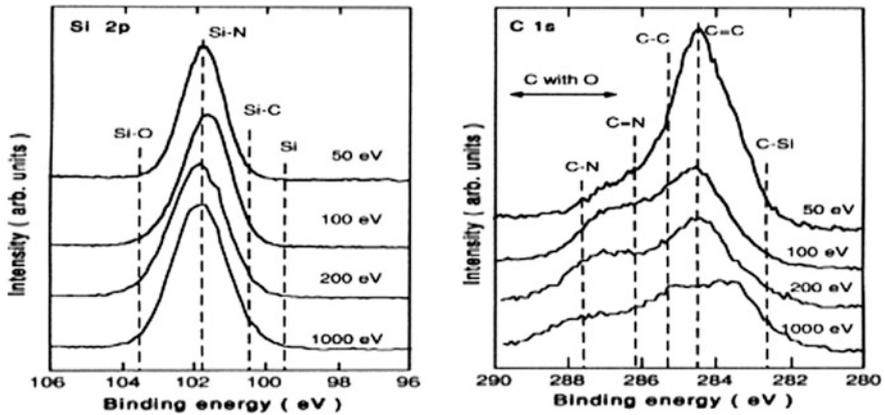


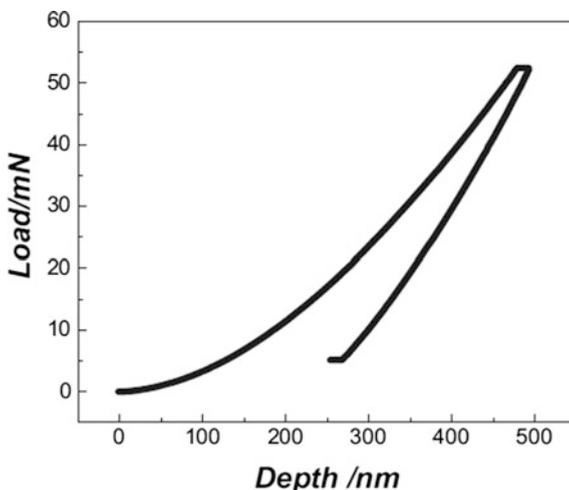
Fig. 4 N(1s) is appeared at 398 eV, C(1s) is appeared at 284 eV, Si(2p) is appeared at 101 eV [19]

Fainer et al. investigated plasma chemical decomposed SiC_xN_y films and found that at 373 K, Si-Si, Si-C-N, and Si-O chemical bonds were obtained corresponding to the energies bonds of 98.7, 101.2, and 102.8 eV. They also found N-Si bond at 397.9 eV, N-C bond at 400.2 eV [20]. Bulou et al. studied PECVD synthesized amorphous SiC_xN_y thin films and observed Si(2p) was decomposed into two components centred at 101.3 and 101.8 eV and C(1s) also has two components assigned to C(1s)/Si bonds (283 eV) and C(1s)/C/H at (284.5 eV) [21]. Bulou et al. investigated MPACVD-deposited amorphous SiC_xN_y thin films using N_2 , Ar, CH_4 and vaporised HMDSN gases and found major part of N atoms is bonded to Si and observed C-Si bonds at 283 eV, C-C and/or C-H bonds 284.5–284.7 eV, respectively [22]. Rumyantsev et al. investigated energy dispersive spectroscopy (EDS) deposited SiC_xN_y thin films and analysed XPS spectra of C1s, N1s and Si2p. Si2p XPS spectrum is deconvoluted into three components Si-O bond at 102.8 eV energy, Si-N-C at 101.1 eV and Si-Si bond at 98.8 eV, N1s spectrum at 397.5 eV energy corresponds to N-Si bond, 399.9 eV energy to N-C (sp^2) bond and the C1s spectrum at 281.9, 283.7 and 284.6 eV, which correspond to C-Si and C-C bonds [23].

2.5 Nanoindentation

The nanoindentation is used for computing the hardness, and the elastic modulus of the material. Generally Berkovich indenter, i.e. three-sided pyramid diamond is used. The principal of nanoindentation lies in bringing a very small tip to the material surface producing an imprint (Fig. 5).

Fig. 5 A typical load depth curve corresponding to the indentation showing the highest hardness of 22 GPa for SiC_xN_y thin films [24]



Jedrzejowski et al investigated mechanical properties of PECVD deposited SiC_xN_y films by using SiH_4 , CH_4 , N_2 , Ar gas mixtures at a temperature of 400 °C. The SiCN thin films exhibits a hardness >30 GPa, Young's modulus >190 GPa, elastic rebound of 85% and a compressive stress of approximately 1 GPa [25]. Bhattacharyya et al. investigated magnetron sputtering deposited SiC_xN_y coatings on silicon (100) by using DC and AC radio frequency current. The hardness of the SiC_xN_y thin films was 16 GPa for both Si substrates and 304 SS, respectively, for DC magnetron sputtering. However films deposited in RF mode showed higher hardness of 22–25 and 240–250 GPa modulus on Si and SS substrates [26]. Chen et al. studied the mechanical properties of SiC_xN_y materials prepared by microwave PECVD and electron cyclotron resonance PCVD. They found the hardness and the effective modulus of the SiCN crystals and amorphous SiC_xN_y as 30 and 321.7 GPa, and 22 and 164.4 GPa, respectively [27]. Saha et al. fabricated SiC_xN_y having iron dispersed in it and found that it resisted oxidation up to 500 °C and had a hardness of 5–7 GPa [28]. Leo et al. investigated the influence of process parameters like temperature and pressure and found that the hardest SiC_xN_y ceramics having hardness as 23 GPa were achieved when pure CERASET™ is thermally set at a temperature of 240 °C, cross linked at 2,000 psi (13.8 MPa) and pyrolyzed at 1,200 °C [29].

3 Conclusion

The results of various characterizations show the property of SiCN to withstand high temperature, it retained its amorphous property upto 1500 °C but with increase in temperature it converted into crystalline. Its oxidation resistance was also

validated by results. Thermal stability and enhanced mechanical properties was also reported. The hardness and the effective modulus of the SiCN crystals and amorphous SiC_xN_y was found to be 30 and 321.7 GPa, and 22 and 164.4 GPa respectively.

References

1. Ryu, H., Wang, Q., Raj, R.: Ultrahigh-temperature semiconductors made from polymer-derived ceramics. *J. Am. Ceram. Soc.* **93**, 1668–1676 (2010)
2. Bhattacharyya, A., Mishra, S., Mukherjee, S., Das, G.: A comparative study of Si–C–N films on different substrates grown by RF magnetron sputtering. *J. Alloys Compd.* **478**, 474–478 (2009)
3. Bai, X., Zhang, C., Tao, A.: Superhard multilayers and nanocomposite coatings. *Prog. Nat. Sci.* **105**, 97–107 (2005)
4. Hoche, H., Pusch, C., Riedel, R., Fasel, C., Klein, A.: Properties of SiCN coatings for high temperature applications—comparison of RF-, DC- and HPPMS-sputtering. *Surf. Coat. Technol.* **205**, S21–S27 (2010)
5. Leo, A., Andronenko, S., Stiharu, I., Bhat, R.: Characterization of thick and thin film SiCN for pressure sensing at high temperatures. *Sensors* **10**, 1338–1354 (2010)
6. Chen, Y., Li, C., Wang, Y., Zhang, Q., Xu, C., Weid, B., An, L.: Self-assembled carbon–silicon carbonitride nanocomposites: high-performance anode materials for lithium-ion batteries. *J. Mater. Chem.* **21**, 18186–18190 (2011)
7. Mera, G., Riedel, R., Poli, F., Müller, K.: Carbon-rich SiCN ceramics derived from phenyl-containing poly (silylcarbodiimides). *J. Eur. Ceram. Soc.* **29**, 2873–2883 (2009)
8. Fainer, N., Golubenko, A., Rummyantsev, Y., Kesler, V., Ayupov, B., Rakhlin, V., Voronkova, M.: Tris (diethylamino) silane—a new precursor compound for obtaining layers of silicon carbonitride. *Glass Phys. Chem.* **38**, 15–26 (2012)
9. Chen, Y., Li, C., Wang, Y., Zhang, Q., Xu, C., Weid, B., An, L.: Self-assembled carbon–silicon carbonitride nanocomposites: high performance anode materials for lithium-ion batteries. *J. Mater. Chem.* **21**, 18186–18190 (2011)
10. Wang, Y., Ding, J., Feng, W., An, L.: Effect of pyrolysis temperature on the piezoresistivity of polymer-derived ceramics. *J. Am. Ceram. Soc.* **94**, 359–362 (2011)
11. Mera, G., Tamayo, A., Nguyen, H., Sen, S., Riedel, R.: Nanodomain structure of carbon-rich silicon carbonitride polymer-derived ceramics. *J. Am. Ceram. Soc.* **93**, 1169–1175 (2010)
12. Chen, Y., Yang, X., Cao, Y., Gan, Z., An, L.: Quantitative study on structural evolutions and associated energetics in polysilazane-derived amorphous silicon carbonitride ceramics. *Acta Mater.* **72**, 22–31 (2014)
13. Prasad, R., Mera, G., Morita, K., Müller, M., Kleebe, H., Gurlo, A., Fasel, C., Riedel, R.: Thermal decomposition of carbon-rich polymer-derived silicon carbonitrides leading to ceramics with high specific surface area and tunable micro- and mesoporosity. *J. Eur. Ceram. Soc.* **32**, 477–484 (2012)
14. Bhattacharyya, A., Mishra, S., Mukherjee, S.: Correlation of structure and hardness of rf magnetron sputtered silicon carbonitride films. *J. Vac. Sci. Technol.* **28**, 505 (2010)
15. Kaspar, J., Mera, G., Nowaka, A., Zajaca, M., Riedel, R.: Electrochemical study of lithium insertion into carbon-rich polymer-derived silicon carbonitride ceramics. *Electrochim. Acta.* **56**, 174–182 (2010)
16. Fainer, N., Kosyakov, V.: Phase composition of thin silicon carbonitride films obtained by plasma enhanced chemical vapour deposition using organosilicon compounds. *J. Struct. Chem.* **56**(1), 163–174 (2015)

17. Gao, Y., Mera, G., Nguyen, H., Morita, K., Kleebe, H., Riedel, R.: Processing route dramatically influencing the nanostructure of carbon-rich SiCN and SiBCN polymer-derived ceramics. Part I: Low temperature thermal transformation. *J. Eur. Ceram. Soc.* **32**, 1857–1866 (2012)
18. Demin, V., Smirnova, T., Borisov, V., Grachev, G., Smirnov, A., Khomyakov, M.: Physical chemical properties of silicon carbonitride films prepared using laser plasma deposition from hexamethyldisilazane. *Glass Phys. Chem.* **41**, 232–236 (2015)
19. Yamamoto, K., Koga, Y., Fujiwara, S.: XPS studies of amorphous SiCN thin films prepared by nitrogen ion-assisted pulsed-laser deposition of SiC target. *Diamond Relat. Mater.* **10**, 1921–1926 (2001)
20. Fainer, N., Golubenko, A., Rumyantsev, Y., Kesler, V., Maximovskii, E., Ayupov, B., Kuznetsov, F.: Synthesis of silicon carbonitride dielectric films with improved optical and mechanical properties from tetramethyldisilazane. *Glass Phys. Chem.* **39**, 77–88 (2013)
21. Bulou, S., Brizoual, L., Miska, P., Poucques, L., Hugon, O., Belmahi, M., Bougdir, J.: The influence of CH₄ addition on composition, structure and optical characteristics of SiCN thin films deposited in a CH₄/N₂/Ar/hexamethyldisilazane microwave plasma. *Thin Solid Films* **520**, 245–250 (2011)
22. Bulou, S., Brizoual, L., Miska, P., Poucques, L., Hugon, R., Belmahi, M., Bougdira, J.: Structural and optical properties of a-SiCN thin film synthesised in a microwave plasma at constant temperature and different flow of CH₄ added to HMDSN/N₂/Ar mixture. *Surf. Coat. Technol.* **205**, S214–S217 (2011)
23. Rumyantsev, Y., Fainer, N., Maximovskii, E., Ayupov, B.: Elemental composition analysis of silicon carbonitride thin films by energy dispersive spectroscopy. *J. Struct. Chem.* **51**, S179–S185 (2010)
24. Bhattacharyya, A., Mishra, S., Mukherjee, S.: Correlation of structure and hardness of rf magnetron sputtered silicon carbonitride films. *J. Vac. Sci. Technol. A.* **28**, 505 (2010)
25. Jedrzejowska, P., Cizekb, J., Amassiana, A., Sapihaa, J., Vlcek, J., Martinua, L.: Mechanical and optical properties of hard SiCN coatings prepared by PECVD. *Thin Solid Films* **447–448**, 201–207 (2004)
26. Bhattacharyya, A., Das, G., Mukherjee, S., Mishra, S.: Effect of radio frequency and direct current modes of deposition on protective metallurgical hard silicon carbon nitride coatings by magnetron sputtering. *Vacuum* **83**, 1464–1469 (2009)
27. Chen, L., Chen, K., Weia, S., Kichambare, P., Wu, J., Luc, T., Kuo, C.: Crystalline SiCN: a hard material rivals to cubic BN. *Thin Solid Films* **355–356**, 112–116 (1999)
28. Saha, A., Shah, S., Raj, R., Russek, S.: Polymer-derived SiCN composites with magnetic properties, rapid communications. *J. Mater. Res.* **18**, 2549–2551 (2003)
29. Leo, A., Andronenko, S., Stiharu, I., Bhat, R.: Characterization of thick and thin film SiCN for pressure sensing at high temperatures. *Sensors* **10**, 1338–1354 (2010)

Improved Normalization Approach for Iris Image Classification Using SVM

Mahaboob Shaik

Abstract With the rapid improvement of information technology, security and authentication of individuals has become a greater significance. Iris recognition is one of the best solutions in providing unique authentication for individuals based on their IRIS structure. Iris normalization meant to extract the iris region and represent it in spatial domain, Daughman's rubber sheet model is so far a standard and efficient method of implementing this process. In this paper, a low complex, simpler and improved version of rubber sheet model is proposed. The main aim of this method is to minimize the complex computations that were involved in the conventional rubber sheet model and to provide an equivalent performing approach with very less computations. Classification performance is evaluated with CASIA and IIT Delhi IRIS databases using SVM classifier.

Keywords IRIS · Normalization · Rubber sheet model

1 Introduction

In present scenario, keeping the data secured and authenticated is most difficult and very important to any organization or individual. Most of the researchers are focusing and developing different new applications on biometrics. Among all the biometrics iris recognition is playing vital and unique role. This works according to the visual features of a person like furrows, freckles, corona, and rings. Due to the high variation of randomness in the above-mentioned features, recognition of iris is considered as highly problematic approach [1].

M. Shaik (✉)

Department of ECE, JJT University, Jhunjhunu, Rajasthan, India
e-mail: 44.maha@gmail.com

M. Shaik

Department of EEE, MJCET, Hyderabad, India

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_15

139

This organ is externally visible; it has unique quality of epigenetic pattern which stables throughout the adult life. This versatile character makes very attractive to use this as an identifying individual biometric application. By using different image processing techniques, the unique iris pattern can be extracted from a digitalized eye image, one can encode them into a biometric template and can be stored in database.

A working model for automated iris recognition system with generic information formulated, which was a successful system [2]. Other most of the popular recognition models were also proposed [3–7]. The Daughman’s system was tested under several studies, the result rate always lead to zero failure rate. By doing experiments on millions of irises, Daughman’s system always gives perfect identification of individuals. By doing experiment on 520 iris images Wildes et al. who established the prototype system also gave flawless performance reports. By taking the database of 6000 eye images another researcher Lim et al. proposed a model by having recognition of 98.4%.

The main objective of this paper is to present low complexity based IRIS normalization method to minimize the complex computations that were involved in the conventional rubber sheet model and to provide an equivalent performing approach with very less computations. Here, we use Log Gabor filters for extracting the features from normalized images. Feature extraction of this IRIS images was implemented using Log Gabor filters. A performance comparison is presented with CASIA [8] and IIT Delhi [9] IRIS databases using SVM classifier.

The proposed model follows as, Sect. 1 gives a brief introduction about the physical dimensions and the importance of IRIS recognition system continuing with Sect. 2 which clear depicts the implementation of Daugman’s rubber sheet model. Section 3 explains neatly about the proposed approach of minimizing the complexity involved in the Daugman’s model, Sect. 4 presents the performance analysis of the proposed approach along with other classifiers and databases.

2 Rubber Sheet Model (Daugman’s)

Here, Daughman’s integro-differential operator is used to locate the coordinates of iris boundary. First,

- r_{\min} is the possible minimum,
- r_{\max} is maximum radius.

And the λ = iris searching parameter. For the nonideal database of WVU, 50% of the image size will be set for r_{\max} , and 33% of r_{\min} will be set for r_{\max} . Using λ , the operator will locate the change in maximum pixel intensity value (J) in the image which lies in the allocated radius with a circular integral of (x_0, y_0) , the original image blurred by a Gaussian kernel (G) with the given radius r (Fig. 1).

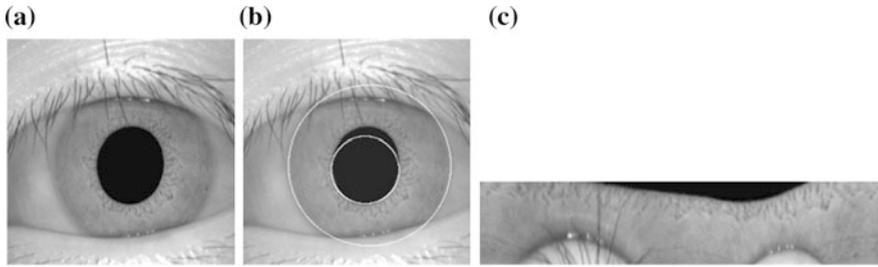


Fig. 1 a Input iris. b Iris localization. c Normalized image

$$J(r, x_0, y_0) = G_\sigma(r) * \left(\frac{d}{dr} \right) \int \frac{I(x, y)}{2\pi r} ds \quad (1)$$

$$G_\sigma(r) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-r^2}{2\sigma^2}\right) \quad (2)$$

In this model, the remap can be done by each point within the iris region by taking a pair of polar coordinates. This model will be accountable or responsible for imaging distance, pupil dilation, and nonconcentric pupil misplacement.

3 Proposed Iris Recognition System

The proposed approach involves the following steps; (a) Iris Segmentation (b) Normalization of Iris (c) Feature Extraction (d) Classification.

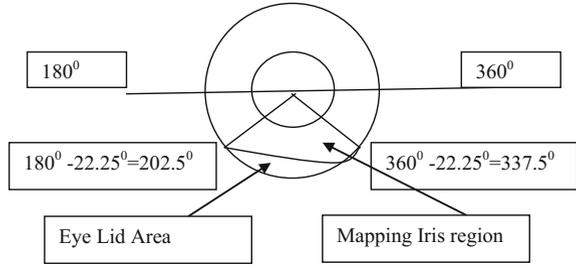
3.1 Iris Segmentation

In iris recognition system, segmentation is the basic step. Here from captured eye image we will isolate the iris region. Canny method locates edges by searching maxima of the gradient in iris image. The gradient is calculated by using derivative of Gaussian filter. Here two thresholds are used to detect the weak and strong edges. This approach is tougher to detect the true weak edges.

3.2 Iris Normalization

The proposed improved normalization approach involves in partitioning the circular region obtained through Daugman's Model. The approach divides the 360° circular

Fig. 2 Mapping area for the proposed approach



region into a trapezoidal portion as shown in Fig 2. This approach mainly considers the mapping area between 202.5° and 337.5° in the circular region. This region when segmented looks like a trapezoid as shown in Fig. 3c.

The proposed approach is as follows:

- Step 1: considering lower half iris, i.e., lower half circle: Starting from angle 180 to 360 (Lower circular portion).
- Step 2: Consider only 75% of the lower half circle, eliminating 25%, i.e., 22.5% from left and 22.5% from right, means from left 180 to $202.5 = 22.5\%$, and from right, i.e., from $360 - 22.5 = 337.5$, i.e., 22.5% deducted from right.
- Step 3: Now get the shape of trapezoid, and convert that trapezoid to rectangle by removing the triangular shapes from both the sides of Trapezoid.
- Step 4: Finally get a small rectangular normalized iris image for which the features are calculated.

3.3 Feature Extraction

In feature extraction, the fundamental information in iris pattern will generate a binary template which is used for classification and recognition. Iris feature set is generated by combing or grouping the pattern of iris with 1D Log Gabor filter [10]. This will be represented in mathematical terms as

$$G(f) = \exp\left(\frac{-\log\left(\frac{f}{f_0}\right)^2}{2\log\left(\frac{\sigma}{f_0}\right)}\right), \tag{3}$$

where f_0 = center frequency and σ = bandwidth of the filter.

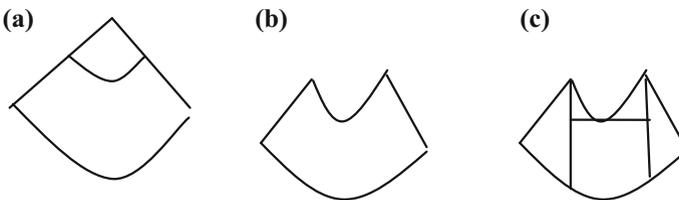


Fig. 3 a Cropped region from Fig. 2. b After removal of upper part. c Trapezoidal portion

3.4 SVM Classifier

Support vector machine is a wide-ranging classification tool which is used in many successful applications and here, it is used as classification tool.

4 Performance Analysis

The performance evaluation of this method is validated with CASIA Iris Image Database Version 1 [8]. In which for the experiment, 20 users of each three images are considered for training and two for testing, similarly the IIT Delhi database in

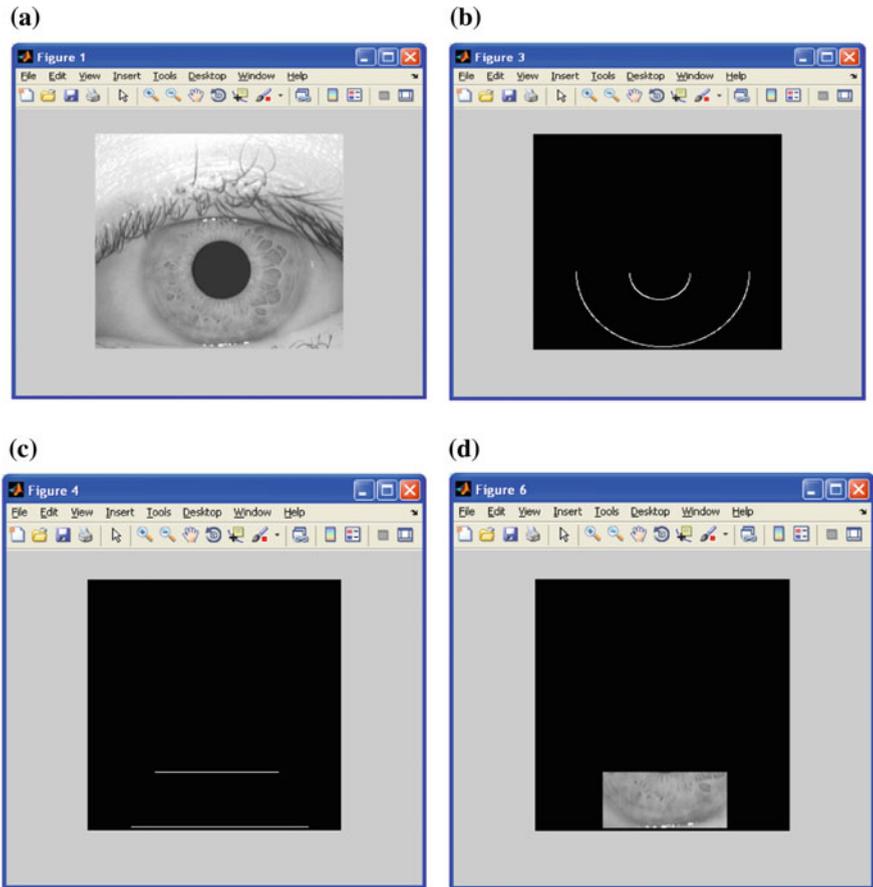


Fig. 4 a Input image. b Extraction of half circular region. c Conversion into trapezoidal region. d Extraction of iris in rectangular form

which 20 users three images of each are considered for training and the two for testing. Feature vector of each image was generated using the concept explained in Sect. 3. When classified with conventional rubber sheet model, the proposed approach shows an improvement of 7% decrement in processing time with almost same recognition rate (Figs. 4, 5 and 6).

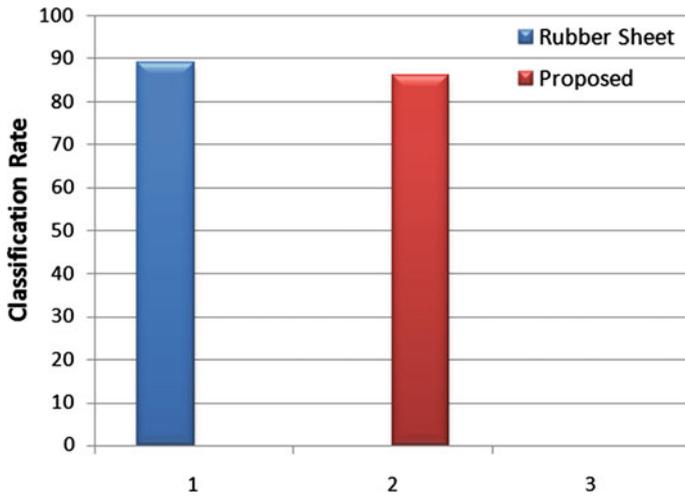


Fig. 5 Classification rate analysis for proposed normalization approach using SVM

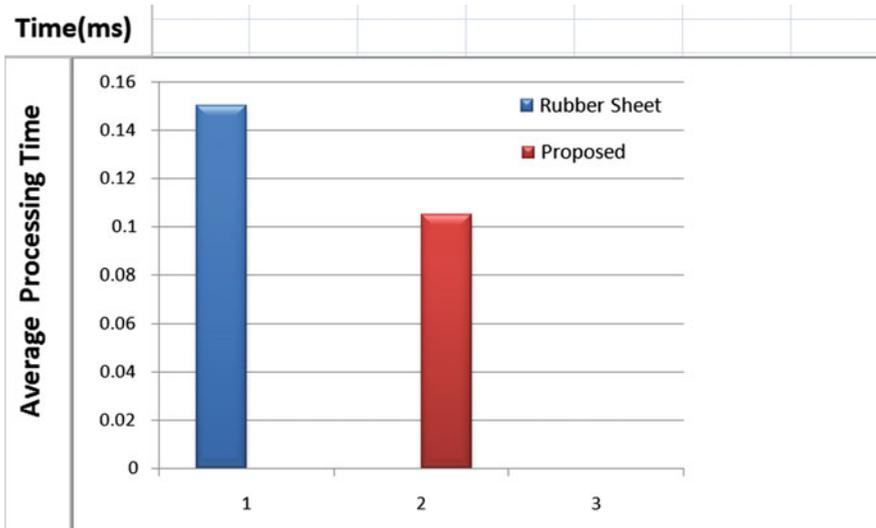


Fig. 6 Average processing time for normalization

For the evaluation, two parameters are considered first the classification rate and second average processing time for normalization. In the earlier analysis, it was found that classification is decreased from 89–86% with the proposed approach which is not a significant loss as per our knowledge. However, the average processing is decreased by 30% using on 70% of the CPU processing time, this is a significant improvement.

5 Conclusion

A new improved method of iris normalization is proposed in this method and the performance is validated in terms of classification rate and average processing time by evaluating with different datasets of iris images. The experimental findings prove that the proposed approach is consuming 70% of the processing when compared against the conventional rubber sheet model which makes it significant for further research analysis. The future scope lies in analyzing this approach with multiclass classifiers.

References

1. Seetharaman, K., Ragupathy, R.: Iris recognition based image authentication. *Int. J. Comput. Appl.* **44**(7) (2012)
2. Daugman, J.: How iris recognition works. In: *Proceedings of 2002 International Conference on Image Processing*. Vol. 1 (2002)
3. Wildes, R.: Iris recognition: an emerging biometric technology. In: *Proceedings of the IEEE*. Vol. 85, No. 9 (1997)
4. Wildes, R., Asmuth, J., Green, G., Hsu, S., Kolczynski, R., Matey, J., McBride, S.: A system for automated iris recognition. In: *Proceedings IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, pp. 121–128 (1994)
5. Boles, W., Boashash, B.: A human identification technique using images of the iris and wavelet transform. *IEEE Trans. Sig. Process.* **46**(4) (1998)
6. Lim, S. Lee, K., Byeon, O., Kim, T.: Efficient iris recognition through improvement of feature vector and classifier. *ETRI J.* **23**(2), Korea (2001)
7. Noh, S., Pae, K., Lee, C., Kim, J.: Multi-resolution independent component analysis for iris identification. *The 2002 International Technical Conference on Circuits/Systems, Computers.* (2002)
8. <http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp>
9. http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm
10. Azizi, A., Pourreza, H. R.: Efficient iris recognition through improvement of feature extraction and subset selection. *Int. J. Comp. Sci. Inf. Secur.* **2**(1) (2009)
11. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)

Investigation of Titanium Aluminium Nitride (TiAlN): A Review

Soham Das and Bibhu Prasad Swain

Abstract Titanium aluminium nitride (TiAlN) thin films are now commonly used as protective layer in various fields as these films have high thermal stability and sound corrosion resistance. Due to high hardness and wear resistance at elevated temperature, it is one of the most important thin film coating material for cutting tools. Various deposition methods such as physical vapour deposition (PVD), chemical vapour deposition (CVD), sputtering technique, etc., are used by many researchers to synthesize TiAlN coatings. To analyse morphological, structural, and mechanical characteristics, various tests such as SEM, AFM, xrd, Raman, and nanoindentation have been carried out by many researchers. This paper basically summarizes the properties of TiAlN monolayer, nanolayer, and multilayer thin films.

Keywords Titanium aluminium nitride · PVD · Microstructure
Mechanical properties

1 Introduction

For past few decades, nitride coatings have been emerged as one of the hardest and high temperature-resistant material in mechanical and industrial components [1]. In the recent years, it has also been emerged as one of the very suitable material for medical implants [2]. Among a wide variety of materials used in various fields, titanium nitride (TiN) suited the most for most of the requirements, but this material

S. Das (✉)

Department of Mechanical Engineering, Sikkim Manipal Institute of Technology,
Majitar, Sikkim, India
e-mail: sohamengg10@gmail.com

B. P. Swain

Centre for Material Science and Nano Technology, Sikkim Manipal
Institute of Technology, Majitar, Sikkim, India
e-mail: bibhuprasad.swain@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_16

147

is subjected to oxidation at high temperature. To overcome this problem, aluminium is mixed into the TiN network to form TiAlN coatings which in the last few years have shown improvement in the quality of films [3]. One of the most important properties which TiAlN possesses is its resistance to oxidation at high temperature up to 700–800 °C, which in case of titanium nitride with NaCl structure, undergoes oxidation at around 500 °C.

1.1 Titanium Aluminium Nitride

Nowadays, single-layer and multilayer TiAlN coatings are intensively investigated due to its sound mechanical and tribological properties which make this a suitable coated material for cutting hard materials with good degree in accuracy. TiAlN is a face-centred cubic (fcc) crystal system commonly used as cutting tool, forming tools, semiconductor devices, etc. The main advantage of TiAlN films is high adhesion strength and thick layer of Al₂O₃ film which prevents oxygen diffusion in the coated film [4].

2 Morphology of TiAlN

2.1 Atomic Force Microscopy (AFM)

Khelifi et al. deposited and characterized different layers of TiN and TiAlN coatings on steel substrate via magnetron sputtering system. Characterization through AFM led the researchers to conclude better surface finish of nanolayer coatings than single-layer nitride coating [5] (Fig. 1).

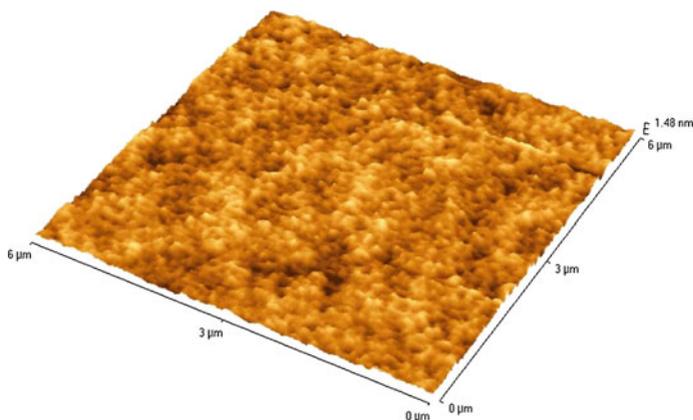


Fig. 1 Surface roughness of a sample obtained from AFM [5]

Dani et al. investigated the structure of TiAlN coatings prepared by magnetron sputtering method over silicon plate using a compound $Ti_{0.5}Al_{0.5}N$ target at varying N_2 pressure, bias voltage. AFM analysis revealed improved coating morphology with increasing voltage. Also it indicates coatings with pore-free surfaces with no micro-cracks [6]. Park et al. successfully replaced conventional TaN heater in inkjet printers by TiAlN film, over a Si (100) wafer at 400 °C in the presence of TiN and AlN as targets. 3D AFM studies revealed the highest RMS surface roughness at mid-level power density followed by highest power density. Later surface morphology found to be decreased slightly which further validated the mechanism of variation of grain size [7]. Buranawong et al. observed steep change in $AlTi_3N$ grain size (bigger triangle shape) over Si wafers with a slight change of current from 600 to 700 mA. Surface roughness for the deposited samples at different current sources was found to be 9.0, 11.0, and 17.4 nm [8]. Huang et al. investigated the after-effect of surface morphology on the properties of TiAlN films prepared over TiC material. Analysis of results revealed parent material and coating roughness within 10–62 nm. Further they also found substrate inhomogeneity and preferred nucleation growth as responsible parameter for coating surface roughness [9]. Chawla et al. deposited Ti–Al–Cr based coatings INCOLOY under reactive nitrogen atmosphere. AFM surface morphology (2D and 3D) results revealed TiAl-based coating with lower roughness value than Al–Cr based coatings. Also the grain size provided by AFM analysis reported a lesser particle size (10 nm) for TiAlN as compared to (25 nm) for AlCrN coating [10].

2.2 Scanning Electron Microscopy (SEM)

Sui et al. fabricated TiAlN, TiAlN-Si-based single-layer and multilayer coating using sputtering method. A change in the morphology of the deposited film from columnar to nanocrystal was observed by diffusing silicon elements. Multilayered structure of Ti–Al–Si based films also came up with reduced spalling problem [11] (Fig. 2).

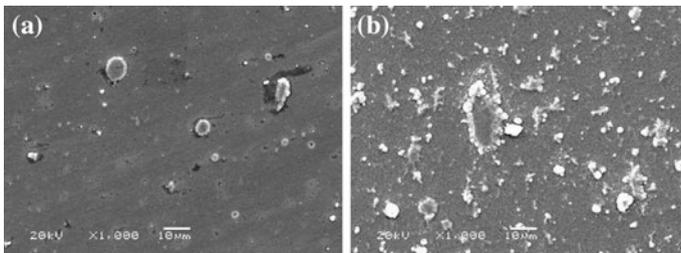


Fig. 2 Morphology of TiN and TiAlN [12]

Zouina et al. investigated the mechanical and chemical behaviour of TiAlN films in presence of NaCl (3.5 wt%). Microstructure has been studied using SEM micrographs, where coating thickness was found to have columnar structure with an estimated thickness of 1.2 μm . Also the coating was found to be quite dense, uniform and homogeneous [13]. Dani et al. investigated sputtering coated TiAlN films at different N_2 pressure and bias voltage. SEM images reveal coating thickness up to 5 μm , and the deposition rates up to 115.5 nm/min also coatings were observed with compact and pore-free surface [6]. Pei et al. investigated the stability of Ti_2AlN at pressure of 5 GPa and at different temperatures of 700–1600 $^\circ\text{C}$ by using SEM with EDS technology. For 1400 $^\circ\text{C}$ and 5 GPa, Ti_2AlN was found to be very stable with good quality microstructure. Further investigation also indicated direct decomposition of Ti_2AlN into TiN and TiAl at 5 GPa and 1500 $^\circ\text{C}$ [14]. Girleanu et al. examined $\text{Ti}_{1-x}\text{Al}_x\text{N}$ films on steel substrate for different aluminium compositions using radiofrequency magnetron reactive process. TEM along with FEA results indicated small grain size of TiN films, while elongated fibres were observed for TiAlN at 0.86% wt of Al., they also observed destruction of coating with further addition of aluminium [15]. Chawla et al. investigated the influence of varying Al content on the characteristics of $\text{Ti}_{1-x}\text{Al}_x\text{N}$ coating prepared over glass by sputtering technique. Structural analysis showed reduction in grain size with increasing aluminium content which was also found to be responsible for reduced RMS value of deposited thin film [16]. Lie et al. deposited microcrystalline diamond (MCD), nanocrystalline diamond (NCD), diamond-like carbon, and TiAlN layer over carbide product. FESEM results indicated coatings with good uniformity, where NCD was found to exhibit very small grain size as compared to MCD. Al_2O_3 particles up to a size of 1.5 μm were observed in TiAlN coating [17]. Zhang et al. proposed a new method for improved mechanical properties of cam body by depositing TiAlN thin layer over it using ion sputtering deposition technique. SEM images revealed clear interface and compacted film. The effect of substrate bias voltage over deposition was also demonstrated, where high substrate voltage was found responsible for damage of thin films [18]. Yan et al. synthesized TiAlN/TiB₂ multilayers with various modulation ratio on Al_2O_3 (111) substrate using ion beam assistant deposition. Images obtained through SEM and TEM clearly indicated the existence of TiAlN crystalline phase embedded over TiB₂ amorphous structure. They observed light and dark coloured TiAlN and TiB₂ layers with equal width which led them to the conclusion that the thickness ratio of deposited films have same modulation ratio [19].

3 Structure of TiAlN

3.1 X-ray Diffraction (XRD)

Xudong Sui et al. fabricated TiAlN and TiAlSiN using magnetron sputtering method. They observed a change from columnar growth to nanocrystal growth with the addition of Si element. Xrd images revealed the diffraction peaks of TiAlN in corresponding (111), (200), and (220) crystal planes. With addition of silicon element in TiAlN, the diffraction peaks were found to be stretched with reduced peak intensity. Later they found that TiAlN coating possesses hexagonal structure and TiAlSiN possesses mixed cubic and hexagonal structure [11] (Fig. 3).

Presence of aluminium in TiN enhances the solar radiation absorption capability of deposited thin film over stainless steel substrate, where the deposited TiAlN coating was found to contain small amount of crystalline phase with fine grained structure by Feng et al. [21]. Zirconium (Zr) plays very vital role regarding the improvement in crystallographic structure of TiAlN since the addition of Zr in TiAlN changes the columnar structure of film into compact isometric structure, which further enhances the texture coefficient of Zr/TiAlN approximately up to one [22]. Dejun et al. analysed and reported that addition of elemental aluminium in TiN improves hardness of TiAlN since the replacement of some Ti grains with Aluminium grains creates lattice distortion with increased grain extremities and dislocations. Also they reported that further addition of elemental silicon (Si) in TiAlN refines the grain structure with further improved hardness and wear resistance since the presence of Si_3N_4 along with TiAlSiN converted to wearable SiO_2 phase at elevated temperature when deposited through cathode arc ion plating process [23]. Peng et al. reported the presence of different diffraction peaks at

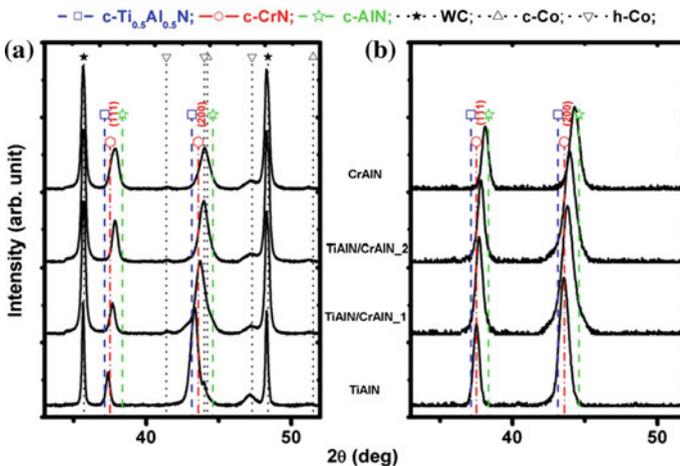


Fig. 3 XRD patterns of TiAlN, CrAlN coatings [20]

-36.7° (111), -42.8° (200), -61.8° (220), -74.1° (311) for TiAlN, -37.8° (111) for AlN and 70.6° (103) for α -Ti phases when deposited TiAlN films on Ti6Al4V alloy deposited through magnetron sputtering technique [24]. Rousseau et al. found higher machining capability of PVD and CVD coated TiAlN thin films over High speed steel substrate due to higher N_2 bombardment during deposition, which further leads to proper substrate diffusion. Through XRD authors found the major peaks at (111) and (200) regarding TiN and TiAlN, respectively [25]. Substrate bias (V_b) plays a very important role in the deposition of titanium aluminium nitride based films when synthesized through magnetron sputtering in existing Ar/ N_2 mixture, reported by Huang et al. They found that for both the coatings initially with increase in V_b , the diffraction peak from (111) shifted to a dominant (200) peak but with further increase in V_b they observed reversal of peak to (111) with small intensity and poor crystallinity. Further they explained that peaks at (200) possess lowest surface energy and appear relatively at lower stress level (low V_b), whereas peaks at (111) possess highest surface energy and appear relatively at higher stress level (high V_b) [26]. Wang et al. analysed the oxidation behaviour of TiAlN/SiN coatings on borosilicate glass deposited through magnetron sputtering method. Xrd images revealed successful deposition of TiAlN with corresponding diffraction peaks at 37.1° with (111) crystalline planes with columnar crystal structure but for SiN no, such peaks were observed and due to higher film density of SiN it was found that after oxidation at higher temperature SiN shows better oxidation resistance than TiAlN [27]. Pinot et al. compared the properties of TiAlN coatings prepared through sputtering using titanium aluminium targets with previously deposited TiAlN deposited films from mosaic (M) target comprises of titanium and aluminium material. Films deposited with higher Al content showed cubic xrd fingerprints, where the existences of different crystal lattice were concluded regarding sintered Al mixed films. With both the targets, they found decrement in mechanical properties with increasing friction damage due to growth competition between hcp and fcc crystal structure [28]. Twu et al. investigated the effect of titanium-based, chromium-based, and Ti-A based nitride films on cyclic load resisting behaviour for aluminium and medium carbon steel. Study revealed coating phases at (111), (200) and (220) peaks and CrN peaks were observed at (111) and (200) peaks with fcc structure [29]. Generally, the diffraction peak intensities of TiAlN coated films are found to be at (111) or (200) crystalline plane reported by Shang et al. [30]. Kumar et al. examined the microstructure of TiAlN/AlCrN coatings deposited on carbide material using vapour deposition technique. During analysis, TiN (111) plane found to have the dominant orientation with B1-NaCl crystal structure. TiN, AlN, and CrN exhibited polycrystalline structure with mixed orientations. Since the diffraction peaks were found to be unbroaden, they concluded the coatings to be nanocrystalline. Later, they also revealed a strain free coating due to no peak shift. The average crystallite size (D_p) was found to be 0.94 and the crystallite sizes of TiN, AlN, and CrN were calculated as 44.14, 35.65, and 17.98, respectively [31]. Zhan et al. reported successful deposition of TiAlON/TiAlN/TiAl with fcc crystal structure over HSS substrate through ion plating process. With variation in O_2 partial pressure they found that

the surface phase structure of above stated thin films are same, whereas the peak intensity of Al and Ti were found to enhance partially by enhancing O₂ pressure. Due to low oxygen content at coating surface, authors found it very difficult to explain the reasons regarding the existence of Ti and Al oxide on the thin film [32]. Sprute et al. demonstrated the influence of PVD-based TiAlN multilayer and monolayer coatings on SiC grinding, diamond grinding papers regarding their initial and final residual stress. Test results revealed TiAlN diffraction peaks at (220) with fcc-NaCl structure. Further, substrates with TiAlN multilayer coatings found to exhibit higher residual stress than the substrate with monolayer coatings [33]. Generally, it has been found that the hardness of TiAlN coating is maximum when the deposited coating exhibits smaller grain size with fcc crystal structure [34]. Deng et al. examined the influence of niobium and carbon ion on the hardness of TiAlN layers prepared through arc ion plating process. Test results indicated increased diffraction intensity ratio with Nb and C ion addition from where small change in structure also found. Authors also found NbN and TiC binary phase formation with reduced lattice parameters [35]. Aluminium and vanadium has a vice versa effect on TiN coatings deposited by magnetron reactive sputtering on M2 high speed steel reported by wang et al. they found reduction in lattice constant of TiN coatings with Al addition whereas the lattice constant increased with vanadium addition. Test results also indicated an improvement in columnar structure of TiN, TiAlN, TiAlVN with the addition of Al and V. Lastly some traces of ϵ (Fe₃N–Fe₂N) phase were also observed due to presence of Fe in the deposited coatings [36].

4 Compositional and Electronic Structure

4.1 X-ray Photoelectron Spectroscopy (XPS)

Yi et al. deposited multilayer TiAlN coating over Titanium aluminium vanadium based substrate. From XPS analysis, it was found that Ti 2p peak exhibited high binding energy correspond to 2p_{3/2} and 2p_{1/2} electrons whereas for TiN films, the peaks were found to be situated at 454.6 and 461.2 eV, respectively [24] (Fig. 4).

Ananthakumar et al. successfully deposited TiN/TiAlN multilayers of 2 mm thickness. Spectroscopy study revealed coexistence of all elements associated with the coating. It was also found that Ti 2p peak exhibits binding energy more than 400 eV at different orbitals [37]. Barshilia et al. used (DC) sputtering system to synthesize TiAlN coatings over drill bits in presence of N₂ + Ar plasma. Authors evaluated the performance of coated drill bit in presence and absence of coolant. From XPS analysis, they observed Ti peaks relatively at higher binding energy. Also a good bonding structure was observed within the coating [38]. Wosińska et al. investigated the microstructural and mechanical properties of PVD-coated (Ti, Al,Si)N film over ceramics substrate. Several tests such as SEM, XPS, and nanoindentation were carried out to investigate the coating properties. For

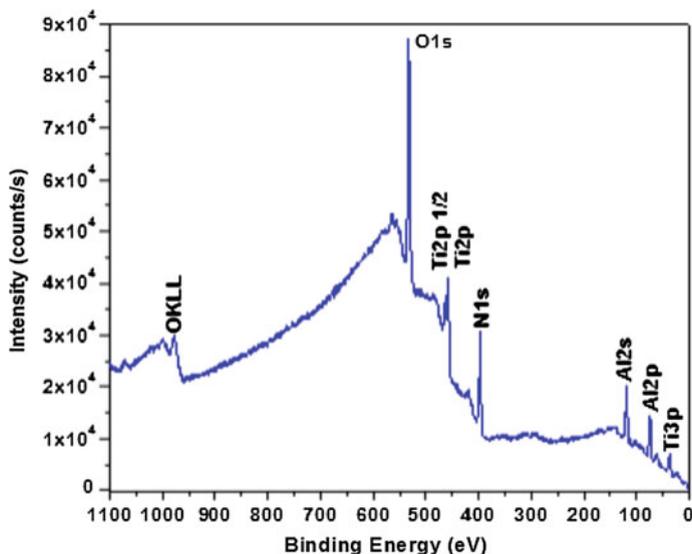


Fig. 4 XPS image of TiAl for different BE's [24]

mechanical properties, a comparison was carried out between uncoated substrate and gradient (Ti,Al,Si)N coated substrate. For 2.6 μm thickness on the substrate, the microhardness of (Ti,Al,Si)N coating (microhardness: 2650 HV and roughness: 0.18 μm) was found to be higher than uncoated substrate (microhardness: 1980 HV and roughness: 0.07 μm) in every aspect [39]. Rizzo et al. reported the xps results of single-layer, intermixed multilayer, and nano-multilayer TiAlN/TiN film deposited through reactive magnetron sputtering system. Results indicated the presence of three different components namely TiON, TiAlN, and TiN compounds, and N_2 surface absorbates in the range of 397–400 eV binding energy in N1S region. They also reported the presence of equal amount of Ti and Al percentage for the samples due to same deposition parameters. For multilayer and nanolayer films, low titanium and high aluminium content were also observed [40]. Gao et al. investigated the bonding status of titanium through xps while depositing TiAlN/ Al_2O_3 multilayer coatings using magnetron sputtering method. Test results indicated the presence of stoichiometric TiN and TiO_x peaks at binding energy of 453.8 and 457.8 eV in $\text{Ti}2p_{3/2}$ orbitals, where the attachment of Ti with oxygen was found to be more than nitrogen probably due to higher affinity [41].

4.2 Raman Spectroscopy

Ananthakumar et al. successfully deposited TiN/TiAlN multilayers of 2 μm thicknesses using reactive DC magnetron sputtering method. Result revealed that

coatings exhibited peaks at 209 and 318 cm^{-1} for different vibrational mode. Vibration of metal atoms during deposition was found to be the responsible factor for the spectrum in the frequency range of 200–300 cm^{-1} [37] (Fig. 5).

Subramanian et al. compared the properties of Ti–N, Ti–O–N, and Ti–Al–N regarding surface roughness, corrosion in simulated body fluid and cytotoxicity for biomedical applications. Raman spectra for all the coatings were found to within of 100–1100 cm^{-1} . For TiN films, vibration of big size Ti ions and smaller N_2 ions was responsible for scattering in acoustic range and optic range. Also the results witnessed the presence of anatase structure at 200, 398, and 641 cm^{-1} in the case of TiON coatings [42]. Feiler et al. doped Si and B with TiAlN for the improvement of TiAlN coatings by cathodic arc evaporation technique. Both anatase and rutile phase titania were detected through Raman spectroscopy, when oxidized TiAlN films for 60 min and 900 °C. But in later stage when oxidized for 300 min rutile phase found to be the dominant factor over anatase phase. In case of Ti–Al–B–N and Ti–Al–Si–N films, same conditions had been observed with a slight shifting of peaks of Raman spectra [43]. Subramanian et al. deposited 1 μm TiAlN on MS substrates with equal titanium and aluminium content. Through RAMAN spectra, the characteristics peaks at 312.5, 675, and 1187.5 cm^{-1} at different optical modes of TiAlN were observed [44].

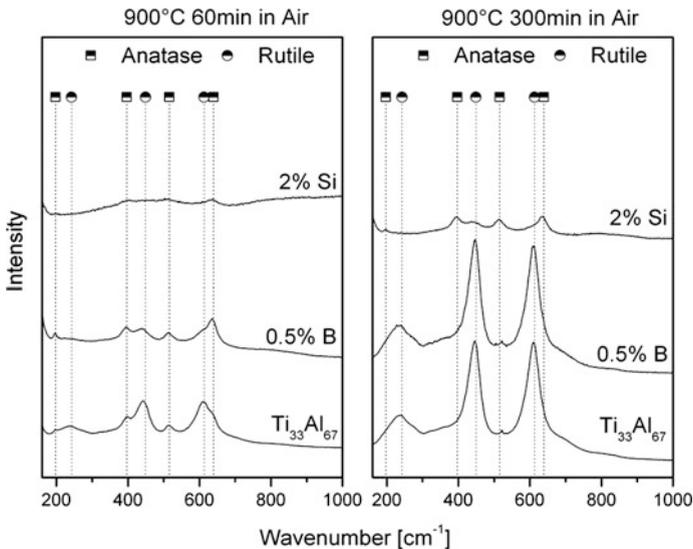


Fig. 5 Raman spectra of TiAlN films [44]

5 Conclusion

Above literature clearly indicates the importance of TiAlN thin film coatings in various fields (manufacturing, medical) as these coating provides high hardness, wear resistance at elevated temperature with enhanced service life along with the parent material. In most of the cases, these films have been found to exhibit fcc columnar crystal structure where the diffraction peaks were generally observed at (111) and (200). The average nanohardness value of TiAlN films has been found as 30–35 GPa with higher modulus of elasticity and low friction coefficient.

References

1. Gago, R., Soldera, F., Hubner, R., Lehmann, J., Munnik, F., Vazquez, L., Redondo-Cubero, J. A., Endrino, L.: X-ray absorption near-edge structure of hexagonal ternary phases in sputter-deposited TiAlN films. *J. Alloy. Comp.* **561**, 87–94 (2013)
2. Major, B., Major, R., Bruckert, F., Lackner, J.M., Ebner, R., Kustos, R., Lacki, P.: New gradient coatings on TiN and TiCN basis for biomedical application to blood contact. *Adv. Mater. Sci.* **7**, 63–70 (2007)
3. Gnoth, C., Kunze, C., Hans, M., Baben, M., Emmerlich, J., Schneider, J.M., Grundmeier, G.: Surface chemistry of TiAlN and TiAlNO coatings deposited by means of high power pulsed magnetron sputtering. *J. Phys. D Appl. Phys.* **46**, 8 (2013)
4. Subramanian, B., Ananthakumar, R., Jayachandran, M.: Microstructural, mechanical and electrochemical corrosion properties of sputtered titanium–aluminum–nitride films for bio-implants. *Vacuum* **85**, 601–609 (2010)
5. Khlifi, K., Ben, A., Larbi, C.: Mechanical properties and adhesion of TiN monolayer and TiN/TiAlN nanolayer coatings. *J. Adhes. Sci. Technol.* **28**, 85–96 (2014)
6. Dani, S., Sava, S.S.: Relation between coating parameters and structural and mechanical properties of magnetron sputtered TiAlN coatings. *Arab. J. Sci. Eng.* **39**, 5025–5034 (2014)
7. Park, M.-H., Kim, S.-H.: Temperature coefficient of resistivity of TiAlN films deposited by radio frequency magnetron sputtering. *Trans. Nonferrous Mat. Soc.* **23**, 433–438 (2013)
8. Buranawong, A., Witit-anun, N., Chaiyakun, S., Pokaipisit, A., Limsuwan, P.: The effect of titanium current on structure and hardness of aluminium titanium nitride deposited by reactive unbalanced magnetron co-sputtering. *Thin Solid Films* **519**, 4963–4968 (2011)
9. Huang, R.-X., Qi, Z.-B., Sun, P., Wang, Z.-C., Wu, C.-H.: Influence of substrate roughness on structure and mechanical property of TiAlN coating fabricated by cathodic arc evaporation. *Phys. Procedia* **18**, 160–167 (2011)
10. Chawla, V., Puri, D., Prakash, S., Chawla, A., Sidhu, B.S.: Characterization and comparison of corrosion behavior of nanostructured TiAlN and AlCrN coatings on superfer 800H (INCOLOY 800 H) substrate. *J. Miner. Mater. Charact. Eng.* **8**, 715–727 (2009)
11. Sui, X., Li, G., Qin, X., Yu, H., Zhou, X., Wang, K., Wang, Q.: Relationship of microstructure, mechanical properties and titanium cutting performance of TiAlN/TiAlSiN composite coated tool. *Ceram. Int.* **42**, 7524–7532 (2016)
12. Sun, G.D., Yi, D.L., Liu, C.H.: Comparison of surface properties of TiN and TiAlN coating prepared by Arc Ion plating for the improvement of life-time extension of tool steel. *Adv. Mater. Res.* **887–888**, 1096–1100 (2014)
13. Zouina, A., Djafera, A., Saoula, N., Aknouche, H., Guedoua, B., Madaoui, N.: Deposition and characterization of titanium aluminum nitride coatings prepared by RF magnetron sputtering. *Appl. Surf. Sci.* **350**, 6–9 (2015)

14. Pei, A.N., Zhilei, H.E., Jiaqian, Q., Ziyang, L., Yongjun, L., Zili, K., Duanwei, H.E.: Stability of titanium-aluminium Nitride (Ti₂AlN) at high pressure and high temperatures. *J. Wuhan Univ. Technol. Mater. Sci. Ed.* **26**, 914–919 (2011)
15. Girleanu, M., Pac, M.J., Louis, P., Ersen, O., Werckmann, J., Rousselot, C., Tuilier, M.H.: Characterization of nano-structured titanium and aluminium nitride coatings by indentation, transmission electron microscopy and electron energy loss spectroscopy. *Thin Solid Films* **519**, 6190–6195 (2011)
16. Chawla, V., Chandra, R., Jayaganthan, R.: Effect of phase transformation on structural, electrical and hydrophobic properties of nano-composite Ti_{1-x}Al_xN films. *J. Alloy. Compd.* **507**, L47–L53 (2010)
17. Lei, X., Wang, L., Shen, B., Sun, F., Zhang, Z.: Comparison of chemical vapor deposition diamond-, diamond-like carbon- and TiAlN-coated micro-drills in graphite machining. *J. Eng. Manuf.* **227**, 1299–1309 (2013)
18. Zhang, H., Wang, S., Guo, P., Wang, M.: Microstructure and wear properties analysis of TiAlN film deposited on Cam profile using ion sputtering. *Tribol. Trans.* **56**, 968–976 (2013)
19. Yan, J.Y., Sun, Y.D., Li, D.J., Liu, M.Y., Dong, L., Cao, M., Gao, C.K., Wang, N., Deng, X. Y., Gu, H.Q., Wan, R.X., Sun, X.: High-temperature stability of TiAlN/TiB₂ multilayers grown on Al₂O₃ substrates using IBAD. *Surf. Coat. Technol.* **229**, 105–108 (2013)
20. Li, P., Chen, L., Wang, S.Q., Yang, B., Du, Y., Li, J., Wu, M.J.: Microstructure, mechanical and thermal properties of TiAlN/CrAlN multilayer coatings. *Int. J. Refract. Met. Hard Mater.* **40**, 51–57 (2013)
21. Feng, J., Zhang, S., Lu, Y., Yu, H., Kang, L., Wang, X., Liu, Z., Ding, H., Tian, Y., Ouyang, J.: The spectral selective absorbing characteristics and thermal stability of SS/TiAlN/TiAlSiN/Si₃N₄ tandem absorber prepared by magnetron sputtering. *Solar Energy* **111**, 350–356 (2015)
22. Xian, G., Zhao, H.B., Fan, H.Y., Du, H.: Structure and mechanical properties of Zr/TiAlN films prepared by plasma-enhanced magnetron sputtering. *Rare Met.* **300**, 717–724 (2014)
23. Dejun, K., GuiZhong, F.U.: Nanoindentation analysis of TiN, TiAlN, and TiAlSiN coatings prepared by cathode ion plating. *Technol. Sci.* **58**, 1–9 (2015)
24. Yi, P., Peng, L., Huang, J.: Multilayered TiAlN films on Ti6Al4V alloy for biomedical applications by closed field unbalanced magnetron sputter ion plating process. *Mater. Sci. Eng.* **59**, 669–676 (2016)
25. Rousseau, A.F., Partridge, J.G., Mayes, E.L.H., Tona, J.T., Kracica, M., McCulloch, D.G., Doyle, E.D.: Microstructural and tribological characterisation of a nitriding/TiAlN PVD coating duplex treatment applied to M2 high speed steel tools. *Surf. Coat. Technol.* **272**, 403–408 (2015)
26. Huang, H., Li, Z., Wang, M., Xie, C.: Microstructural and mechanical properties of TiAlN and Ti₃AlN films deposited by reactive magnetron sputtering. *Mater. Sci. Forum* **816**, 283–288 (2015)
27. Wang, R., Li, J., Wang, Y., Hu, J., Wu, H.: High temperature oxidation behavior and mechanical properties of TiAlN/SiN decorative films on borosilicate glass by magnetron sputtering. *Thin Solid Films* **584**, 72–77 (2015)
28. Pinot, Y., Pac, M.J., Henry, P., Rousselot, C., Odarchenko, Y.I., Ivanov, D.A., Bouillet, C.U., Ersen, O., Tuilier, M.H.: Friction behavior of TiAlN films around cubic/hexagonal transition: A 2D grazing incidence X-ray diffraction and electron energy loss spectroscopy study. *Thin Solid Films* **577**, 74–81 (2015)
29. Twu, M.J., Hu, C.C., Liu, D.W., Hsu, C.Y., Kuo, C.G.: Effect of TiN, CrN and TiAlN coatings using reactive sputtering on the fatigue behaviour of AA2024 and medium carbon steel specimens. *J. Exp. Nanosci.* **11**, 581–592 (2015)
30. Shanga, H., Li, J., Shaoa, T.: Mechanical properties and thermal stability of TiAlN/Ta multilayer film deposited by ion beam assisted deposition. *Appl. Surf. Sci.* **310**, 317–320 (2014)
31. Kumar, T.S., Prabu, S.B., Manivasagam, G.: Metallurgical characteristics of TiAlN/AlCrN coating synthesized by the PVD PROCESS on a cutting insert. *J. Mater. Eng. Perform.* **23**, 2877–2884 (2014)

32. Zhan, J., Cai, J., Zhao, W., Zhang, Z.: Microstructure and mechanical properties of TiAlON/TiAlN/TiAl films. *Appl. Mech. Mater.* **628**, 93–97 (2014)
33. Sprute, T., Tillmann, W., Grisales, D., Selvadurai, U., Fischer, G.: Influence of substrate pre-treatments on residual stresses and tribo-mechanical properties of TiAlN-based PVD coatings. *Surf. Coat. Technol.* **260**, 369–379 (2014)
34. Longn, Y., Zeng, J., Yu, D., Wu, S.: Microstructure of TiAlN and CrAlN coatings and cutting performance of coated silicon nitride inserts in cast iron turning. *Ceram. Int.* **40**, 9889–9894 (2014)
35. Deng, B., Pei, J., Tao, Y.E.: Microstructure, mechanical and tribological properties of the TiAlN coatings after Nb and C dual ion implantation. *Mater. Sci. Forum* **789**, 455–460 (2014)
36. Wang, C.F., Ou, S.F., Chiou, S.Y.: Microstructures of TiN, TiAlN and TiAlVN coatings on AISI M2 steel deposited by magnetron reactive sputtering. *Trans. Nonferrous Met. Soc. China* **24**, 2559–2565 (2014)
37. Ananthakumar, R., Subramanian, B., Kobayashi, A., Jayachandran, M.: Electrochemical corrosion and materials properties of reactively sputtered TiN/TiAlN multilayer coatings. *Ceram. Int.* **38**, 477–485 (2012)
38. Barshilia, H.C., Yogesh, K., Rajam, K.S.: Deposition of TiAlN coatings using reactive bipolar-pulsed direct current unbalanced magnetron sputtering. *Vacuum* **83**, 427–434 (2009)
39. Wosińska, L.A.L., Mięka, J., Gołombek, K., Gawarecki, T.: Investigation of hard gradient PVD (Ti, Al, Si)N coating. *J. Achiev. Mater. Manuf. Eng.* **24**, 79–85 (2007)
40. Rizzo, A., Mirengi, L., Massaro, M., Galiotti, U., Capodice, L., Terz, R., Tapfer, L., Valerini, D.: Improved properties of TiAlN coatings through the multilayer structure. *Surf. Coat. Technol.* **235**, 475–483 (2013)
41. Gao, C.K., Yan, J.Y., Dong, L., Li, D.J.: Influence of Al₂O₃ layer thickness on high-temperature stability of TiAlN/Al₂O₃ multilayers. *Appl. Surf. Sci.* **285**, 287–292 (2013)
42. Subramanian, B., Muraleedharan, C.V., Ananthakumar, R., Jayachandran, M.: A comparative study of titanium nitride (TiN), titanium oxy nitride (TiON) and titanium aluminum nitride (TiAlN), as surface coatings for bio implants. *Surf. Coat. Technol.* **205**, 5014–5020 (2011)
43. Feiler, M.P., Zechner, J., Penoy, M., Michotte, C., Kathrein, M.: Improved oxidation resistance of TiAlN coatings by doping with Si or B. *Surf. Coat. Technol.* **203**, 3104–3110 (2009)
44. Subramanian, B., Ashok, K., Kuppasami, P., Sanjeeviraja, C., Jayachandran, M.: Characterization of reactive DC magnetron sputtered TiAlN thin films. *Cryst. Res. Technol.* **43**, 1078–1082 (2008)

The Study of GPR with Conducting Surfaces to Determine the Operating Frequency in Single-Fold Reflection Profiling

Amitansu Pattanaik and Rajat Sharma

Abstract Mapping of different soil surfaces is possible by the means of ground penetrating radar (GPR). To optimise the interpretation of such radar surveys carried out at various sites, we turned our attention to the accuracy of GPR on different geophysical surfaces. The properties of soil significantly influence the GPR performance, modifying, in particular, the propagation velocity, skin depth, phase constant and attenuation. Thus, setting the parameters for a GPR survey is not easy. Also, it is difficult to decide the right operating frequency for GPR operations. In order to get a higher penetration depth, the user will have to compromise with the resolution of the GPR or vice versa. Here, we try to estimate an appropriate operating frequency for GPR survey, if the properties of the soil are known.

Keywords Ground penetrating radar · GPR · Single-fold reflection profiling
Propagation velocity · Skin depth

1 Introduction

The ground penetrating radar is becoming increasingly successful in geophysics, civil engineering, archaeological investigation, defence, etc. It is a tool that is used to access information about subsurface properties. GPR techniques are potentially powerful, non-destructive techniques for the measurement of surface features. It helps the measurement of surface dielectric constant, water content, target thickness, depth, etc. GPR can also be used against dangerous geological hazards. It is because events like subsurface faulting, contamination of groundwater (man-made

A. Pattanaik (✉)

Defence Terrain Research Laboratory (DTRL), DRDO, Metcalfe House,
Delhi 110054, India
e-mail: amitansu@yahoo.com

R. Sharma

Department of Physics, Alagappa University, Karaikudi 630003, Tamil Nadu, India
e-mail: rajat.2890@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_17

159

or natural) or underground cavities can easily be studied by a GPR. It can also characterise subsurface contaminations produced by hazardous materials. Also, GPR is a very helpful tool for landslide studies and predictions [1-3]. The success of a GPR survey is influenced by the data collection parameters. It controls the properties of the electromagnetic wave propagated in the ground and the reflection of the reflected wave at the surface. Another factor that influences the success of a GPR survey is the properties of the deposits, through which the wave is propagated. Chemical or physical properties of the subsurface deposits like compaction, water content, and presence of conducting materials, such as soluble salts, clays and metals can influence the properties of the electromagnetic and design parameters of the GPR [4-14].

In this paper, estimation of phase velocity and skin depth has been done using the wave number equations when the GPR is operated under different frequencies. At those frequencies, designing parameters will be measured and an appropriate operating frequency is estimated for GPR applications. Resolution is an important factor, thus high frequencies are used.

2 Theoretical Background

Theory of a GPR is explained by the Maxwell's equations. Full solution for the electromagnetic waves is moderately used in GPR processing. The approximation of low loss conditions is broke and in order to make more accurate results, the full solution for the wavenumber equation should be used. It is used to estimate the phase velocity and the skin depth. GPR is most efficient when the losses associated with conduction current are minimal and the displacement current dominates [15].

But first, terms α and β should be defined. They are called attenuation and phase constant, respectively [16]. They are given by Eqs. (1) and (2), respectively:

$$\alpha = \omega \sqrt{\frac{\mu\epsilon}{2} \left\{ \sqrt{\left(1 + \frac{\sigma^2}{\epsilon^2\omega^2}\right)} - 1 \right\}} \quad (1)$$

And;

$$\beta = \omega \sqrt{\frac{\mu\epsilon}{2} \left\{ \sqrt{\left(1 + \frac{\sigma^2}{\epsilon^2\omega^2}\right)} + 1 \right\}}, \quad (2)$$

where ω is called the angular frequency (it is equal to $2\pi f$; f is frequency). Electrical conductivity (S/m) is represented by σ , and ϵ is called the dielectric permittivity. Dielectric permittivity can be expressed as $\epsilon_0\epsilon_r$, where ϵ_0 is called the dielectric permittivity of free space (8.854×10^{-12} F/m). ϵ_r is the dimensionless quantity, called the relative dielectric permittivity. Similarly, μ is the magnetic

permeability and expressed as $\mu_0\mu_r$, where μ_0 is called the magnetic permeability of free space ($4\pi \times 10^{-7}$ H/m). μ_r is the dimensionless quantity, called the relative magnetic permeability. Most near surface materials μ_r is almost one [17], we will assume that to be the case for this paper.

The complete expressions of skin depth and phase velocity are given by Eqs. (3) and (4), respectively [18].

$$\delta = \frac{1}{\omega \sqrt{\frac{\mu\epsilon}{2} \left\{ \sqrt{\left(1 + \frac{\sigma^2}{\epsilon^2\omega^2}\right)} - 1 \right\}}} \quad (3)$$

And:

$$v = \frac{1}{\sqrt{\frac{\mu\epsilon}{2} \left\{ \sqrt{\left(1 + \frac{\sigma^2}{\epsilon^2\omega^2}\right)} + 1 \right\}}} \quad (4)$$

The loss tangent is given by Eq. (5):

$$\tan \theta = \frac{\sigma}{\epsilon\omega} \quad (5)$$

Loss tangent is the tangent of the loss angle. For GPR, the conduction current is very small because under low loss conditions the loss tangent $\ll 1$. Similarly under the quasi-static case, the loss tangent $\gg 1$ and the displacement current is very small. When the system is operating between the quasi-static and low loss condition, it is operating in the transition band [18].

When there is no loss of energy from the electric field across the material, then the material is called a perfect dielectric material. They can be gaseous, liquid or solid materials. Since there are no propagation losses, thus there is no consideration of attenuation. A material in which electrical hysteresis losses are present is called an imperfect dielectric material. In the situation of ground penetrating radar, these losses are linked with conductivity effects [19]. Table 1 provides the electrical conductivity and the relative dielectric permittivity of few clay samples [20]. Electrical conductivity increases with the increase in water content of the sample.

3 Variations

At first, we plot the variation of propagation velocity and relative dielectric permittivity with varying frequencies (Fig. 1). In GPR surveys, the propagation velocity is an important factor in the determination of depth of the target, wavenumber, refractive index, two-way transit time, etc. Thus, it determines the speed at which the survey is conducted.

Table 1 Relative dielectric constant and electrical conductivity of clay sample

Material	Water content (% by volume)	Clay content (% by weight)	Main minerals	Electrical conductivity σ ($\mu\text{S/cm}$)	Relative dielectric permittivity ϵ_r
Kaolinitic clay	23	20	Kaolinite chloride	40	11
	30			60	12.5
	36			100	15
Eddy clay	32	14	CaCO ₃ quartz	140	19
	36			240	22
	38			560	26
Houston black clay	38	15.1	CaCO ₃ quartz	700	33
	43			640	35
	46			1400	38
Beaumont clay	46	16.2	Quartz muscovite	1840	42
	48			1950	45
	50			1800	47

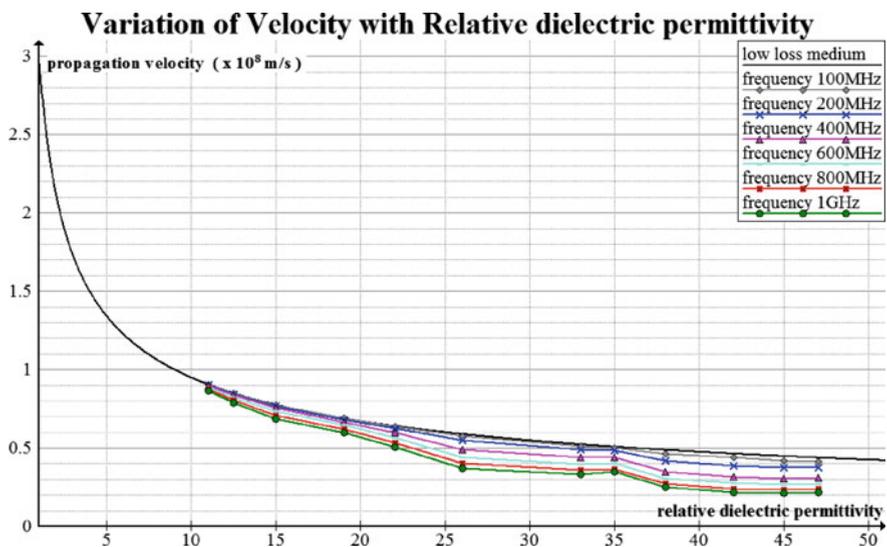


Fig. 1 Propagation velocity/relative dielectric permittivity

Then, the skin depth is compared with relative dielectric permittivity for the same frequencies (Fig. 2). The depth to which a GPR can operate is of real importance. If the target is deeper than the range of GPR, then it has no practical application.

And then the loss tangent is varied with relative dielectric permittivity for the same frequencies (Fig. 3). Loss tangent is proportional to the dielectric loss. It signifies that the electromagnetic energy is converted to heat energy because the materials are lossy.

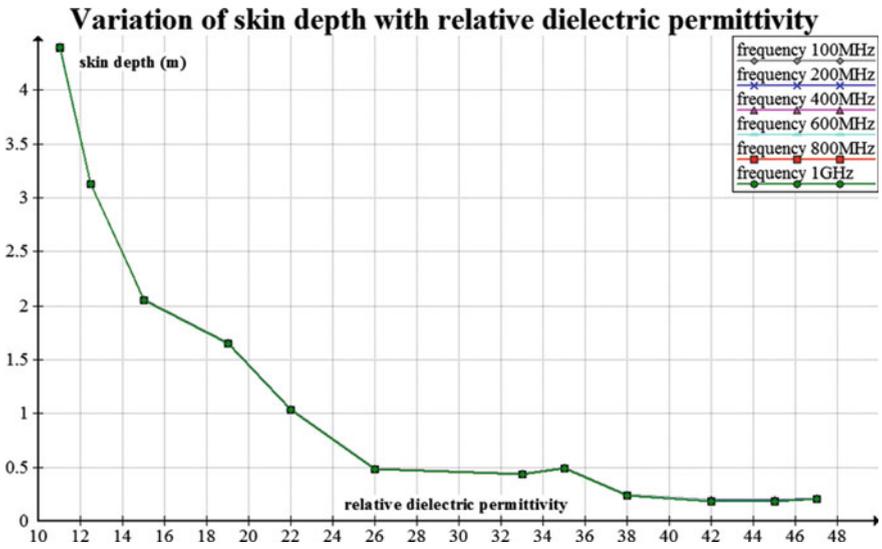


Fig. 2 Skin depth/relative dielectric permittivity

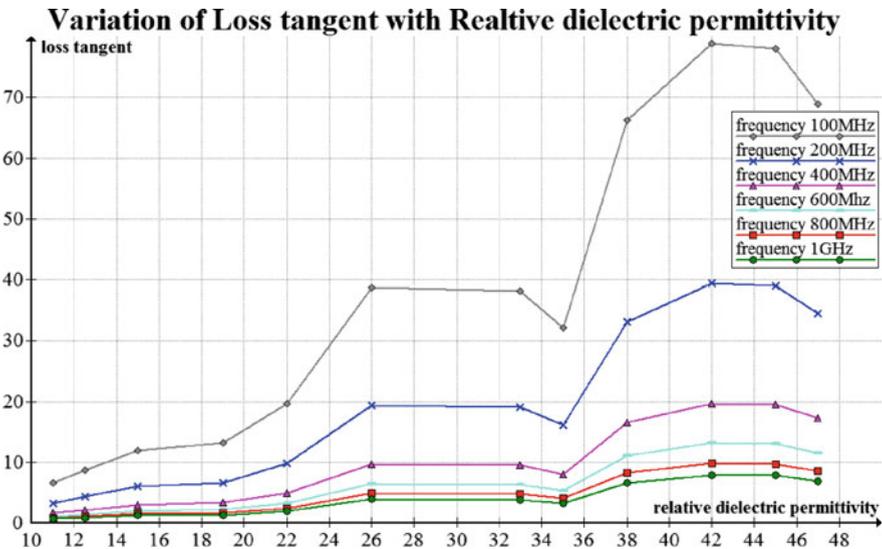


Fig. 3 Loss tangent/relative dielectric permittivity

4 Designing Survey

Single-fold reflection profiling is one of the most common methods of GPR surveying. In this method of profiling, the system is moved along a survey line to plot reflection versus positions. A fixed antenna system is mostly used for this method of profiling. In order to get a higher penetration depth, the user will have to compromise with the resolution of the GPR or vice versa. There will be no practical application if the GPR is able to have a good resolution but unable to detect the target. Range resolution is defined as the half power point of the normalised sine function [21, 22]. It is the ability of radar to resolve between two closely spaced targets. The variation of range resolution with relative dielectric permittivity is given in Fig. 4.

The sampling interval, i.e. the time interval between points on a record waveform is controlled by the Nyquist sampling concept. According to the Nyquist sampling concept, the sampling interval in the record cannot be higher than half the period of the highest frequency signal. The maximum sampling interval for different operating frequencies is given in Fig. 5.

The operating frequency influences the spacing. It is also influenced by dielectric properties of the surface. Nyquist sampling interval is the limit for spacing. Variation of station spacing with dielectric materials operating under different frequencies is given in Fig. 6.

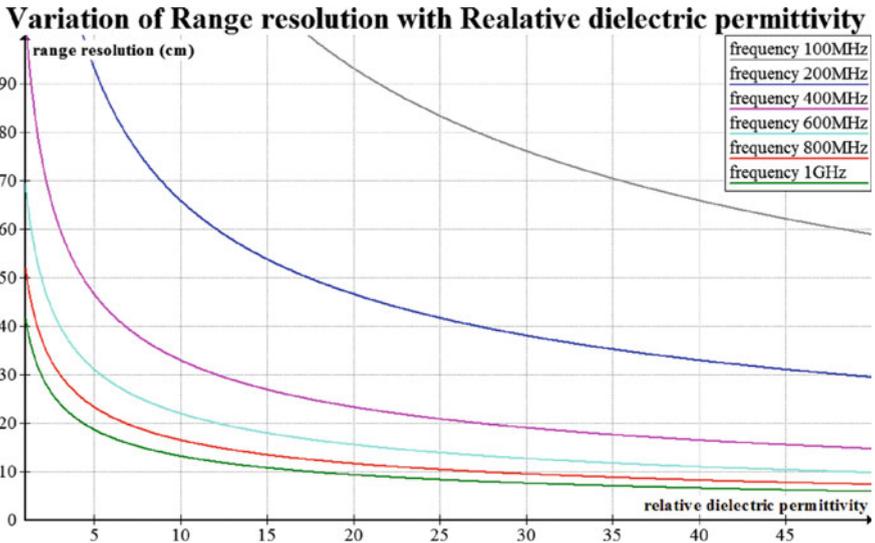


Fig. 4 Range resolution/relative dielectric permittivity

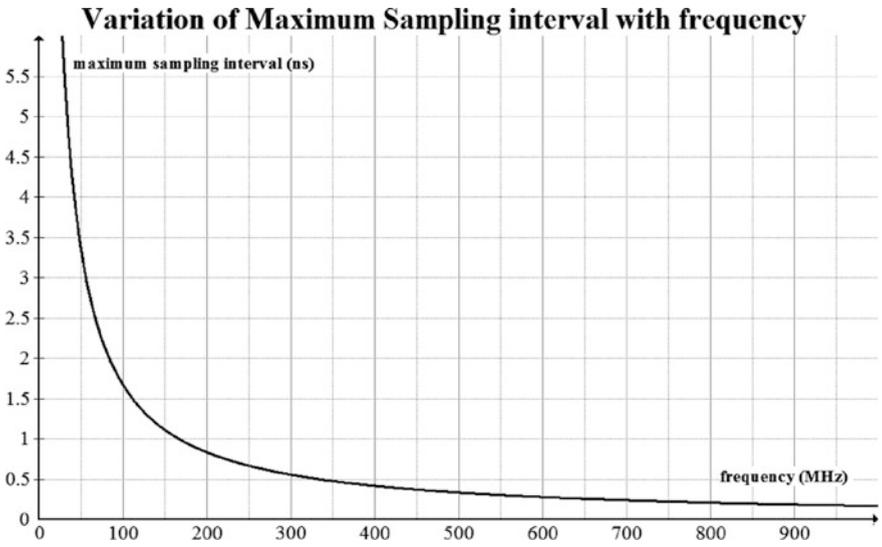


Fig. 5 Maximum sampling interval/frequency

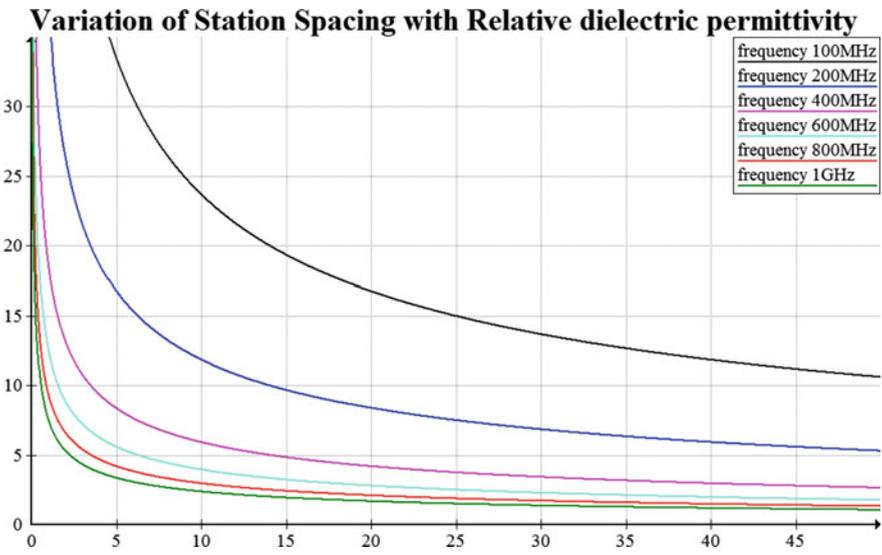


Fig. 6 Station spacing/relative dielectric permittivity

5 Summary/Conclusion

For a low loss medium, propagation velocity is independent of the operating frequency but for conducting mediums, it is important to include conductivity and dielectric permittivity information in GPR processing. With increasing conductivity, the variation of velocity from low loss condition increases.

Figure 2 suggests that skin depth is independent of the operating frequency (overlapping plots). But there is a healthy change in skin depth with increasing dielectric permittivity and conductivity. In general practice, the actual depth of the target should not be greater than 50% of the maximum range. In our study, the maximum skin depth was around 4.5 m. Thus, the maximum the GPR can detect will be around 2 m. GPR hardly have any practical application with dielectric >33 and conductivity >700 $\mu\text{S/m}$. In Fig. 3, the loss tangent almost becomes constant at higher frequencies.

Thus, a frequency range above 200 MHz is good for GPR survey. But there are situations in which GPR cannot be used. Thus, some common sense and a logical thought process are needed to conduct high-quality GPR survey. It is better to define a generic target type before selecting the operating frequency.

References

1. Senson, A.K.: Applications of ground penetrating radar in assessing some geological hazards: example of ground contamination, faults, cavities. *J. Appl. Geophys.* **33**, 177–193 (1995)
2. Bui, D.T., Pradhan, B., Lofman, O., Revhaug, I., Dick, O.B.: Spatial prediction of landslide hazards in Hoa province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena* **96**, 28–40 (2012)
3. Wiegand, C., Kringer, K., Geitner, C., Rutzinger, M.: Regolith structure analysis—a contribution to understanding the local occurrence of shallow landslides (Austrian Tyrol). *Geomorphology* **183**, 5–13 (2013)
4. Beauvais, A., Ritz, M., Parisot, J.-C., Bantsimba, C., Dukhan, M.: Combined ERT and GPR methods for investigating two-stepped lateritic weathering systems. *Geoderma* **119**, 121–132 (2004)
5. Daniels, J.J., Roberts, R., Vendl, M.: Ground penetrating radar for the detection of liquid contaminants. *J. Appl. Geophys.* **33**, 195–207 (1995)
6. Kruse, S.E., Schneider, J.C., Campagna, D.J., Inman, J.A., Hickey, T.D.: Ground penetrating radar imaging of cap rock, caliche and carbonate strata. *J. Appl. Geophys.* **43**, 239–249 (2000)
7. Aranha, P.R.A., Augustin, C.H.R.R., Sobeira, F.G.: The use of GPR for characterizing underground weathered profiles in the sub-humid tropics. *J. Appl. Geophys.* **49**, 195–210 (2002)
8. Diamanti, N., Peter Annan, A.: Characterizing the energy distribution around GPR antenna. *J. Appl. Geophys.* **99**, 83–90 (2013)
9. Jiao, Y., McMechan, G.A., Pettinelli, E.: In situ 2D and 3D measurements of radiation patterns of half-wave dipole GPR antennas. *J. Appl. Geophys.* **43**, 69–89 (2000)
10. Chen, H.-W., Huang, T.-M.: Finite-difference time-domain simulation of GPR data. *J. Appl. Geophys.* **40**, 139–163 (1998)

11. Cassidy, N.J., Millington, T.M.: The application of finite-difference time-domain modelling for the assessment of GPR in magnetically lossy materials. *J. Appl. Geophys.* **57**, 295–308 (2009)
12. Lai, W.L., Kind, T., Wiggenhauser, H.: Using ground penetrating radar and time-frequency analysis to characterize construction materials. *NDT&E Int.* **44**, 111–120 (2011)
13. Mahmoudzadeh, M.R., Frances, A.P., Lubczynski, M., Lambot, S.: Using ground penetrating radar to investigate the water table depth in weathered granites—Sardon case study, Spain. *J. Appl. Geophys.* **79**, 17–25 (2012)
14. Doolittle, J.A., Collins, M.E.: Use of soil information to determine application of ground penetrating radar. *J. Appl. Geophys.* **33**, 101–108 (1995)
15. Reppert, P.M., Dale Morgen, F., Nafi Toksoz, M.: Dielectric constant determination using ground penetrating radar reflection coefficients. *J. Appl. Geophys.* **43**, 189–197 (2000)
16. Von Hippel, A.R.: *Dielectric Materials and Application: Dielectric Materials and Applications*. MIT press, Cambridge (1954)
17. Lazaro-Macilla, O., Gomez Trevino, E.: Ground penetrating radar inversion in 1D: an approach for the estimation of electric conductivity, dielectric permeability. *J. Appl. Phys.* **43**, 199–213 (2000)
18. Stratton, J.A.: *Electromagnetic Theory*, International Series in Pure and Applied Physics, p. 615. McGraw-Hill Book co. (1941)
19. Hatch, M.A., Heinson, G., Munday, T., Thiel, S., Lawrie, K., Clarke, J.D.A., Mill, P.: The importance of including conductivity and dielectric permittivity information when processing low-frequency GPR and high-frequency EMI data sets. *J. Appl. Geophys.* **95**, 75–87 (2013)
20. Tsui, F., Matthews, S.L.: Analytical modelling of the dielectric properties of concrete for subsurface radar applications. *Constr. Build. Mater.* **11**(3), 149–161 (1997)
21. Saarenketo, Timo: Electrical properties of water in clay and silty soils. *J. Appl. Geophys.* **40**, 73–88 (1983)
22. Eaves, J.L., Reedy, E.K.: *Principles of Modern Radar*, pp. 226–228, 409–420. Van Nostrand Reinhold Company, New York (1987)

Investigation of Titanium Silicon Nitride: A Review

**Spandan Guha, Asish Bandyopadhyay, Santanu Das
and Bibhu Prasad Swain**

Abstract Nano-tribological studies are required to develop in-depth understanding on interfacial phenomena of different materials that are used in different industrial (e.g. aerospace industry, coatings, tool hardening electronics devices etc.) applications. In a recent year, tribological studies on transition metals increase because of their promising mechanical properties, hardness, oxidation resistant and wear resistant characteristics. Titanium silicon nitride is considered to be one of the most prominent materials among all the transition metals. Silicon provides excellent resistance to oxidation and titanium guarantees hard coatings. So, combination of two can be developed a useful coating material which can provide wear resistance even at elevated temperature.

Keywords Titanium silicon nitride · CVD · PVD · Microstructure
Mechanical properties

S. Guha (✉)

Department of Mechanical Engineering, Sikkim Manipal Institute of Technology, Majitar,
Sikkim, India

e-mail: spandan.guha@yahoo.com

A. Bandyopadhyay

Department of Mechanical Engineering, Jadavpur University, Kolkata, India

e-mail: asishbanerjee@yahoo.com

S. Das

Department of Mechanical Engineering, Kalyani Government Engineering College, Kalyani,
India

e-mail: sdas.me@gmail.com

B.P. Swain

Centre for Material Science and Nano Technology, Sikkim Manipal Institute of Technology,
Majitar, Sikkim, India

e-mail: bibhuprasad.swain@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication
and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_18

1 Introduction

In past decades, thin films are increasingly used for various industrial, scientific and technological applications. Interests in transition metals have grown considerably because of its resilience, high wear resistant, high strength and high toughness [1–3]. Titanium as a transition metal is considered being one of the hardest, wear resistant and decorative coating materials [1–4]. Among a wide variety of transition metal nitride available in industries, titanium nitride (TiN) is the most common and important material as titanium- and nitrogen-based metal nitride possesses very good commercial interest because of their extreme hardness, wear resistant, corrosion resistant, thermal and electrical properties which fit well with most of the industrial requirements. Besides those advantages, binary metal also shows some of the disadvantages like poor oxidation at elevated temperature, intrinsic brittleness, etc. Those disadvantages restrict binary metal from specific technological applications [1]. This problem has been sorted out by mixing silicon or aluminium into the TiN group of materials to form new family of ternary nitride (e.g. TiSiN, TiAlN), in which for past few years much more progress has shown a remarkable development in the quality of such films [1, 5–7].

1.1 *Titanium Silicon Nitride*

Advances of ternary films nitride have been made in the last decade because of their growing demand in industrial applications. Ternary metal nitride (e.g. TiSiN, TiAlN, etc.) possesses properties like low coefficient of friction, high hardness, moderate wear resistant, corrosion protection and significant melting point, etc., as compared to binary-based coatings. Nowadays most of the cutting tools are coated with TiN binary nitride coating to improve tool's performance. But TiN became oxidized at elevated temperature. Sputtered TiN also associated with two major drawbacks as the microelectronics device size started to shrink it subjected to poor step coverage and columnar polycrystalline microstructure. Deposition of silicon into it indicates clear improvement of mechanical properties in terms of hardness and abrasion resistance as well as it possesses excellent abrasion resistance. Silicon nitride is also considered as one of the best metal nitride material because of its thermal, mechanical and chemical properties but still it cannot be used in some applications because of lack of toughness and reliability. Toughness, density and reliability of silicon nitride can be enhanced by the addition of TiN into it. All over titanium silicon nitride as a ternary nitride films provides better performance compared to binary metal nitride [8–13].

2 Morphology of TiSiN

2.1 Atomic Force Microscopy (AFM)

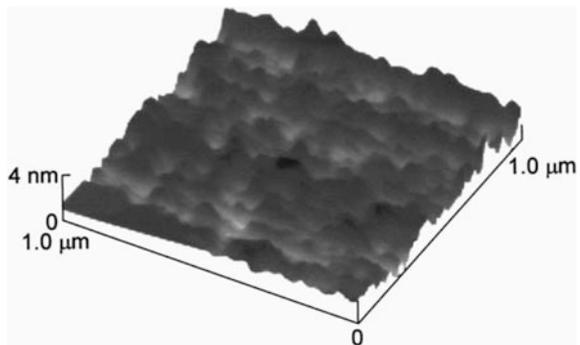
Yokota et al. [14] used ion beam assisted deposition technique to deposit TiN on silicon wafer. Deposition chamber consisted of electron cyclotron resonance ion source to ionize nitrogen. Irregular surface roughness of 4 nm was observed in the AFM micrograph of a single TiN sample. It was found that surface roughness increases when N-ion beam current decreases and it became round in shape with increasing substrate temperature (Fig. 1).

AFM analysis of the sample prepared by Lemus and Drew [15] suggests that the roughness (R_a) of polished titanium (Ti) surfaces (53.4 nm) is slightly less than silicon nitride (Si_3N_4) (63.4 nm) but for unpolished Si_3N_4 and Ti, the value of roughness is 584 and 1140 nm, respectively.

Shen et al. [16] measured roughness of their samples by atomic force microscopy (AFM). They found that addition of Si into TiN causes grain size development and reduce roughness of surface.

Yokota et al. [17] deposited TiN films on Si wafers by using PVD (ion beam assisted deposition) technique. Electron cyclotron resonance is (ECR) used to ionize nitrogen gas in high neutral ratio of nitrogen ion to titanium. Deposition was carried out in N_2 gas flow rate and temperature was maintained at 800 °C. Nitrogen ion particle also used to bombard the TiN film surface. AFM technique used to investigate surface morphology of the films. The probe of AFM scanned the surface area of $10 \times 10 \mu\text{m}^2$ at a rate of 5 Hz/line. They found that bombardment causes breakage of chemical bonds between Ti and N and increased ion current causes variation in deposited TiN.

Fig. 1 AFM image on the surface of the TN-10-rTiN films [15]



2.2 Scanning Electron Microscopy (SEM)

Bellosi et al. [3] studied microstructure development of Si_3N_4 -TiN composites related with sintering process. They used three different grades of Si_3N_4 and TiN powders for sintering process. In first case, silicon nitride powder with sintering aids (3 wt% Al_2O_3 and 8 wt% Y_2O_3) homogenized for 48 h in water and then dried in other case 20–40 vol% was added and homogenized in isobutyl alcohol. Three different densification methods were used for sintering, e.g. hot pressing (HP) in high temperature and pressure at 1800 °C and 30 MPa, respectively, pressure less sintering (PS) in N_2 atmosphere at 1850 °C for 60 min, gas pressure sintering within 1800–1850 °C for cycle A and B respectively for 60 min. They also found from microstructural analysis by scanning electron microscopy that in first case TiN powder of grain sizes 0.2–7 μm were found where as in second case particle size of TiN reduces up to 0.2–3 μm without having any large particles.

Yasutomi et al. [18] tried to get reaction bonded net shaped electroconductive Si_3N_4 -TiN in this experimental work. They used Si and TiN powders with for the production of TiN and Si_3N_4 . At first, they suspended powder mixture into methanol (50 vol%) for a period of 24 h with Si_3N_4 balls. They varied composition of Si/TiN ratio to control TiN in nitride body within 20–70 vol%. Then mixture was dried in room temperature followed by kneading with polyethylene-based thermoplastic resin binder. Then mixture was crushed under 0.5 mm. Then it was dewaxed in argon filled furnace. At last, it passed through the N_2 gas atmosphere at 1350 °C. From SEM analysis, they found that nitride body became more porous as the TiN content increases.

Three different PVD techniques (rf and dc magnetron sputtering, reactive evaporation and ion assistant deposition) were used by Grigorov et al. [19] to deposit titanium nitride films on crystalline (100) silicon of 20 Ω cm substrate. They performed this experiment under crystallization temperature (850–900 °C). Two distinct titanium layer of 60 nm thickness was deposited on silicon substrate in N_2 atmosphere. They specified the layer prepared in presence of Ar ion bombardment as B_0 and without Ar ion bombardment as B_+ . From SEM, they found that a columnar growth of B_0 layer occupied 25–30% of layer's volume and fine-grained structure of B_+ layers with a grain size of 150 nm.

Herrmann et al. [20] in order to find out the dependence of densification behaviour on TiN content and its grain size; they prepared samples of Si_3N_4 (LC 12S, HCST), TiN of various grade or Ti powders to investigate the influence of carbon content in densification. They also used Y_2O_3 powders as sintering additives and prepared two series of samples. First group of materials contained Y_2O_3 (5 wt%) and Al_2O_3 (2 wt%) powder as additives. TiN grade C, Ti(C,N)-30/70 or Ti(C,N)-50/50 is also used as dispersoid after preparing it in isopropanolic solution in a attrition mill. Second group of materials is prepared under the constant ratio of Si_3N_4 to sintering additives (Y_2O_3 , Al_2O_3 : Si_3N_4). In two steps of preparation, first they mixed Si_3N_4 and sintered the additives in attrition mill and in second step they dried it. Ball mill was used to mix TiN powders. Additives were also baked for 8 h at 280 °C in

air and up to 500 °C in Ar gas atmosphere for TiN–Si₃N₄ and TiN–Si(C, N) composites, respectively. They confirmed from SEM analysis that TiN contributes in densification as the grain sizes change during sintering.

Gogotsi et al. [8], in order to find the effect of TiN addition on creep behaviour of hot pressed doped silicon nitride, used mixture of hot pressed silicon nitride and TiN as a sample for study. Uniform distribution of TiN particles was obtained after ball milling of Si₃N₄ along with oxide additives and TiN powder. They measured creep of samples in air within the range of 1100–1340 °C and they could not find any crack in the specimen's surface tested below 1250 °C by SEM analysis.

Chemical stability of TiN on different conditions of silicon nitride was investigated by Huang et al. [9]. The effect of TiN on mechanical properties and microstructure of Si₃N₄ also scrutinizes in this work. To prepare the sample for experiment, they mixed silicon nitride powder with yttria and alumina in a polyurethane bottle. Highly pure silicon nitride balls were mixed with ethanol for periods of 22 h in this bottle. They milled TiN powders of 5 μm size in this mixture with the help of a PE balls in presence of ethanol with a charge to ball ratio of 1:8. After analysed by SEM micrographs, they could not find any interfacial reaction between TiN and Si₃N₄ and up to 1800 °C interfacial regions between TiN and Si₃N₄ remain intact (Fig. 2).

Huang et al. [10] investigated the mechanical properties, microstructure, morphology and fracture behaviour of hot pressed silicon nitride mixed with two different sizes of TiN particles. They prepared their samples as in [9]. Only exception is that they used high purity alumina balls instead of silicon nitride balls and maintained the ratio of ball, charge and vehicle at 6:1:5. Then they hot pressed the sample at 1800, 1850 and 1900 °C. SEM micrographs reveal intactness between the interfaces of TiN and Si₃N₄ even at very high temperature. This proved chemical stability of TiN in Si₃N₄ at 1850 °C. They also found that width of propagation crack increases with decreasing TiN content.

Metal organic chemical vapour deposition technique is used to grow titanium silicon nitride films. In this work, Smith and Custer [11] maintained temperature between 300 and 450 °C and in this temperature tetrakis titanium, ammonia and silane combined together to form films of average compositions near the tie line of TiN–Si₃N₄. They used thermally oxidized bared wafers of Si to deposit the films.

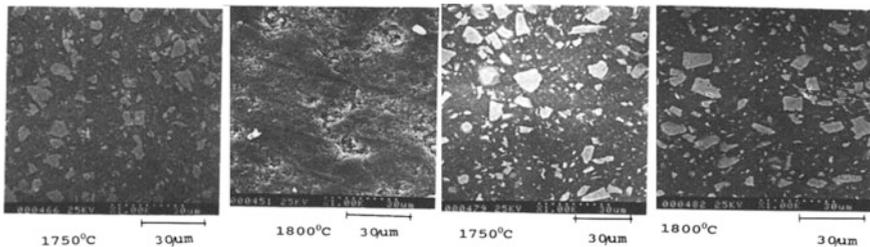


Fig. 2 SEM photograph of microstructure [24]

After analysing the deposited films by SEM, they found that the density of deposited films reduces up to 25% compared to bulk TiN.

In order to compare the R-curve behaviour of Si₃N₄-40 wt% TiN with monolithic Si₃N₄ Choi et al. [21] used indentation strength method. They used Si₃N₄, TiN, Al₂O₃, Y₂O₃ powders as starting materials. For monolithic Si₃N₄, they mixed 97 wt% Si₃N₄, 2 wt% Y₂O₃, 1 wt% Al₂O₃ powders and for Si₃N₄-TiN composites they mixed powders with ratio of 57 wt% Si₃N₄, 40 wt% TiN, 2 wt% Al₂O₃. Then they used chemical routing to prepare the sample. Thin film of gold and moisture free silicon oil deposited on the indentation surface to reduce the moisture-assisted sub critical crack growth. SEM analysis reflects on crack deflection around TiN grains which is bridging by Si₃N₄ grains.

With an objective to improve wear resistance properties of TiN coatings Diserens et al. [13] deposited Titanium silicon nitride films by PVD (reactive unbalanced magnetron sputtering) process. To allow independent regulation of each source, they equipped PVD reactor with titanium and silicon as targets. Cross-sectional morphology of the sample was investigated by SEM analysis. They found that addition of Si into Ti causes transforming of columnar structure into finely grained structure.

Min et al. [22] used metal organic atomic layer deposition technique at 180 °C to grown titanium silicon nitride films. They maintained Si content 18% and Si deposition thickness 0.22 nm/cycles during silane supply. Silane was supplied separately in different pulses (e.g. titanium, ammonia, etc.) within the pressure range of 0.27–13.3 Pa. SEM micrograph of TiSiN films found the step coverage of MOALD even for 0.3 µm films with very less negative slope.

High speed steel is used as a substrate material to deposit TiSiN films by Park et al. [23] with an objective to find the effect of silicon addition on mechanical and microstructural properties of TiN in presence of gaseous mixture (e.g. SiCl₄, TiCl₄, H₂, N₂ and Ar, etc.). During deposition pressure, temperature and rf power were maintained at 1 Torr, 500 °C and 60 W, respectively and deposition time 120 min. By varying the evaporation pressure, they controlled the inlet gas ratio of TiCl₄/SiCl₄ while inlet gas ratio of N₂/H₂/Ar was maintained at 15/40/45. SEM micrographs of undoped TiN film show the presence of columnar grains from substrate to surface whereas for Ti-S-N films a fine microstructure with destruction of columnar grains appeared.

Lemus and Drew [15] studied the different directions of diffusion bonding Ti foil interlayer during joining of Si₃N₄. They used Si₃N₄-rod ceralloy-147 31 N and commercially pure Ti foil as starting materials. The combination of Si₃N₄/Ti foil/Si₃N₄ was hot pressed for different holding times at temperature ranging in between 1200 and 1500 °C. Si₃N₄ which is used in hot pressing is β-Si₃N₄ contained of Y₂O₃ and Al₂O₃ as additives. From SEM analysis, they could not find any bonding between Ti and Si₃N₄ up to 1400 °C.

Blugan et al. [24] investigated the effect Ti addition on fracture toughness, strength and wear properties of commercial Si₃N₄. In this work, they varied the TiN wt% (10, 20, and 30) to study the effect on fracture toughness. Alumina (Al₂O₃) and yttria (Y₂O₃) were used as a doping material to dope Si₃N₄ and Si₃N₄-TiN

composite materials. They investigated the microstructure of plasma etched specimen by SEM process. They observed that the average grain size of β - Si_3N_4 increases with increased TiN content up to 20 wt%. But average grain size of TiN maximum for 10 wt% of TiN was observed. They also found that Si_3N_4 can retain their sub micrometre grain size with addition of TiN.

Lee and Lim [25] used ion arc plating technique to deposit TiN films on silicon wafer at 450 °C. Before deposition, they sputtered the surface of substrate using Ar ion at -700 V bias voltages for duration of 10 min. Base and working pressure was maintained at 1×10^{-5} and 3×10^{-3} Torr, respectively. Time of deposition was 20 min. Arc and sputter current was 50 and 1 A, respectively. They also maintained ratio of N_2 gas to Ar gas at 0.3. Produced particles from laser ablation are collected on carbon tape to investigate by SEM process. They observed from FESEM and TEM that size of the particles for first 5 laser shots was within the range of 70–90 nm. Size of particle increases up to 50–100 nm as the laser shots increase. Si substrate was completely exposed by TiN layer after 25–30 laser shots. Silicon particle completely appeared from 30 to 50 laser shots.

Commercially available TiN powders and amorphous silicon nitride (a- Si_3N_4) of 1.7–8.1 mol% had been sintered at 6.5 GPa to prepare dense nano composite. In this experiment, Blab et al. [26] maintained temperature at 1300 °C during high pressure sintering. They also processed the powders in dry N_2 gas because of their high affinity in oxygen. SEM images of the samples reveal the presence of various clusters of pores inside thick matrix at 1300 °C and it decreases with increasing content of Si_3N_4 . But there is a presence of cluster pores sporadically with maximum silicon nitride content of 8.1 mol%.

3 Structure of TiSiN

3.1 X-ray Diffraction (XRD)

Belloso et al. [3] sintered Si_3N_4 -TiN composites to study microstructural development of it. They used three different grades of Si_3N_4 powder (e.g. LC 12, Starck, FRG) and TiN powder (grade C, Starck, FRG). Three different densification techniques are used by them for sintering. From XRD analysis, they came to know that hot pressing sintering and gas pressure sintering produce very dense structure but pressure less sintering produces porous structure. They also find out a complete conversion of α to β Si_3N_4 in HP samples only.

Yasutomi et al. [18] mixed TiN and Si powder to produce net-shaped reaction bonded Si_3N_4 -TiN. They carried out different steps to produce Si_3N_4 -TiN. They found out from X-ray diffractometry that the reaction between metallic Si and nitrogen formed TiN, α and β - Si_3N_4 in nitride bodies.

Grigorov et al. [19] tried to prevent the silicon diffusion in TiN films of columnar [B_0] and fine-grained structure [B_+]. They deposited the titanium nitride films on

silicon substrate by different methodologies. Presence of δ -TiN, α -TiN in B_o layers and δ -TiN, TiN in B_+ layers they found different phase of titanium nitride (TiN) from XRD analysis. Crystalline sizes of B_o layers for pronounced TiN (200) and peaks Ti (011) are 25 and 20 nm, respectively, also they were found from it.

Huang et al. [9] investigated the chemical stability of TiN in Si_3N_4 matrix in different temperature and gaseous environment conditions. They also studied the effect of titanium nitride (TiN) on mechanical properties and microstructure of silicon nitride (Si_3N_4). From XRD analysis, they find that α phase of Si_3N_4 converted into β phase at the temperature of 1800 °C but some α - Si_3N_4 returns back when temperature decreased to 1750 °C.

Huang et al. [10] could not find any phases except Si_3N_4 and TiN from XRD profile of monolithic Si_3N_4 and 20 vol% TiN-containing Si_3N_4 composites up to a temperature of 1900 °C.

Diserens et al. [13] deposited TiSiN films with the help of PVD techniques in quest to improve wear resistance properties of TiN by adding Si into it. PVD reactor equipped with titanium and silicon targets to adjust the Ti:Si ratio in the gas phase. XRD analysis shows that the columnar structure (111) transforms into dense finely grained structure when Si was added to TiN coatings. Park et al. [23] found from XRD analysis that the orientation of undoped TiN film is very strong (200) whereas TiSiN films has some mixed orientation of (111), (220) and (311). Intensity of peak (200) will decrease with increasing gas mixture.

Vaz et al. [27] to investigate and characterize the properties of (Ti,Si)N systems, (Ti,Si)N films deposited on high speed steel and silicon wafers substrate by r.f. reactive magnetron sputtering technique. They carried out deposition in Ar and N_2 gas atmosphere. All the substrates were sputter etched in pure argon gas atmosphere for 15 min in 200 W r.f. powers. The base pressure of deposition chamber was 10^{-4} Pa later raised to 4×10^{-1} Pa during deposition. XRD patterns of the film reveal that both the structures were indexed to cubic structure.

Yokota et al. [14] deposited TiN on alcohol, acetone and solvent naphtha rinsed silicon wafer by ion beam assisted deposition technique. To ionize N, they used an electron cyclotron ion source. Deposition pressure and ratio of N ion beam current to N_2 flow rate were maintained at 10^{-2} Pa and 16.5 mA/20 sccm, respectively. From XRD analysis, strongest and weakest peaks are found at (200) and (111), (220) lattice planes, respectively, for cubic TiN films. They also came to know that cubic TiN increased significantly with increasing temperature when it diffracted from (200) lattice plane.

At the interface of Si_3N_4 /Ti foil/ Si_3N_4 specimen Lemus and Drew [15] found the presence of Ti_5S_3 , TiSi and TiN from XRD analysis. The interface of polished samples was grown thicker parabolically as compared to those samples considered as ground sample.

Shen et al. [16] used reactive unbalanced dc magnetron sputtering technique to grow $Ti_{1-x-y}Si_xN_y$ film on unheated Si (100) substrate. They carried out their deposition in presence of Ar and N_2 gas mixture. Si substrates were ultrasonically cleaned and dried in acetone–methanol mixture and N_2 gas, respectively, before entered into the deposition chamber. The base pressure and working pressure set at 2×10^{-6} Torr and 2 mTorr, respectively. To get homogeneous composition and thickness, they rotated the substrate at a speed of 35 rpm during deposition. They used XRD to determine crystallographic structure of the film. From XRD θ -2 θ scan, they observed a single peak for TiN at (111) and this peak broadened with increasing Si content. They found average grain size of pure TiN 50 nm by using XRD peaks and Scherrer equation.

Blugan et al. [24] investigated the crystalline phase of their polished samples with XRD process with $CuK\alpha_1$ radiation. XRD confirmed presence of βSi_3N_4 and TiN as main crystalline phase as there are no Ti-based crystalline phase has been detected.

Blab et al. [26] prepared dense nanocomposite of TiN and 1.7–8.1 mol% silicon nitride by sintering process. They maintained pressure and temperature during sintering at 6.5 GPa and 1300 °C, respectively. After investigating the crystalline structure of samples by XRD, they found that TiN was present as a crystalline form in all specimens. They also observed significant reduction in lattice parameter of titanium nitride.

4 Compositional and Electronic Structure

4.1 X-ray Photoelectron Spectroscopy (XPS)

Diserens et al. [13] tested chemical nature of their samples by XPS technique and presence of silicon, titanium and nitrogen they found in it. Even they also discovered that partial pressure of nitrogen plays the high influential role for nitridation of silicon. They also came to know that, as the partial pressure of nitrogen is increasing it causes changes in chemical state of silicon.

Park et al. [23] were performed XPS operation to verify the behaviour of Si on TiSiN films consist of 7% Si. They found that peak binding energy of Si 2p and Ti 2p were at 455.6 and 101.8 eV, respectively.

Shen et al. [16] produced $Ti_{1-x-y}Si_xN_y$ thin films on unheated Si (100) substrate by PVD (reactive unbalanced dc magnetron sputtering) method in presence of Ar– N_2 gas mixture. The electron bonding configuration of film was characterized by XPS technique. They observed opposite trend between Si and Ti concentration growth with current. Concentration of Si increased with increasing current up to 2 A whereas Ti concentration decreased.

5 Mechanical Properties

5.1 Nano Indentation

Shen et al. [16] used nanoindentation to measure hardness of their samples. They observed significant growth in hardness for TiN films with Si addition. Initial hardness of TiN found 18.8 GPa and it increases up to 32.8 GPa with increasing Si concentration.

6 Conclusion

Some of the nitride-based ternary thin films (e.g. TiSiN, TiAlN, TiMoN) have huge industrial application because of their properties like significant hardness, less coefficient of friction, good wear resistant, corrosion protection and significantly high melting point, etc., as compared to binary-based nitride coatings. The hardness of TiSiN film was found to be 32.8 GPa [16] when deposited by reactive unbalanced dc magnetron sputtering technique in presence of Ar–N₂ gas mixture. From AFM analysis, minimum surface roughness (R_a) of titanium silicon nitride thin film deposited by ion beam assisted deposition technique was found to be 4 nm [17]. Young modulus of dense nano composite of titanium nitride and silicon nitride increased with decreasing temperature [26].

References

1. Wang, J., Zhou, Y.: Recent progress in theoretical prediction preparation and characterization of layered ternary metal carbides. *Annl. Rev. Mater. Res.* **39**, 415–443 (2009)
2. Chauhan, K.V., Rawal, S.K.: A review paper on tribological and mechanical properties of ternary nitride based coatings. *Proc. Technol.* **14**, 430–437 (2014)
3. Bellosi, A., Guicciardi, S., Tampieri, A.: Development and characterization of electro conductive Si₃N₄–TiN composites. *J. Eur. Ceram. Soc.* **9**, 83–93 (1992)
4. Gago, R., Soldera, F., Hubner, R., Lehmann, J., Munnik, F., Vazquez, L., Redondo-Cubero, A., Endrino, J.L.: X-ray absorption near-edge structure of hexagonal ternary phases in sputter-deposited TiAlN films. *J. Alloy. Comp.* **561**, 87–94 (2013)
5. Satou, M., Andoh, Y., Ogata, K., Suzuki, Y., Matsuda, K., Fujimoto, F.: Coating films of titanium nitride prepared by ion and vapour deposition method. *Jpn. J. Appl. Phys.* **24**, 656–660 (1985)
6. Kiuchi, M., Tomita, M., Fujii, K., Satou, M., Shimizu, R.: Titanium nitride crystal growth with preferred orientation by dynamic mixing method. *Jpn. J. Appl. Phys.* **26**, L938–L940 (1987)
7. Durand-Drouhin, O., Santana, A.E., Karimia, A., Derflinger, V.H., Schutze, A.: Mechanical properties and failure modes of TiAl(Si)N single and multilayer thin films. *Surf. Coat. Technol.* **235**, 260–266 (2003)

8. Gogotsi, Y.G., Grathwohl, G.: Creep of silicon nitride-titanium nitride composites. *J. Mater. Sci.* **28**, 4279–4287 (1993)
9. Huang, J.L., Chen, S.Y., Lee, M.T.: Microstructure chemical aspects and mechanical properties of $\text{TiB}_2/\text{Si}_3\text{N}_4$ and $\text{TiN}/\text{Si}_3\text{N}_4$ composites. *J. Mater. Res. Soc.* **9**, 2349–2354 (1994)
10. Huang, J.L., Lee, M.T., Lu, H.H., Li, D.F.: Microstructure fracture behaviour and mechanical properties of $\text{TiN}/\text{Si}_3\text{N}_4$ composites. *Mater. Chem. Phys.* **45**, 203–210 (1995)
11. Smith, P.M., Custer, J.S.: Chemical vapor deposition of titanium silicon nitride films. *Appl. Phys. Lett.* **70**, 3115–3118 (1997)
12. Loye, H.C., Houmes, J.D., Bem, D.S.: Recent developments in ternary nitride chemistry, *The Chemistry of Transition Metal Carbides and Nitrides*, pp. 154–174. Chapman & Hall (1996)
13. Diserens, M., Patscheider, J., Levy, F.: Improving the properties of titanium nitride by incorporation of silicon. *Surf. Coat. Technol.* **108–109**, 241–246 (1998)
14. Yokota, K., Nakamura, K., Kasuya, T., Mukai, K., Ohnishi, M.: Deposition of titanium nitride films onto silicon by an ion beam assisted deposition technique. *Thin Solid Films* **473**, 340–345 (2005)
15. Lemus, J.S., Drew, R.A.L.: Joining of silicon nitride with titanium foil interlayer. *Mater. Sci. Eng. A* **352**, 169–178 (2003)
16. Shen, Y.G., Ziu, Z.J., Jiang, N., Zhang, H.S., Chan, K.H.: Effect of silicon addition on surface morphology and structural properties of titanium nitride films grown by reactive unbalanced direct current magnetron sputtering. *J. Mater. Res.* **19**, 523–534 (2004)
17. Yokota, K., Nakamura, K., Kasuya, T., Ohnishi, M.: Phase composition and crystalline structure of titanium nitride deposited on silicon by an ion beam assisted deposition technique, *Nuc. Instrum. Meth. Phys. Res.* **206**, 386–389 (2003)
18. Yasutomi, Y., Chiba, A., Sobue, M.: Development of reaction bonded electro conductive silicon nitride titanium nitride and resistive silicon nitride aluminium oxide composites. *J. Am. Ceram. Soc.* **74**, 950–957 (1991)
19. Grigorov, K.G., Grigorov, G.I., Stoyanova, M., Vignes, J.L., Langeron, J.P., Denjean, P., Perriere, J.: Diffusion of silicon in titanium nitride films, efficiency of TiN barrier layers. *Appl. Phys. A* **55**, 502–504 (1992)
20. Herrmann, M., Balzer, B., Schubert, C., Hermel, W.: Densification, microstructure and properties of $\text{Si}_3\text{N}_4\text{-Ti(C, N)}$ composites. *J. Eur. Ceram. Soc.* **12**, 287–296 (1993)
21. Choi, H.J., Cho, K.S., Lee, J.G., Kim, Y.W.: R curve behaviour of silicon nitride-titanium nitride composites. *J. Am. Ceram. Soc.* **80**, 2681–2684 (1997)
22. Min, J.S., Park, H.S., Kang, S.W.: Metal organic atomic layer deposition of titanium silicon nitride films. *Appl. Phys. Lett.* **75**, 1521–1524 (1999)
23. Park, B.H., Kim, Y.I., Kim, K.H.: Effect of silicon addition on microstructure and mechanical property of titanium nitride film prepared plasma assisted chemical vapor deposition. *Thin Solid Films* **348**, 210–214 (1999)
24. Blugan, G., Hadad, M., Rusch, J.J., Kuebler, J., Graule, T.: Fractography, mechanical properties and microstructure of commercial silicon nitride-titanium nitride composites. *J. Am. Ceram. Soc.* **88**, 926–933 (2005)
25. Lee, J.S., Lim, H.B.: Laser ablation of titanium nitride coated on silicon wafer substrate for depth profiling using ICP-MS. *Appl. Surf. Sci.* **327**, 483–489 (2015)
26. Blab, U.W., Barsukova, T., Schwarz, M.R., Kohler, A., Schimpf, C., Petruscha, I.A., Muhle, U., Rafaja, D., Kroke, E.: Bulk titanium nitride ceramics-significant enhancement of hardness by silicon nitride addition nano structuring and high pressure sintering. *J. Eur. Ceram. Soc.* **35**, 2733–2744 (2015)
27. Vaz, F., Rebouta, L., Silva, R.M.C.D., Silva, M.F.D., Soares, J.C.: Characterization of titanium silicon nitride films deposited by PVD. *Vacuum* **52**, 209–214 (1999)

An Efficient MOS Class-E Power Amplifier for Wireless LAN Applications

M. Ananda and A.B. Kalpana

Abstract In the present scenario, use of wireless communication devices is increasing drastically for communication. Because of huge wireless communication, there are chances for loss of signal strength. Power amplifier helps in increasing the strength of weak RF signal of transmitter of wireless communication. There is a demand for high quality, low cost and low power consumption power amplifier for wireless devices. This project proposes a design of an efficient MOS Class-E power amplifier at low power consumption either by reducing the input DC power level or by scaling the MOS device using 180 nm technology for wireless applications such as Bluetooth and Wi-Fi application.

Keywords RF · Class-E · Power amplifier (PA) · PAE · BFC

1 Introduction

The power amplifier is used at the transmitter chain in transceiver of communication system and helps in increasing the strength of weak RF signal before applying to antenna. The signal to PA comes from the processor and is used for message broadcasting, the power amplifier consumes large amount of power in transmitter chain and hence is called as power hungry block.

Power amplifier mainly consists of three blocks [1]: the input matching network, amplifying circuit and the output matching network. From the block diagram of power amplifier shown in Fig. 1, the input matching makes use of inductance L connected between drain of MOS and supply and is called drain inductance which is varied according to resonance frequency of 2.4 GHz. If matching is perfect then

M. Ananda (✉)

Visvesvaraya Technological University, Belagavi, Karnataka, India
e-mail: anandam@pes.edu

A.B. Kalpana

ECE Research Center, BIT, Bangalore, India
e-mail: abkalpana@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_19

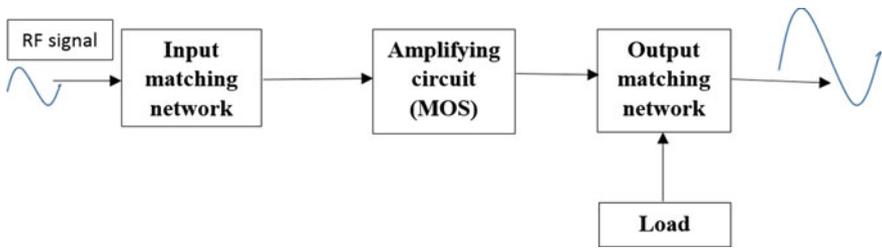


Fig. 1 Block diagram of power amplifier

power should not be reflected back to the input side transmitted from the source. The output matching is responsible for output power of the circuit. A LC tank circuit with a series–parallel connection of inductor and capacitor together is used for output matching network. The LC tank circuit is used for tuning the output part of the circuit to operate at the frequency of 2.4 GHz. The value of the matching elements depends on the size of the transistor used for amplification.

2 Design Methodology

Some of the topologies of the proposed MOS Class-E power amplifier design are:

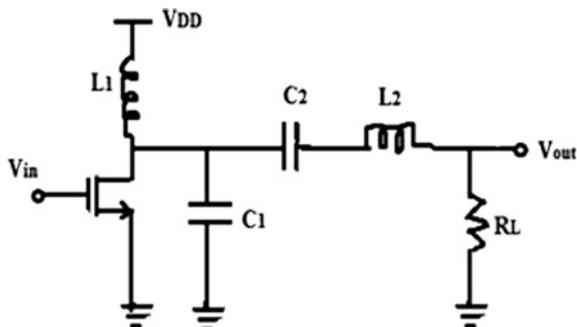
2.1 Basic Configuration of MOS Class-E Power Amplifier

Class-E power amplifier is designed by choosing the appropriate value of inductor L and capacitor C for the matching network. For Class-E design, biasing of transistor is important and is achieved such that, before current starts flowing through the Transistor, voltage goes to zero. It means there is no overlap between voltage and current. To do this, we are using higher order reactance and hence we use very large value of inductor and capacitor.

As depicted in Fig. 2, the following points are to be studied:

- Large Inductor (BFC), L_1 : It helps in constant current flow through transistor. BFL acts as current source and advantage of using BFT is it biases the output to V_{dd} .
- Large Capacitor (BFC), C_1 : Use of BFC avoids Dc current flowing through the load R_L .
- LC Resonant Circuit: At resonance, LC circuit has high impedance. So some current flows through the load and remaining current is bypassed.

Fig. 2 Class-E power amplifier basic configuration



- Load Resistance R : The value of load is 50Ω . Since antenna is pure resistor which has 50Ω . We can change this value by using impedance matching on what power we have to deliver.
- Advantages of using BFC and BFL:
 1. The transistor output capacitance can be absorbed into the tank circuit.
 2. These help in filtering which reduce the frequency on out of band.

3 Design Procedure

3.1 Calculation of Aspect Ratio (W/L) of nMOS

The standard values used for 180 nm technology are threshold voltage (V_{th}) is of 0.7 V and gate-to-source voltage (V_{gs}) of MOSFET is in the range of 5–7% of V_{dd} .

If we take $V_{dd} = 1.8$ V then the minimum and maximum value of V_{gs} is as follows:

$V_{gs(\min)} = 5\%$ of $V_{dd} = 0.09$ V and $V_{gs(\max)} = 7\%$ of $V_{dd} = 0.126$ V. Therefore V_{gs} value will be in between 0.09 and 0.126 V.

For design purpose length (L) of MOSFET is fixed to 180 nm and width (W) of MOSFET is going to variable.

The drain current equation for nMOS, which is in saturation region, is given by:

$$I_D = 1/2 \mu_n C_{ox} (W/L) (V_{gs} - V_{th})^2 \quad (1)$$

Maximum output power is calculated as:

$$P_{out(\max)} = 0.577 V_{dd}^2 / R \quad (2)$$

Output power is given by

$$P_{\text{out}} = V_{\text{dd}} I_{\text{DC}} \quad (3)$$

$$I_{\text{DC}} = P_{\text{out}}/V_{\text{dd}} = 17.78 \text{ mA.}$$

But $C_{\text{ox}} = 8.840 \text{ fF}/(\mu\text{m})^2$ and for nMOS $\mu\text{m} = 8 * 10^{-3}$.

Now by substituting all above values in Eq. (1), we get $W = 276.42 \mu\text{m}$ which is approximately equal to $W = 360 \mu\text{m}$.

And $W/L = 360 \mu\text{m}/180 \mu\text{m}$ which is depicted in Fig. 3.

3.2 Calculation of LC Values for Output Matching Network

We know that $f = 2.4 \text{ GHz}$, $V_{\text{dd}} = 1.8 \text{ V}$ and $P_{\text{out}} = 32 \text{ MW}$.

Therefore, $\omega = 2\pi f = 15.08 \text{ MHz}$

$$L_2 = \frac{\pi V_{\text{dd}}^2}{2\omega P_{\text{out}}} \frac{\pi^2 - 4}{\pi^2 + 4} \quad (4)$$

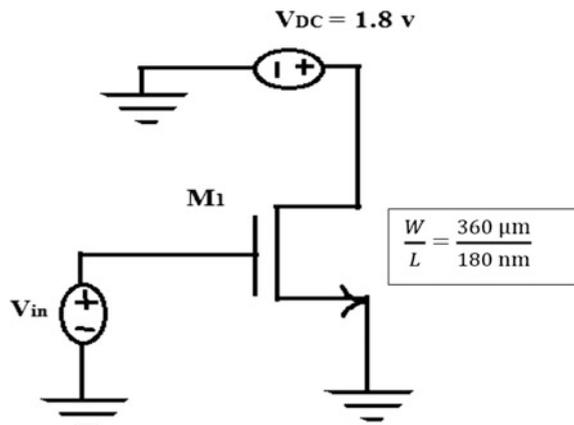
$$\omega = \frac{1}{\sqrt{L_2 C_2}} \quad (5)$$

$$\sqrt{L_2 C_2} = \frac{1}{\omega} \quad (6)$$

Then

$$C_1 = \frac{P_{\text{out}}}{\pi\omega V_{\text{dd}}} \quad (7)$$

Fig. 3 NMOS aspect ratio



The load resistor RL is 50Ω which matches the antenna impedance. The proposed output matching network is depicted in Fig. 4.

3.3 Calculation for Input Matching Circuit

For a design of perfect input matching, an inductor is connected between drain and supply. The inductance L_1 value at drain of MOS is varied in accordance with tuned resonance frequency of 2.4 GHz. The reactance is larger than ten times the reactance of C_1 (Fig. 5)

$$X_{L1} > 10X_{C1}$$

Therefore,

$$2\omega L_1 = 10/2\omega C_1$$

If $C_1 = 29 \text{ pF}$, then $L_1 = 4.9 \text{ nH}$.

Fig. 4 Output matching network of the proposed design

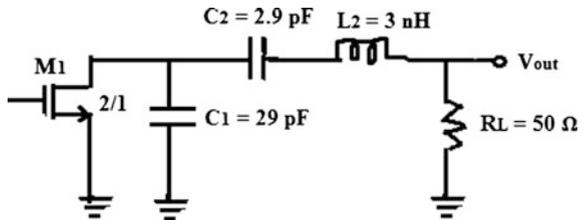
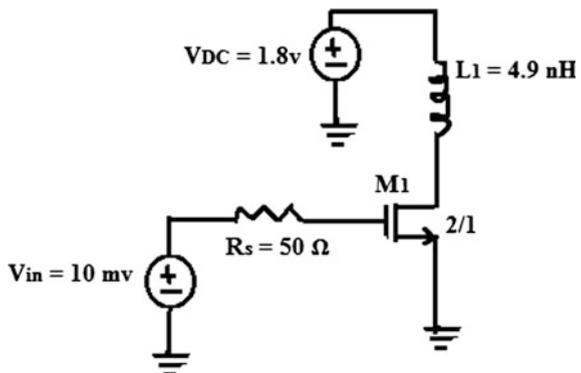


Fig. 5 Input matching network of proposed design



4 Proposed Design of MOS Class-E Power Amplifier

The proposed design of MOS Class-E power amplifier is depicted in Fig. 6.

5 Simulation Results and Performance Analysis

The test setup is shown in Fig. 7, and simulated using 180 nm RF CMOS process and results are validated on Cadence IC 613 [2].

The transient response of designed MOS Class-E power amplifier is shown in Fig. 8. The gain can be calculated as $\text{Gain} = V_{\text{out}}/V_{\text{in}} = 1 \text{ V}/1 \text{ mV} = 20 \text{ dB}$ (Fig. 9).

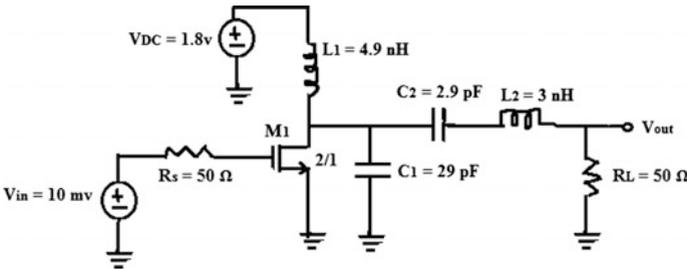


Fig. 6 Proposed Class-E power amplifier

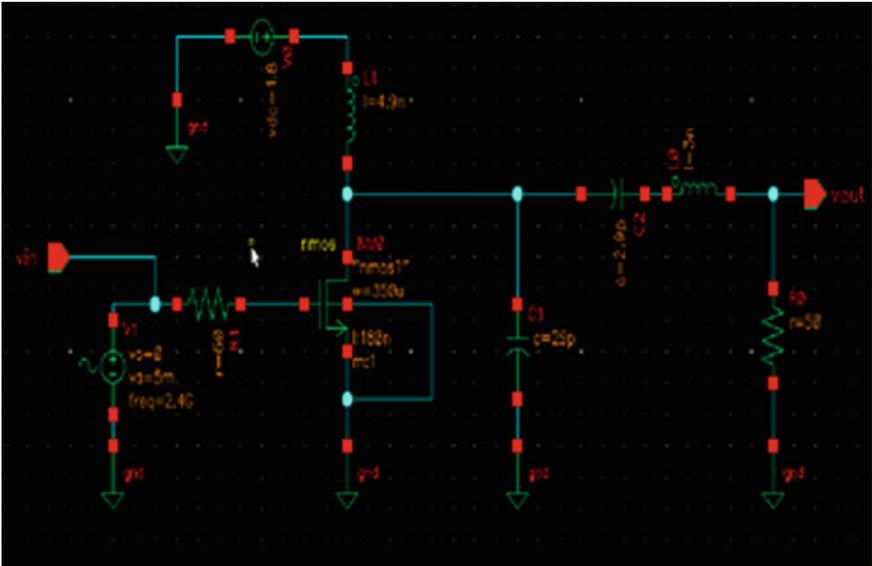


Fig. 7 Test setup of MOS Class-E power amplifier

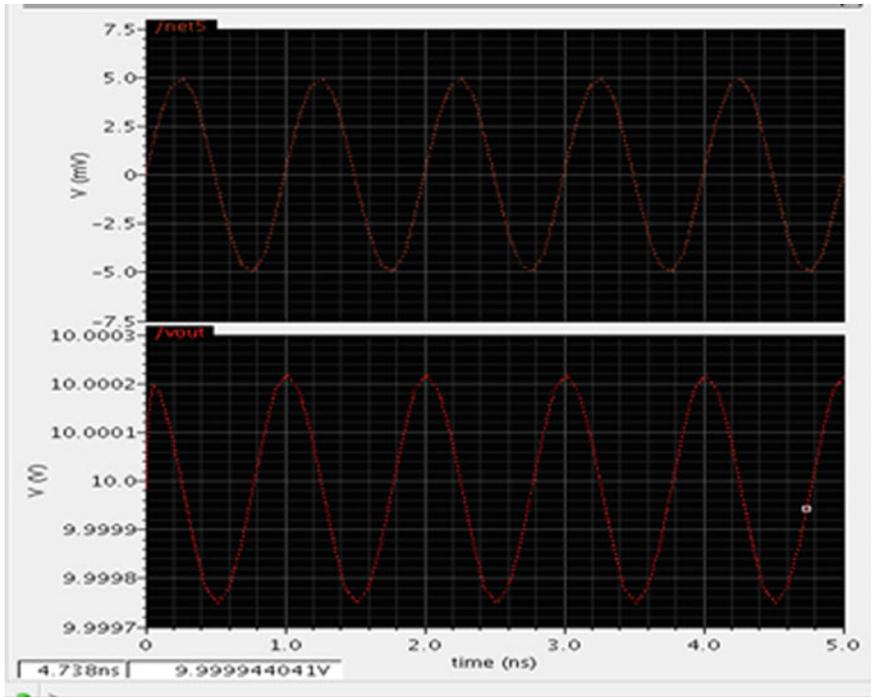


Fig. 8 Transient response of the MOS Class-E PA

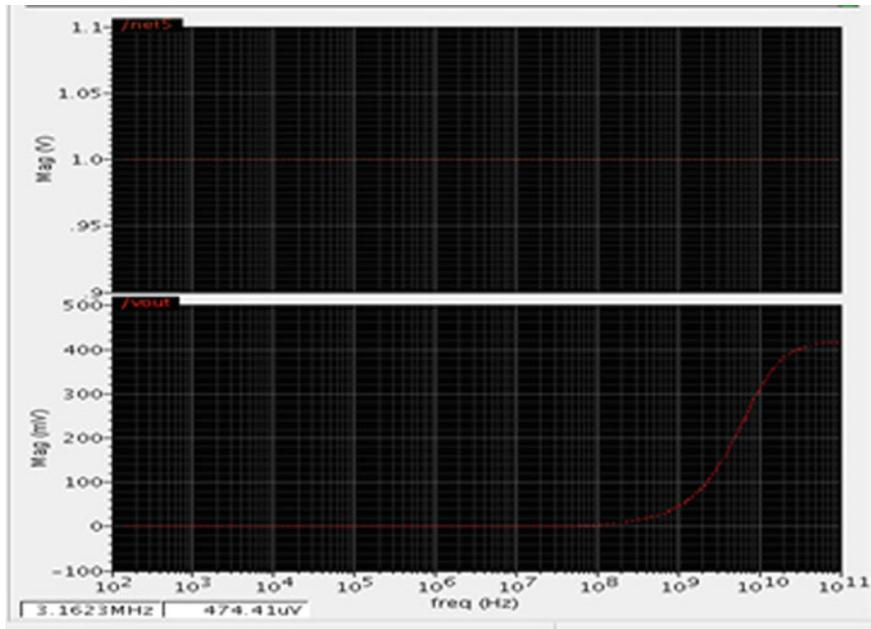


Fig. 9 Gain versus frequency of MOS Class-E PA

From Fig. 10, the output power is calculated is,
 $P_{out} = 10 \log_{10} (31.399 \text{ MW}) = 14.96 \text{ dBm}$, which is approximately equal to 15 dBm (Table 1).

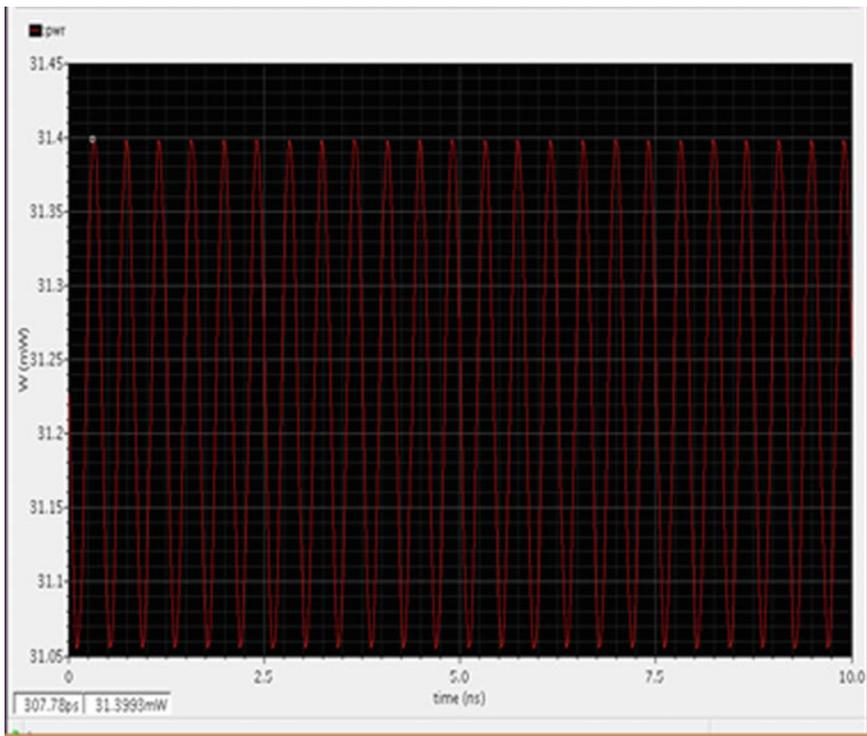


Fig. 10 Output power waveform of MOS Class-E PA

Table 1 Comparison summary

Year	2007	2009	2010	2012	2014	2015–16
References	[3]	[4]	[5]	[6]	[7]	This work
Process technology (nm)	180	180	180	180	180	180
Supply voltage (V)	1.8	1.8	3.3	1.8	1.8	1.8
Frequency (GHz)	1.9	2.4	1.8	2.4	2.4	2.4
Output power (dBm)	2	20	3	20	5.8	15
Efficiency (%)	40	62	31	44.6	55	66.11

6 Conclusion

This paper is designed for efficient power amplifier to get output power of 15 dBm (32 MW) and efficiency of about 70% from input sine wave of amplitude 1 mV, at frequency of 2.4 GHz. But obtained result of output power is 14.96 dBm (31.39 MW) and efficiency is about 66.11% and possible applications of the paper is in short distance communication like Bluetooth and Wi-Fi.

References

1. Razavi, B.: RF Microelectronics, 2nd edn. Prentice Hall Inc. (1998)
2. <http://www.cadence.com/>. Accessed 10 Oct 2015
3. Park, et al.: A 1.9-GHz CMOS power amplifier using three-port asymmetric transmission line transformer for a polar transmitter. *IEEE Trans. Microwave Theory Technol.* **55**(2), 230–238 (2007)
4. Lie, D.Y.C., Lopez, J., Popp, J.D., Rowland, J.F., Wang, G., Qin, G., Ma, Z.: Highly efficient monolithic class-E SiGe power amplifier design at 900 and 2400 MHz. *IEEE Trans. Circuits Syst.-I Regul. Pap.* **56**(7), 1455–1466 (2009)
5. Lee, O., Hwan, K., Kim, H., Lee, D.H., Han, J., Yang, K.S., Lee, C.-H., Kim, H., Laskar, J.: An analysis and design of fully integrated high-power parallel-circuit class-E CMOS power amplifiers. *IEEE Trans. Circuits Syst.-I Regul. Pap.* **57**(3), 724–734 (2010)
6. Li, Z., Torfs, G., Bauwelinck, J.T., Yin, X.: A 2.45 GHz + 20 dBm fast switching Class-E power amplifier with 43% PAE and a 18 dB wide power range in 0.18 μ m CMOS. *IEEE Trans. Circuits Syst.-II* **59**(4), 224–228 (2012)
7. Ture, K., Kilinc, E.G., Dehollain, C.: A fully integrated low-power Class E power amplifier for short range communication. In: *IEEE Proceedings of Mediterranean Microwave Symposium*, pp. 1–5 (2014)
8. Sokal, N., Sokal, A.: Class E—A new class of high-efficiency, tuned single-ended switching power amplifier. *IEEE J. Solid-State Circuits* **10**(3), 168–176 (1975)
9. Sowlati, T., Leenaert, D.: A 2.4 GHz 0.18- μ m CMOS self-biased cascade power amplifier. *IEEE J. Solid-state Circuits* **40**(2), 1318–1324 (2003)
10. Razavi, B.: *Design of Analog CMOS Integrated Circuits*. Tata McGraw-Hill (2002)
11. Ding, Y., Harjani, R.: A high-efficiency CMOS +22-dBm linear power amplifier. *IEEE J. Solid-state Circuits* **40**(2), 1875–1900 (2005)
12. Lee, T.H.: *The Design of CMOS Radio Frequency Integrated Circuits*. Cambridge University Press, New York (2004)
13. Razavi, B.: RF transmitter architectures and circuits. In: *IEEE Custom Integrated Circuits Conference* (1999)
14. <http://www.ncjrs.gov/pdf/nij/191160.pdf>. Accessed 15 Nov 2015
15. http://www.cadence.com/rl/resource/datasheets/virtuoso_vse_fam_ds.pdf. Accessed 20 Oct 2015

Security in Cloud Computing-Based Mobile Commerce

Ferdousi Khatun and Md. Ruhul Islam

Abstract Cloud computing is a burgeoning technology shift impact in gall industries. Due to rapid development in Internet and wireless technology, electronic business transaction has exploded around the world. In brief Mobile commerce or M-commerce means accessing the Internet for business transaction using handheld devices. Cloud integrated mobile application provides the functions for building and consuming the future of M-commerce applications. Nowa ays explosive growth has been taken place in M-commerce due to drastic use of handheld devices like mobile, smartphone, tablet, etc., as it provides user mobility, portability. But some important challenge faced by M-commerce is security, limited storage and processing power, complex device configuration, low network band width, etc. By incorporating cloud computing in mobile commerce will give lots of advantage for business industry as well as the user with lower cost and high benefits, security, flexibility, scalability, mobility, and efficiency.

Keywords M-commerce · Cloud computing · Mobile cloud computing

1 Introduction

In current era of globalization various handheld devices (smart phones, tablet PC, iPad, etc.) becoming an essential part of human being as it provides the easy and efficient way of communication and anytime, anywhere, anyhow accessibility [1, 2]. Few years ago mobile devices was only used for phone calls, massaging of text and pictures. But various services are available due to advances of wireless technology. Mobile devices or handheld devices like smartphones, tablets, etc., are easy

F. Khatun (✉) · Md. Ruhul Islam (✉)

Computer Science and Engineering Department, Sikkim Manipal Institute of Technology, Majhitar, Sikkim, India
e-mail: ferdousi9@yahoo.com

Md. Ruhul Islam
e-mail: ruhulislam786@gmail.com

to carry rather than PC or Laptop and available at reasonable price, thus mobile device users are increasing daily. The handheld mobile devices are still facing some issues regarding limited storage and bandwidth, less battery life, mobility, security, etc. M-commerce need high attention in security-related issues as it deals with online transaction where various user information are shared through Internet. Mobile Cloud computing (MCC) is combined technology which integrates mobile computing, wireless network and cloud computing technology, has become one of the latest industry buzz. The most important and most common MCC application is M-commerce or Mobile commerce which means accessing the Internet for business transaction using handheld devices. More and more business applications and research organization share trying for cloud implementation to get business benefits as it provides dynamic computing capacity and completely centralize approach, Cloud computing enables M-commerce organizations to rent required infrastructure rather than purchase hardware and software which decreases the overall cost of the system, this is beneficial for large-scale M-commerce outlets. The M-commerce applications can be categorized into some sections like shopping, finance, entertainment, e-auction, education, advertising, etc. [1]. But some important challenges faced by M-commerce application are security, limited storage and processing power, complex device configuration, low network bandwidth, etc. Cloud integrated M-commerce solves the issues as Cloud computing provides high security, larger storage space, fast computation, pay per use policy and most importantly on-demand services [3].

2 The Cloud Computing and Cloud Delivery Services

The Cloud computing is a mesh networking which values virtualization to permit computing application and data to be flexibly supplied from a pool of hardware resources. It provide various service model such as Communication as a service (CaaS), Platform as a service (PaaS), Software as a service (SaaS), Monitoring as a service (MaaS), Infrastructure as a service (IaaS), Security as a Service [4–6].

CaaS is responsible for management of H/W and S/W required for delivering VOIP Services, IM and video conferencing. It is fully integrated enterprise class unified communication and has facility like real time presence, soft phones, unified messaging, mobility, etc.

PaaS presents infrastructure upon which program developers can construct new application or continue already present application without having to buy a new server. PaaS multiplied the number of user by evolving, sustaining, and establishing www application.

SaaS is the most widely used, renowned and adds flavor of cloud computing. It can be characterized as a program where the applications are hosted by service provider and it is accessible to clients through Internet. The cloud applications in SaaS are accessed via an interface such as web browser as per requirement and pay for use policy. These services model provide the security to the clients as per

service-level agreement. The addition of visibility and control within SaaS applications with Aperture gives a full end-to-end security solution. These solutions are provided without changing any additional hardware, software, or network [7]. IaaS presents infrastructure of hardware for data center and program assets over the Internet. Clients have direct access to CPU processing, various hardware, virtual servers, network, Internet connectivity, and the storage devices in IaaS. Clients of IaaS are able to install and use the OS, software, on their virtual machine [4]. MaaS is cloud service model under anything as a service (XaaS). The MaaS framework deploys the monitoring capability for other services and application within cloud. It offers real time 24/7 monitoring to protect important information over the Internet. Various internal and external threats are protected by MaaS [8].

The most important cloud business model is Security as a Service upon which most of the service provider incorporate their security services within a collaborative infrastructure based on agreement. Authentication, intrusion detection, security event management, anti malware/spyware, confidentiality, etc., are offered by security as a service model.

3 Mobile Cloud Computing

Mobile cloud computing (MCC) is an infrastructure that combines wireless technology, mobile computing and Cloud computing [9]. In MCC the storage and processing of data is happen outside the handheld device, i.e., in the cloud. MCC is not only bringing mobile computing and its various application and services to mobile users but also extending it to mobile subscribers [3]. The very essential component and heart of MCC is Cloud computing, mobile devices, and wireless technology. MCC has many application areas like Mobile commerce (M-commerce), Mobile ticketing, Mobile Learning, e-auction, Mobile entertainment, Mobile Healthcare, Mobile Gaming, Mobile banking, and many other areas (Fig. 1).

4 Mobile Commerce

Mobile commerce or M-commerce is accessing the Internet for business transaction that is to purchase and sale of product, goods, and to access services using handheld devices [9]. It includes online banking, online buying of goods and services, online bill payment, online ticketing, information transfer and other online transaction related to business. Basically, if you pay for it on your phone, you are participating in Mobile Commerce. Nowadays most of the users spend more time on mobile phone rather than computer, laptop as it provides wireless connectivity, easy to carry, anytime, anywhere, anyhow accessibility, thus a consumer preferably searches for services and products using mobile phones. So, mobile users are increasing

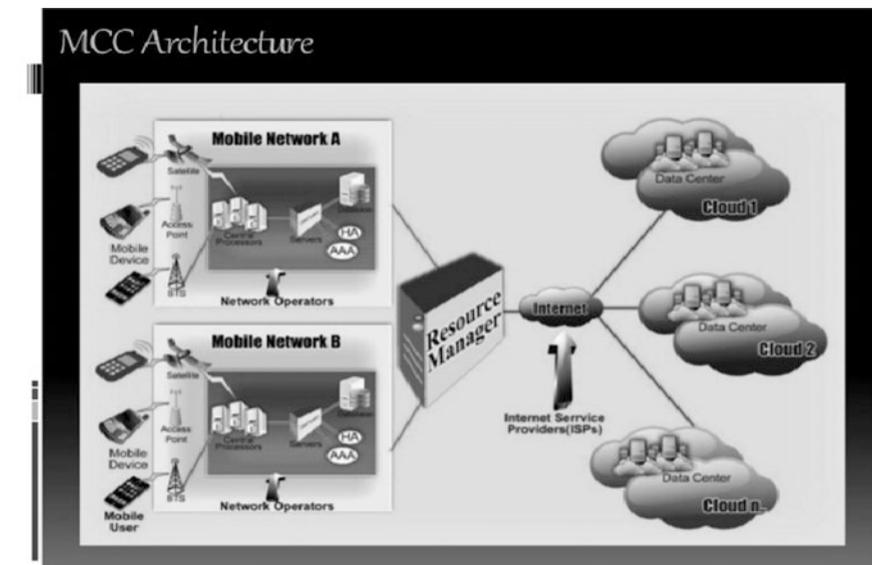


Fig. 1 Typical mobile cloud computing architecture [10]

daily. M-commerce application normally needs mobility such as mobile billing, mobile ticketing, mobile transaction and payment, etc., to achieve some task. Though M-commerce applications is developing still some issues exist regarding security, limited storage and processing power, complex device configuration, low network bandwidth, etc. [11]. The advances in wireless technology solves the problem regarding low network bandwidth as the 3G and 4G gives high-speed data transfer through wireless network [11, 12]. Among various M-commerce application online shopping, mobile ticketing, online banking are mostly used by user [10]. The big companies like Amazon, E-bay, Paytm, etc., are handling huge customer online transaction daily. Companies are processing huge data and customer information like user ID, password of online transaction need high security. The cloud computing, an extension of distributed system that provide huge storage, high security, fast computation, resource sharing, and on-demand services, solves the security related issues.

5 Mobile Commerce Security Issue and Solution

As mobile commerce application uses hand held devices (mobile phone, PDA, tablet PC, smart phone or a custom terminal) to access useful information and made online transactions for buying and selling of services or goods, each and every information has to be kept secured because the handheld devices are connected through cloud, and communicated via wireless network (Fig. 2).

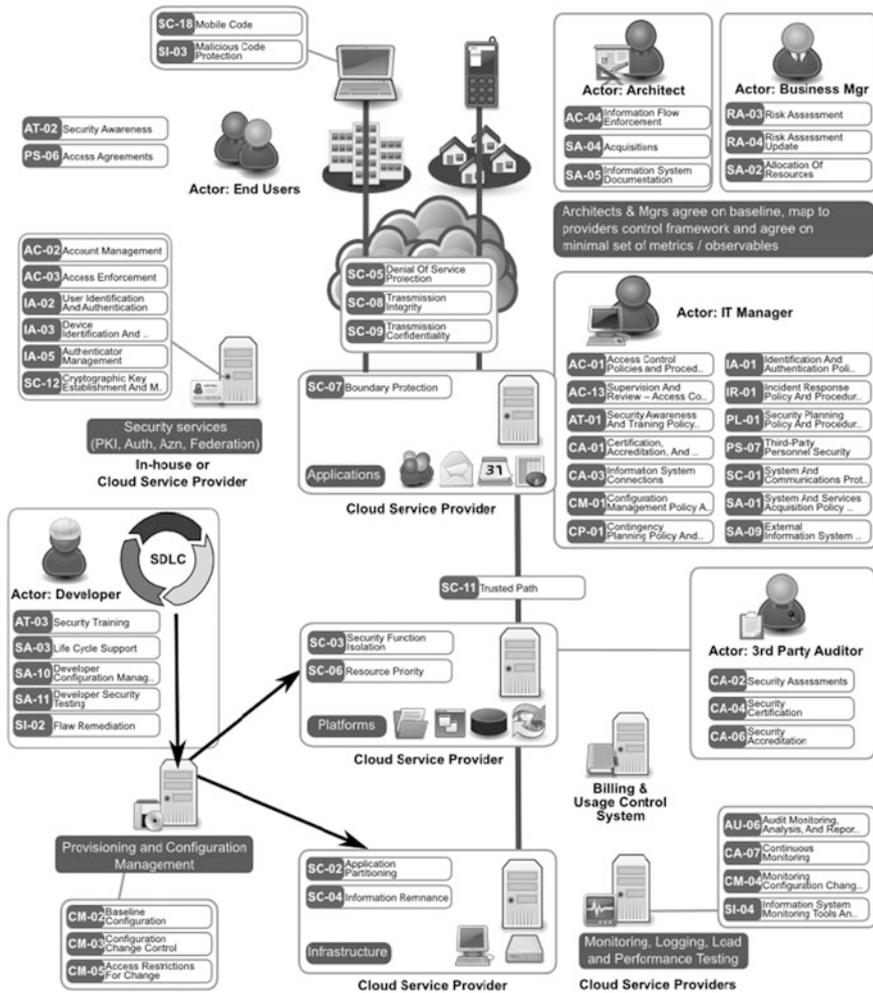


Fig. 2 Cloud security architecture [15]

The M-commerce application required mainly (i) transaction security, i.e., is to protect the transaction parties and their valuable data through an adequate level of security, (ii) Infrastructure security that aims to protect wireless network infrastructure from various attack and (iii) Information security to protect customer’s important and sensitive information from unauthorized access [13, 14]. By incorporating cloud computing in M-Commerce applications provide lots of benefit to business industries in the sense of security because a set of control based technologies and polices are included in Cloud computing to obey the regulatory compliance rules and secure data applications, information, and infrastructure related with Cloud computing use.

5.1 Threat Protection

Mobile devices are itself vulnerable to threat. Protection from malware is a common concern for network security [16]. One of the simple ways to protect the mobile devices is installing anti-malware software that detects and deletes the malware and keeps the device malware free. But mobile device has limited storage and installing and running security software will reduce the processing power and time consuming. By connecting through cloud and using the anti-malware software from cloud save the loss regarding performance of the device.

5.2 Encryption Options

Information authenticity, confidentiality, integrity is guaranteed by many existing technology. Various types of encryption techniques have been widely used in several security frameworks for MCC [16]. Encryption technique is one of the best mechanisms to protect valuable, sensitive customer's data and information (Fig. 3).

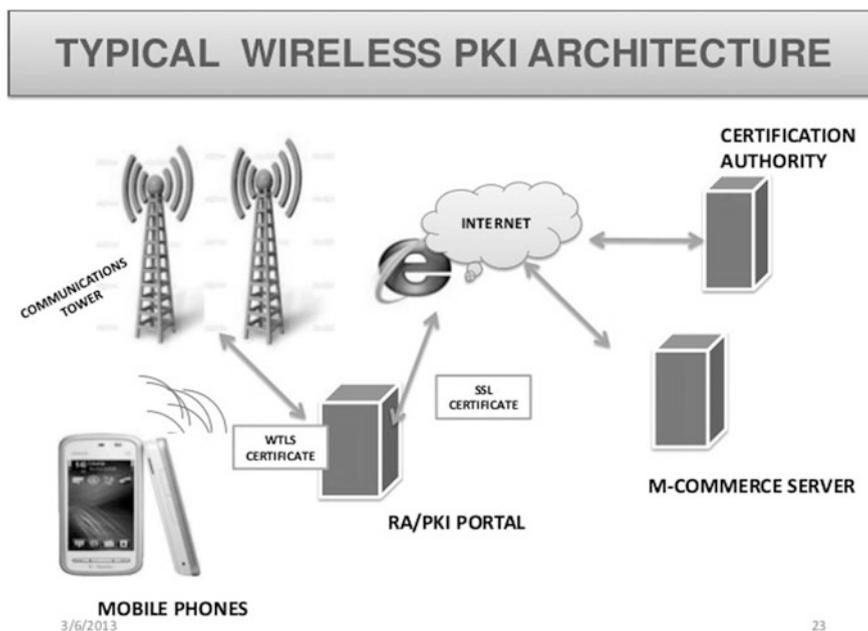


Fig. 3 A typical wireless PKI architecture [17]

5.3 Access Control

One of the important issues is access control to M-commerce systems where all users registered with their personal profile and they have appropriate access right according to their needs. Cloud access control provides proper and reliable authentication, authorization to the user. Thus M-commerce applications are securely accessed.

5.4 Availability and Non Repudiation

M-commerce application is securely executed through cloud by creating a security layer in the cloud. These security layers provide all the security-related services that are threat detection and prevention, create a secure communication channel, user identification and to provide safe, secure storage space to M-commerce user. Availability stands for services should be available all time, that is, an authorised user can reliably access his/her personal information to process his/her transaction. As mobile network unavailability is serious problem unlike wired network, it can be solved by integrating M-commerce to cloud as it provide anytime, anyhow, anywhere availability. On-repudiation is an assurance where after any user made transaction a user cannot deny on it. To meet these security issues normally a digital signature is used with M-commerce transaction such that after making any transaction a user cannot deny later.

6 Conclusion

Cloud computing is an emerging technology through which computing resources, computing application and data are flexibly supplied from a pool of hardware resources over the network, i.e., delivering any and all technology—ranges from CPU processing, storage, computing infrastructure, services, and application, hardware and software resources as a pay per use or on-demand service, i.e., whenever and wherever they required it. Rapid development in mobile technology, wireless technology increases the use of handheld devices. The development in the field of mobile cloud computing generates various opportunities for the business. One of the most common applications of mobile cloud computing is M-commerce. The cloud integrated mobile commerce will give lots of advantage for business industry as well as the user with lower cost high security in case of transaction, information, and infrastructure related to mobile device and wireless network.

References

1. H.T. Dinh, C. Lee, D. Niyato, P. Wang (2015) A survey of mobile cloud computing: architecture, applications, and approaches. *Wirel. Pers. Commun.: Int. J. Arch.* **80**(4)
2. <http://www.cetrom.net/uncategorized/mobile-cloud-computing-real-world-Examples>
3. Kumar, P., Soni, S.: A survey of mobile cloud computing architecture, applications, approaches & current solution providers. *Int. J. Res. Civil Eng.* **2**(7) (2015)
4. Ashraf, I.: An overview of service models of cloud computing. *Int. J. Multi. Curr. Res.* **2** (2014)
5. https://en.wikipedia.org/wiki/Security_as_a_service
6. Malik, S., Chaturvedi, M.M.: Privacy and security in mobile cloud computing: review. *Int. J. Comput. Appl.* **80**(11), 0975–8887 (2013)
7. Khan, A.R., Othman, M., Xia, F., Khan, A.N.: Context-aware mobile cloud computing and its challenges. *IEEE Cloud Comput.* **2**(03), 42–49 (2015)
8. Meng, S., Liu, L.: Enhanced monitoring-as-a-service for effective cloud management. *IEEE Trans. Comput.* **62**(9) (2013)
9. Dharmale, P.N., Ramteke, P.L.: Mobile cloud computing. *Int. J. Sci. Res. (IJSR)* **4**(1) (2015)
10. <http://www.slideshare.net/sairaraja077/mcc-architecture>
11. Yang, X., Pan, T., Shen, J.: On 3G mobile E-commerce platform based on cloud computing. In: *Proceedings of the 3rd IEEE International Conference on Ubi-Media Computing (U-Media)*, pp. 198–201 (2010)
12. Yılmaz, H., Olgaç, S.: Mobile communication and applications in tourism. *Managing the Asian Century*, pp 13–33 (2016)
13. ALL, S., FARAG, W., ROB, M.A.: Security measures in mobile commerce: problems and solutions. In: *The Fourth International Conference on Electronic Business—Shaping Business Strategy in a Networked World* (2008)
14. Bojjagani, S., Sastry, V.N.: STAMBA: security testing for android mobile banking apps. *Adv. Intell. Syst. Comput.* **425**, 671–683 (2015)
15. <http://www.opensecurityarchitecture.org/cms/library/patternlandscape/251-pattern-cloudcomputing>
16. Albuquerque, S.L., Gondim, P.R.L.: Security in cloud-computing-based mobile health. *IT Prof.* **18**(3) (2016)
17. <http://www.slideshare.net/RobbySahoo/real-time-mobile-e-business>

From FET to SET: A Review

Amit Agarwal, P.C. Pradhan and Bibhu Prasad Swain

Abstract Scaling has played an important role in reducing the size of the transistor so as to govern the Moore's law, but we cannot always simply scale down the size of the transistor without some deterioration in the performance of the transistor. These effects are termed as short-channel effects such as drain-induced barrier lowering, threshold voltage shift, leakage current, gate-induced drain lowering, hot carrier effect, etc. In this paper, we have reviewed different gated structures such as single gate, double gate, triple gate and gate all around which will control the electrostatic potential in the channel and reduce these short-channel effects. A molecular transistor, i.e., single electron transistor (SET) is also reviewed in this paper; SET shows better performance and reduced short-channel effects. In this molecular structure, either coulomb blockade or tunneling takes place, due to which it has better control over the flow of the electron.

Keywords Multi-gate MOSFET · Single electron transistor
Short-channel effects

1 Introduction

Metal-oxide-semiconductor field-effect transistor (MOSFET) has been delivering high performance for past decades in ultra-large-scale integration (ULSI) technology [1]. As to govern Moore's, MOSFET has been scaled down, but scaling of the

A. Agarwal (✉) · P.C. Pradhan
Department of Electronic and Communication Engineering,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: amiteng2007@gmail.com

P.C. Pradhan
e-mail: prashant.c.pradhan@gmail.com

B.P. Swain
Centre for Material Science and Nano Technology,
Sikkim Manipal Institute of Technology, Majitar, Sikkim, India
e-mail: bibhuprasad.swain@gmail.com

MOSFET is at its physical limits [2, 3]. Beyond this limits, if MOSFET is scaled down, then the performance of the device deteriorates. Due to these limits, there is high deterioration in the performance of the devices, these effects are termed as short-channel effects [4–6]. Some of these short-channel effects (SCE) are hot carrier effects, gate oxide tunneling, subthreshold swing, gate-induced drain leakage, threshold voltage roll-off, drain-induced barrier lowering, and voltage threshold shift [7, 8].

Short channel effect caused due to charge sharing will further lower the source channel barrier height and hence a higher threshold voltage roll-off. Degradation of subthreshold causes a high OFF-state current. Due to high drain voltage, high electric field is generated; which leads to reduction of the barrier height for carriers at the edge of source, this effect is termed as DIBL. When voltage drop between drain and source increases, it lowers the barrier of channel to source junction, as the barrier is reduced between channel and source, electron can easily pass through the barrier and transfer to channel area. Due to this, there is lowering of threshold voltage and control over the channel is lowered.

So, different types of structured devices have been detailed in this paper, i.e., multi-gate and molecular structure. In multi-gate structure, more than one gate are used, i.e., for a single-gate MOSFET there is only one gate, double-gate MOSFET has two gates on two sides of the channel, triple-gate MOSFET has three gates at three sides of the channel and in case of gate-all-around MOSFET, the entire channel is surrounded by gate. Due to these multi-gates, the electrostatic potential in the channel can be controlled and these short-channel effects can be reduced. Molecular structured transistor, i.e., single electron transistor has also been reviewed, in which a channel is replaced by a molecular structure creating a quantum dot [9]. This structure has a unique feature of coulomb blockade and tunneling. When the electron does not have enough energy to cross the barrier, then the device is in coulomb blockade state and no conduction takes place, but when the electron has gained some energy then, this electron can easily tunnel through the barrier and conduction takes place in the device. In this device there is better control, no leakage current, and the short-channel effects are reduced.

2 Short-Channel Effects

As we scale down the size (in nm) of the transistor, the size (in nm) of the channel, drain, and source are equal. Due to this, the performance of the transistor degrades. These effects are termed as short-channel effects. Following are some of the short-channel effects [10].

2.1 Mobility Degradation

Mobility degradation can be described by lateral and vertical field effect. In short channels, when there is increase in lateral field, these high fields affect the channel

mobility and therefore it becomes field dependent and velocity saturation occurs, due to which current saturation occurs. For short channels, when vertical electric field increases, near at the surface, carriers scattering takes place. Due to this, there is reduction in the surface mobility.

2.2 Subthreshold Swing

As channel length is reduced, threshold voltage decreases and subthreshold swing increases. As channel length decreases, lateral fields end up with more charge into the channel; this charge steals the charge which would have been terminated by voltage from the gate. Due to this stealing the channel-source barrier is reduced, this barrier is the one which controls electron conduction from source to drain [11].

2.3 Threshold Voltage Variation with Channel Length

As channel is decreased below 100 nm, there is decrease in threshold voltage, as charge in depletion region is also accompanied by charges of source and drain. Due to this, gate needs less charge to inject in this region and as a result, threshold voltage decreases, this effect is referred as charge sharing effect [5]. Since drain-to-source current is proportional to $(V_{GS} - V_T)$, so with the decrease of threshold voltage, drain-to-source current starts increasing. When gate-to-source applied voltage is zero, MOSFET is in cutoff mode, as threshold voltage is small negative, there is small leakage current [12].

2.4 Drain-Induced Barrier Lowering (DIBL)

With the increase of drain voltage, depletion region of the drain region increases, which interacts with channel-source junction and potential barrier is reduced. This effect is known as drain-induced barrier lowering (DIBL). When the junction barrier gets reduced, there is very simple injection of electrons into the channel, due to this there is no drain current control by the gate voltage [13].

2.5 Drain Punch Through

When a very high drain voltage is applied as compared to source voltage, drain depletion region will be extended to source, which causes a very high current to flow despite zero gate voltage applied. This effect is known as drain punch through. As channel length is decreased, punch through voltage decreases quickly [14].

2.6 Hot Carrier Effect

As dimension of transistor is very small, electric fields increases. Due to this, hot carrier effects appear for very small channel length. When the electric field increases in reverse bias drain junction, this can lead to carrier multiplication and impact ionization. This carrier generated may add up with the substrate current, it may also progress to source, which will lower the barrier and electron injection can take place easily [15]. These high energy electrons may also tunnel through the barrier into the oxide; these electrons get trapped into the oxide, which changes I - V characteristics and threshold voltage of the device.

2.7 Impact Ionization

During a very high longitudinal field, an electron gets very high velocity due to which there is generation of electron-hole pairs by impact ionization, this impact on atoms and ionizing them. Many of electrons get attracted by drain and hole enter substrate which forms parasitic substrate current. Due to this, there is base region developed between drain and source, where source acts as emitter and hole acts as collector, which is same as in npn transistor [16]. As this hole is collected by source, there is voltage drop of 0.6 V in the substrate material which is created by this hole current and reverse-biased source substrate p - n junction will start conduction. There is injection of electron from source to substrate, analogous to emitter to base. These electrons gain enough energy and move toward drain, creating an electron pair. The device may be permanently damaged, if the generated electrons due to high fields may escape the field of the drain and travel to substrate, and affects other devices.

3 Single- and Double-Gate MOSFET

Single-gate SOI MOSFETs as shown in Fig. 1a consist of a source, drain, one gate, channel, and an oxide which separate gate and the channel. These are on top layer of the buried oxide, i.e., silicon on insulator. These SOI devices have shown lower parasitic capacitances with higher packing density [17]. This device is having single gate to control the electrostatic potential on the channel. This devices exhibit short-channel effects when the dimension is below 100 nm. So, the structure is modified with two gates on two sides (one front gate and one back gate) and the channel lies between these two gates and is termed as double-gate SOI MOSFET as shown in Fig. 1b. Due to use of two gates, the electrostatic potential in the channel can be controlled in better way compared to single-gated structure and therefore some reduction of short-channel effects [18]. But as we further decrease the size of the transistor, then this device again starts to exhibit short-channel effects.

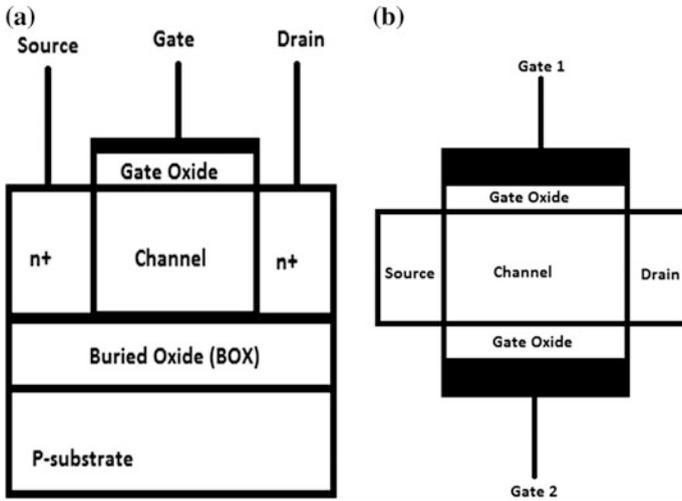


Fig. 1 Schematic view of single-gate SOI MOSFET [36] (a) and double-gate SOI MOSFET [36] (b)

4 FinFET

FinFET allows further scaling of channel length, it may surpass SOI MOSFET. The unique characteristic of finFET is that its gate straddles a thin fin shaped body, formed on three sides of the channels, along the top and on the vertical sidewall surfaces of the fin [19–21]. With the use of three gates which surrounds the channel exhibit brilliant electrostatic control. The width of the fin should be approximately half of the channel length and minimum width dimension of the fin, is as per lithography. Because of the vertical nature of finFET, it provides larger device width per wafer area, due to which finFETs can be densely packed. In FinFET one of the challenges is to fabricate uniform narrow fins. As the fins are narrow, high access resistance is created thereby reducing the ON current. Implementation of high strain into fins, so as to boost ON current is another challenging issue.

5 Triple-Gate MOSFET

Triple-gate MOSFET is a thin film, with three gates on the three sides of the channel. A thick buried oxide layer is placed on top of the substrate. On the buried oxide layer, source, gate, channel, and drain are located. Due to this, there is isolation between channel and substrate. As there are three gates, these gates have control over the channel and reduce the short-channel effects [22] (Figs. 2 and 3).

Fig. 2 Schematic of triple-gate SOI MOSFET [36]

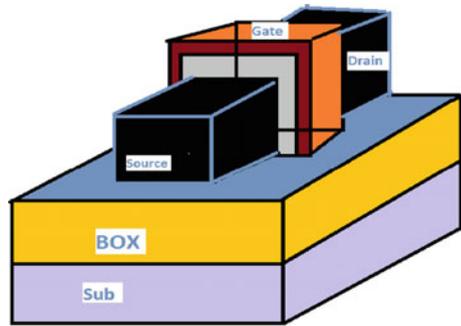


Fig. 3 Schematic of FinFET [21]

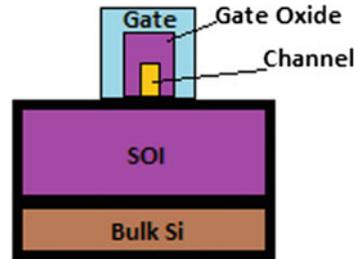
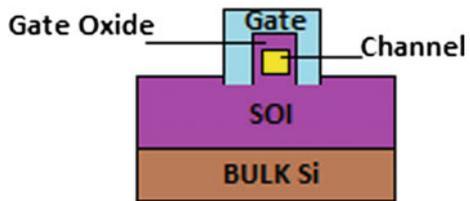


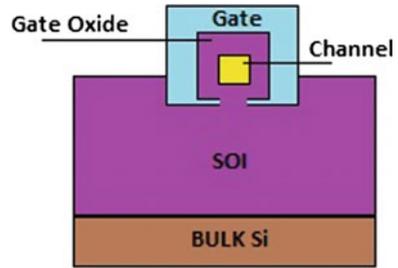
Fig. 4 Schematic of Pi-gate [21]



6 Pi-Gate MOSFET

The Pi-gate SOI MOSFET have the capabilities to replace the existing devices, as it exhibits better control over short-channel effects (SCEs) and leads to reduction in gate leakage currents, which reduces power dissipation and as the geometry of the Pi-gate MOSFET is in between triple and quadruple gate SOI MOSFET [23]. In Pi-gate MOSFET, the gate is grown into a buried oxide, and enhancing electrostatic potential control in the channel and protecting it from the electric field lines which are originating from the drain [24]. The Pi-gate SOI MOSFET can be easily manufactured as compared to double-gate or a gate-all-around structure (Figs. 4 and 5).

Fig. 5 Schematic of Omega Gate [21]



7 Omega MOSFET

Omega-FET is another good candidate for field-effect transistor and exhibits better scaling down characteristics, where a gate is almost wrapped around the body [25]. As gate all around (GAA) offers the better gate control it faces complexity in fabrication process and other difficulties. Omega-FET is almost similar to GAA with respect to electrical performance, but it uses a manufacturing process similar to FinFET. Omega gate MOSFETs possess the best scaling down characteristics, but it suffers from severe corner effects and difficulty in processing due to poly silicon gate residue under the bottom corners of the body. Gate-all-around MOSFET has best scaling down characteristics but the steps for processing are difficult [26].

8 Gate-All-Around MOSFET

As the transistor size is decreased below a 15 nm, the above multi-gated devices fail to reduce the short-channel effects, so an alternative approach is to use a gate-all-around (GAA) device (square or cylindrical). In these GAA devices, channel is totally surrounded with gate; either its square or cylindrical in shape [27–30].

So GAA is best device configuration to control electrostatic potential for a narrow channel length transistor and minimize leakage current of the device when device is in off-state condition and makes the device to operate with less dissipation. Several GAA geometries are possible either in horizontal or vertical configuration [31]. Benefit of GAA structure for a transistor is not only to allow transistor to be scaled down, but it also helps to achieve a high packing density due to enhanced current drive [32].

The two different shapes in GAA are cylindrical and square-shaped gate. In square-shaped GAA, the gate and the channel are square in shape, but in case of cylindrical structure, the gate and the channel are cylindrical in shape. Among these GAA structures, square GAA exhibits high current driving capability than that of cylindrical GAA. So, square GAA is one of the most promising structures presently available to continue scaling down, due to greater control over the electrostatic potential in the channel and reduces the short-channel effects.

Fig. 6 Schematic of Pi-gate [21]

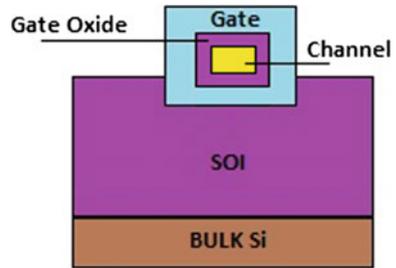
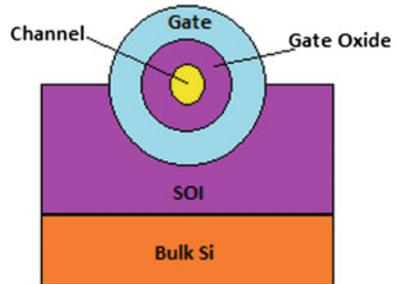


Fig. 7 Schematic of Omega Gate [21]



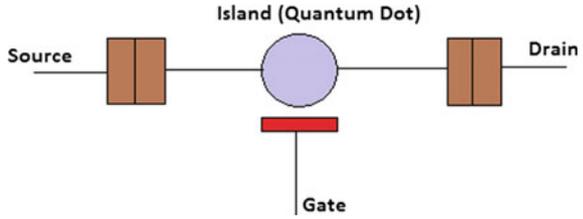
In spite of such advantages of square GAA, it exhibits a very unattractive trait, i.e., corner effect. Electrostatic blend of two adjacent gates on the corners of GAA leads to this corner effects. This corner effect degrades the performance of the device by increasing off-state leakage current. By rounding of the corner regions, this effect can be reduced. Rounding of the corner is fragile process.

Another method to reduce this corner effects is to implement cylindrical GAA, since the shape is cylindrical, so there are no corners, due to which such effects will not exist. Scaling down transistor dimension has been one of the chief factors for improvement in the cost and IC performance, and has shown a rapid growth in the semiconductor industry. But still GAA devices exhibit short-channel effects (Figs. 6 and 7).

9 Molecular Technology—SET Gate MOSFET

Molecular structure based transistor presently is one of the most eligible technology to replace GAA MOSFET and govern Moore's law. One of the molecular transistors is single electron transistor. SET consists of source, drain, gate, and island. SET uses two tunnel diodes as shown in Fig. 8. In case of MOSFET there was quantum wire as channel, but in case of SET, there is quantum dot as an island (Island is a quantum dot). Island is sandwiched between source and drain and

Fig. 8 Schematic of single electron transistor [9]



controlled by a gate. There are two phenomena taking place in the SET: coulomb blockade and tunneling.

Coulomb blockade, i.e., when potential of a electron is less than the barrier, then SET is in coulomb blockade state, i.e., there is no transfer of electron from source to drain, but when electron acquires enough energy then tunneling phenomenon takes place, and electron gets transferred from source to drain [9, 33, 34]. In SET, electron conduction takes place one by one and gate voltage only changes the potential of the island for better controlling for the tunneling. It also exhibits both negative and positive transconductance [35].

10 Conclusion

Due to scaling down of transistor below 100 nm, there is deterioration in the performance of the transistor. These deteriorating parameters are known as short-channel effects. Some of short-channel effects are hot carrier effects, threshold voltage roll-off, gate oxide tunneling, subthreshold swing, gate-induced drain lowering, drain-induced barrier lowering, and voltage threshold shift. In this paper, we have reviewed single-gate, double-gate, triple-gate, Pi-gate, Omega gate, gate-all-around MOSFET, and molecular structure based transistor (SET). As we increase the number of gates, control over the electrostatic potential in the channel also increases. GAA MOSFET shows better performance over other single, double, triple, Pi and Omega gate, when channel length is reduced below 15 nm. But still the short-channel effect is not fully reduced and fabrication process is quite complicated. So, molecular structure based transistor, i.e., SET has shown better characteristics than GAA MOSFET. Since SET has unique characteristics of coulomb blockade, i.e., when potential of a electron is less than the island barrier, then SET is in coulomb blockade state, i.e., there is no transfer of electron from source to drain, but when electron acquires enough energy then tunneling phenomenon takes place, and electron get transferred from source to drain. SET has better control over the flow of electron and also reduces the leakage current.

References

1. Bardeen, J., Brattain, W.H.: The transistor, a semi-conductor triode. *Phys. Rev.* **74**, 230–231 (1948)
2. Moore, G.E.: Progress in digital integrated electronics. *Electron Devices Meeting 1975*, pp. 11–13 (1975)
3. Baccarani, G., Wordeman, M., Dennard, R.: Generalised scaling theory and its application to a $\frac{1}{4}$ micrometer MOSFET design. *IEEE Trans. Electron. Devices*, pp. 452–462 (1984)
4. Naveh, Y., Likharev, K.K.: Shrinking limits of silicon MOSFETs: numerical study of 10 nm scale devices. *Superlattices Microstruct.* 111–123 (2000)
5. Chaudhry, A., Kumar, M.J.: Controlling short-channel effects in deep submicron SOI MOSFET's for improved reliability: a review. *IEEE Trans. Device Mater. Rel.* 99–109 (2004)
6. Packan, P.: Pushing the limits. *Science* 285 (1999)
7. Garduño, S.I., Cerdeira, A., Estrada, M., Alvarado, J., Kilchystka, V., Flandre, D.: Contribution of the carrier tunnelling and gate induced drain leakage effects to the gate and drain currents of fin-shaped field effect transistor. *J. Appl. Phys.* 1–7 (2011)
8. Liu, Z.H.: Threshold voltage model for deep submicrometer MOSFETs. *IEEE Trans. Electron Devices.* 40 (1993)
9. Likharev K.: Single-electron devices and their applications. *Proc. IEEE.* **87**, 606–632 (1999)
10. Yu, B., Wann, C.H.J., Nowak, E.D., Noda, K., Hu, C.: Short channel effect improved by lateral channel-engineering in deep-submicrometer MOSFETs. *IEEE Trans. Electron Devices* 627–634 (1997)
11. Zhang, Q., Zhao, W.: Low subthreshold swing tunnel transistors. *IEEE Trans. Device Lett.* 297–300 (2006)
12. Banna, S.R., Chan, P.C.H., Ko, P.K., Nguyen, C.T., Chan, M.: Threshold voltage model for deep-submicrometer fully depleted SOI MOSFET's. *IEEE Trans. Electron Devices* 1949–1955 (1995)
13. Wolf S.: *The Submicron MOSFET*, vol. 3 of *Silicon Processing for the VLSI Era*. Lattice Press, Sunset Beach (1995)
14. Zhu, J., Martin, R., Chen, J.Y.: Punch through Current for submicrometer MOSFET's in CMOS VLSI. *IEEE Trans. Electron Devices* 145–151 (1988)
15. Fischetti, M.V., Laux, S.E., Crabbe, E.: Understanding hot-electron transport in silicon devices—is there a shortcut. *J. Appl. Phys.* 1058–1087 (1995)
16. Ning, T.H., Osburn, C.M., Yu, H.N.: Emission probability of hot electrons from silicon into silicon dioxide. *J. Appl. Phys.* **48**, 286–290 (1997)
17. Chaudhry, A., Kumar, M.J.: Controlling short-channel effects in deep submicron SOI MOSFET's for improved reliability: a review. *IEEE Trans. Device Mater.* **4**, 99–109 (2004)
18. Balestra, F., Cristoloveanu, S., Benachir, M., Brini, J., Elewa, T.: Double-gate silicon-on-insulator transistor with volume inversion: a new device with greatly enhanced performance. *IEEE Electron Device Lett.* **18**, 410–412 (1987)
19. Park, T., Choi, S., Lee, D.H., Yoo, J.R., Lee, B.C., Kim, J.Y., Lee, C.G., Chi, K.K., Hong, S. H., Hyun, S.J., Shin, Y.G., Han, J.N., Chung, U.I., Moon, J.T., Yoon, E., Lee, J.H.: Fabrication of body-tied FinFETs (Omega MOSFETs) using bulk si wafers. In: *Technical Digest of Symposium on VLSI Tech* (2003)
20. Xiong, W.W.: Multigate MOSFET technology. In: *Colinge J.P. (ed.) FinFETs and Other Multi-Gate Transistors*. Springer, US, pp. 84–91 (2008)
21. Colinge, J.P.: *FinFETs and Other Multi-Gate Transistor*. Springer, Berlin (2008)
22. Schwarz, M., Holtij, H., Kloes, A.: MOS: a new physics-based explicit compact model for lightly doped short-channel triple-gate SOI MOSFETs. *IEEE Trans. Electron Devices* **59**, 349–358 (2012)
23. Park, J.-T., Colinge, J.-P., Diaz, C.H.: Pi-gate SOI MOSFET. *IEEE Electron Device Lett.* **22**, 405–406 (2001)

24. Daugé, F., Pretet, J., Cristoloveanu, S., Vandooren, A., Mathew, L., Jomaah, J., Nguyen, B. Y.: Coupling effects and channels separation in FinFETs. *Solid State Electron.* **48**, 535–542 (2004)
25. Jiang, Y., Singh, N., Liow, T.Y., Lim, P.C., Tripathy, S., Lo, G.Q., Chan, D.S.H., Kwong, L.: Omega-gate p-MOSFET with nanowire like SiGe/Si core/shell channel. *IEEE Electron Device Lett.* **30**, 392–394 (2009)
26. Barraud, S., Coquand, R., Cassé, M., Koyama, M., Hartmann, J.-M., Maffini-Alvaro, V., Comboroure, C., Vizioz, C., Aussenac, F., Faynot, O., Poiroux, T.: Performance of omega-shaped-gate silicon nanowire MOSFET with diameter down to 8 nm. *IEEE Electron Device Lett.* **33**, 1526–1528 (2012)
27. Jimenez, D., Saenz, J.J., Iniguez, B., Sune, J., Marsal, L.F., Pallares, J.: Modeling of nanoscale gate-all-around MOSFETs. *IEEE Electron Device Lett.* **25**, 314–316 (2004)
28. Song, J.Y.: Design optimization of gate-all-around (GAA) MOSFETs. *IEEE Trans. Nanotechnol.* **3**, 186–191 (2006)
29. Iniguez, B., Fjeldly, T.A., Lazaro, A., Danneville, F., Deen, J.: Compact modelling solutions for nanoscale double-gate and gate-all-around mosfets. *IEEE Trans. Electron Devices* **53**, 2128–2142 (2006)
30. Smaani, B., Latreche, S., Iniguez, B.: Compact drain-current model for undoped cylindrical surrounding-gate metal-oxide semiconductor field effect transistors including short channel effects. *J. Appl. Phys.* **114**, 224507–224512 (2013)
31. Haensch, W., Nowak, E.J., Dennard, R.H., Solomon, P.M., et al.: Silicon CMOS devices beyond scaling. *IBM J. Res. Dev.* **50**, 339 (2006)
32. Li, C., Zhuang, Y., Han, R.: Cylindrical surrounding-gate MOSFETs with electrically induced source/drain extension. *Microelectron. J.* **42**, 341–346 (2011)
33. Acharya, M.: Development of room temperature operating single electron transistor using FIB etching and deposition technology. Michigan Technological University (2009)
34. Mahapatra, S., Ionescu, A.M.: Hybrid CMOS single-electron transistor device and circuit design. Artech House, pp. 129–165 (2006)
35. Ihn, T.: Semiconductor Nanostructures: Quantum States and Electronic Transport. Oxford University Press (2009)
36. Mohammadi, H., Abdullah, H., Dee, C.F.: A Review on Modeling the Channel Potential in Multi-Gate MOSFETs, vol. 43, pp. 861–866. Sains Malaysiana Publications (2014)

Important Factors for Designing Assamese Prosody with Festival Frame Work

Parismita Sarma and Shikhar Kumar Sarma

Abstract Prosody is a term related with literature as well as speech technology. It is one of the primary parts for design of any natural sounding text to speech synthesis. Prosody is a broad as well as complex way of expressing meaning of a text segment in terms of pitch means fundamental frequency of utterance, loudness or intensity, intonation, and rhythm. The term is related to phonology and used to convey meaning as well as structure or configuration of a speech unit. This paper discusses about the important issues to design a well-versed prosody with respect to above-mentioned properties. Though our aim is to design an Assamese prosody in festival framework, yet we are highlighting the features which must be taken care of regardless of any specific language.

Keywords Prosody · Pitch · Intensity · Intonation · Festival framework

1 Introduction

Twenty-first century is the age of information technology. Nowadays people want to engage in more than one tasks at the same time. For example connecting with a multimedia conference, participating in a conversation users want to work on more than one device which support basic communication system [1]. Speech is one of the easy communication media for human to interact with machine. Speech processing including speech synthesis draws attention of many researchers to design new device which may be connected with speech synthesis. We are working to

P. Sarma (✉)

Department of Information Technology, Gauhati University, Guwahati, Assam, India
e-mail: parismita.sarma@gmail.com

S.K. Sarma

Cotton College State University, Guwahati, Assam, India
e-mail: sks001@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_22

211

design a speech synthesizer for Assamese language. In this paper, we are discussing an important issue necessary to design a synthesizer. Our respiratory system produces air pressure and this pressure comes out of lungs with full of energy. It goes through our articulatory system and produce sound wave, which is called speech. Speech is an acoustic signal, with a number of components. The whole speech technology is divided to different subparts. They are such as speech analysis, speech coding, speech recognition, speech synthesis, etc. Speech conveys information which can be analyzed in different levels. These level includes phonetic, phonological, syntactic, or acoustic [2]. A Text-To-Speech (TTS) system predicts to speak in natural way. Though it is quite difficult or impossible to predict human speaking by means of a machine, yet many efforts are going on by researchers all over the world. Talking calculator is the simplest form of speech synthesizer. The main problem while designing a speech synthesizer is its speaking quality. Joining two words with appropriate pronunciation and time gap in between is also challenging. The problem can be solved with the help of an efficient speech prosody. Prosody can be said as core of a language synthesis process. We are working on speech synthesizer for North East Indian language 'Assamese'. India is rich of total 1652 dialects for native languages and it is a multilingual country [3]. In the post-period of 1990, Indian speech synthesizer research work began to progress and generated real-time synthesizer for some of the Indian language [3].

Assamese is an Indo-Aryan language. Assam is a state situated in northeast part of India. The people belonging to Assam speak in Assamese language. Assamese language has eight vowel phonemes, 21 consonants, and 15 diphthongs [4]. In this paper, we are discussing some important points regarding prosody design of Assamese language in festival framework. Without proper prosody, spoken text becomes dull and monotonic. Our aim is to dispense proper emotions to the text, such that when somebody speaks it, meaning and intention of the speaker gets clear. In Assamese language, it is so important because some sentences with same word arrangement express different meanings depending upon the tune of the speaker. By looking at the text only sometimes, a nonnative reader is unable to guess the meaning of the sentence. Machine takes help of prosody to understand the features like tone, time duration, and prosodic phrasing. For example, an Assamese sentence 'ৰাম ঘৰলৈ গৈছে' means Ram has gone home, it may carry different meanings depending upon utterance tune of the speaker. If the sentence is 'ৰাম ঘৰলৈ গৈছে?', it will mean that Has Ram gone home? For an interrogative sentence depending upon which word is most stressed, query also becomes different. In some languages, for example, English position of the verb is enough to indicate the semantic and intention of the speaker. But in Assamese language it does not happen. Only the tune is sufficient to differentiate an affirmative and interrogative sentence. So when designing an Assamese TTS, much attention should be given to design a proper prosody, then only real meaning will come out of the text the writer wanted to render exactly.

Table 1 Acoustic quality with measuring property of utterance

Features	Place of articulation	Unit	Scale
Fundamental frequency	Vocal folds vibration	F_0 (Hz)	Low, high
Loudness	Air pressure in sub glottal part	Intensity (dB)	Loud, suppress
Length	Speech gesture	duration (ms)	Short, long
Vowel property	Vocal tract configuration	Formants ($F_1, F_2...$)	Reduced, full

2 Design Approach

Prosody designing has a number of sub-phases like phrasing, duration, intonation, and intensity modeling [5]. There are basically two major approaches for prosody design in festival. They are rule based and corpus based [6]. We are interested in corpus-based approach and considered it for our design. In rule-based approach, a complicated rule set is prepared from natural speech. The rules are not sure to work in all cases. Corpus based approach can give more natural speech, because corpus is prepared from real speech units uttered by human. At the time of synthesis the appropriate units are selected and concatenated that sounds like a human utterance. In rule based approach voice waveforms are edited and rules are generated from the analysed waveform for different types of expression. Signal processing has to be done on the speech wave and this lacks the naturality of the speech. In corpus-based approach, a huge well-formed corpus is designed. Corpus has to be annotated with various types of prosodic information. Prosody design in corpus-based approach is iterative type. Corpus is activated automatically for creation of actual prosodic model. At first it is applied on test data. After that if necessary, it can be improved more. Some of the necessary acoustic qualities of utterances for prosody design are given in Table 1.

Our work is based on festival open source software. Festival is an open source software used for speech synthesis system. It is used in many institutes, research organization, and commercial places. Festival has no specific built-in language, but any new language can be added to festival. It has a number of tools with the help of which rapid or interactive development is also possible. Festival is incorporated with many statistical tools, different scripts can be written to build individual models. It is a general system which supports multilingual TTS. Coding done for festival is in C or C++. Basic computer language for programming is scheme scripting language. It is a multi-platform support tool. Both UNIX and Windows can be used for programming. A group of modifiable modules are there: e.g., duration, phrasing, LTS, lexicon, etc. These can be rewritten for a specific language.

3 Intonation Design

Intonation can be explained as speech melody. Intonation is related with *F0* contour or fundamental frequency of utterance. In other words it is pitch. The same sentence may have different pitch contours (may be falling or rising) when they are uttered in two or three different moods. We discuss about pitch movement when we talk about fundamental frequency contour. Covering range may rise or fall, which again has importance during design. Context of every token should be considered when rise or fall occurs. Two most important things of pitch observation are: prominence portion and phrasing. Phrasing is chunks of speech.

Intonation model can be divided into two category of works [6]:

1. Accent or tone assignment: Allocation of tone is done on syllable or word level. Initially which syllable has to accented more is identified and the type also evaluated.
2. *F0* contour generation: When there is accents, *F0* contour can be generated.

By assigning accent, we identify the prominent syllabi or words in the sentence. We can have different types of syllabic intonations, rise, flat, and fall. For an interrogative sentence, naturally there is gradual rise in *F0* contour at end of the sentence. *F0* contour is flat when declarative or affirmative sentences are uttered. Researcher has proved this property for almost every type of sentence. In general, boundary tones are rising or falling as at those points intonation occurs. The boundaries are detected by prosodic phrasing method. In festival, accent is assigned for content words in the stressed syllables. By default this is done in festival. A decision tree called CART is used for this purpose. Festival wagon tools [7] should be used if predicted accent is used.

For example one of our wave file is manalisha_a_1.wav. Our festival output is given below. First column is the ordinal number of the entity, second one is the token number, third column is for syllable itself and fourth one is the stress assigned to the token. When last field is 0, it means no stress, 1 means there is stress. E.g.

```

.....
.....
27 id_30; name syl 1; stress 0;
28 id_32; name syl 2; stress 0;
29 id_34; name syl 3; stress 0;
.....
.....

```

3.1 *F0* Contour Generation

F0 is said to be fundamental tune. From our experiment, it is seen that for male *F0* ranges between 95 to 130 Hz and 135 to 285 Hz for female speaker in normal

utterance. For female, this value is more according to general property of F_0 . F_0 contour shows higher value at beginning of a sentence, gradually it declines during utterance. We can rise up or rise down an utterance if we want to. Intonation conveys dynamic and temporal information to the listener. For a normally uttered sentence, it is observed that F_0 shows almost a flat curve. But for a sentence with interrogative tune, contour may suddenly rise high up or fall down quickly at end. Festival has a number of procedures to generate target F_0 contour. A function written in LISP for festival can generate or specify a list of F_0 contour target point for every syllable. We have used Praat tool to study the fundamental frequency F_0 for a word or sentence. We are discussing about a Assamese sentence ‘ৰাতুল আজি আহিব’, the english meaning is ‘Ratul will come today’ in three different moods affirmative, interrogative and surprise. Figure 1 shows the male recording with its F_0 contour, which was recorded in a sophisticated well-equipped noise free studio. In normal mode, the fundamental frequency (F_0) is found as 96.324494 Hz in duration of 0.726961s. From the figure, it is clear that the curve is flat type. All the figures given below show fundamental frequency (F_0) for a sentence but in three different emotions. For everyone of them, on the x -axis time or duration in seconds and on y -axis frequency in Hz is plotted.

For the same sentence in interrogative mood, the F_0 or fundamental frequency is 111.763313 Hz in 0.969 s duration. The wave with F_0 contour is shown in Fig. 2.

On the other hand, in surprise mood, the fundamental frequency or F_0 contour is found as 168.098928 Hz in total 1.211610 s as shown in Fig. 3.

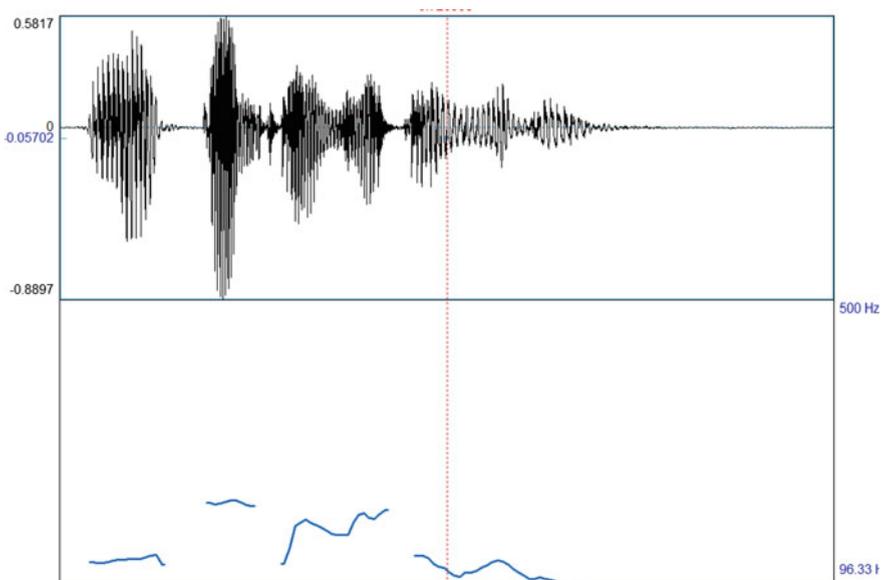


Fig. 1 Pitch or fundamental frequency (F_0) of affirmative sentence ‘ৰাতুল আজি আহিব’

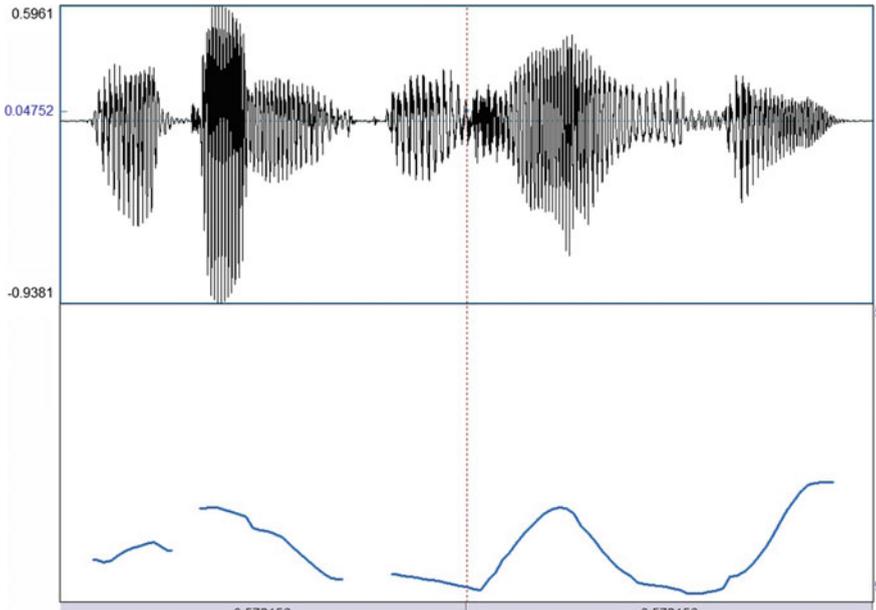


Fig. 2 Pitch or fundamental frequency (F_0) for interrogative sentence ‘বাতুল আজি আহবি?’

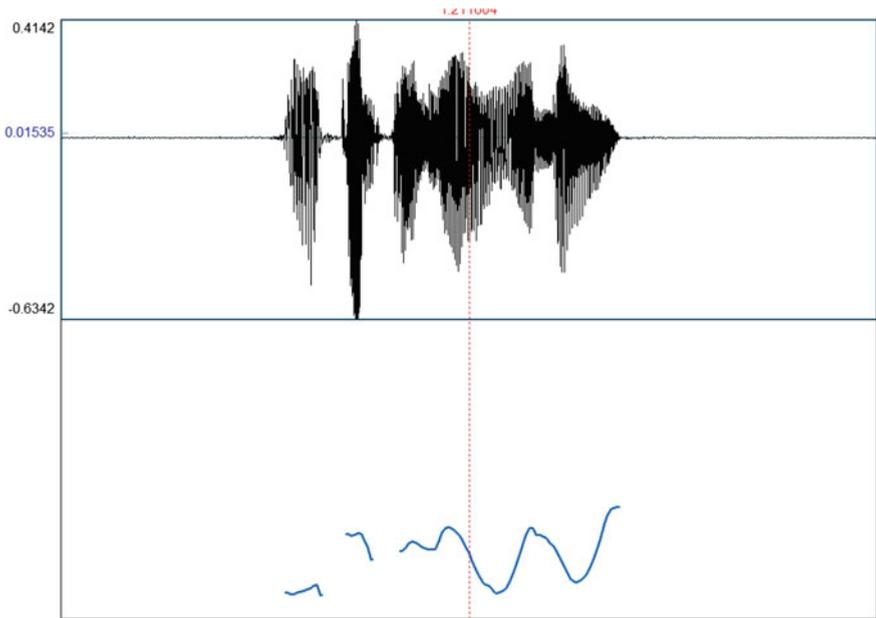


Fig. 3 Pitch or fundamental frequency (F_0) of exclamatory sentence ‘বাতুল আজি আহবি!’

4 Utterance Segmentation

Utterance segmentation means phoneme, syllable, or words. In festival a set of related items are used to represent utterance. Items are represented by a relation [8]. There may be more than one relation in an item. Item means it may be a sentence, word, or phrase. On the other hand, a relation is a list of trees. Each item is capable of expressing an object like word, syllable, etc. A number of features can contribute to an item. Features have a name and value, relation associate the items with its neighboring ones in some meaningful way. The way may be syntax tree, word list, or syllable structure. In this way, every utterances can be selected. An utterance should be prominent enough such that it is separable and has detectable syllable boundary. Another point is within narrow band spectrum fundamental frequency and formants should be detectable. These are called informative utterance [9]. Below we have shown a relation matrix for a series of token in a sentence, expressed as a relation.

Relation Token; ()

18	29	1	0	0	0
1	1	0	18	2	0
19	28	2	0	0	0
2	2	0	19	3	1
20	27	3	0	0	0

The above matrix defines a relation for a token. This structure includes spaces between the words in a sentence as a token. The first number in a row may represent ordinal number, second is the token number, third number is the position in the sentence of that word, etc.

5 Duration

It is observed that assignment of proper duration to every lexicon is quite necessary for proper prosodic pronunciation. Duration is influenced by many factors like near by units, how many total numbers of speech unit, etc. [9]. This model has importance because without proper duration, individual speech units may be overlapped or will be much away. So for proper prosodic pronunciation duration is an important factor.

In festival, there are number of methods to generate proper duration to individual units. As our work is restricted to news reading mood, there is no much consideration of tune for the individual sentence. We have assigned a fix duration of 100 ms to every unit in our work. It is the easiest method. Basically when there is a cause of proper emotional pronunciation then duration with varying values is

needed. Another easier method can be taken as average duration values assigned to every phoneme.

e.g., Festival outputs are as follows:

manalisha/wav/manalisha_a_1.wav	5.1300
manalisha/wav/manalisha_n_1.wav	4.1000
manalisha/wav/manalisha_q_1.wav	3.4100

The last figure in each row gives the duration of the whole sentence.

For individual syllable time can be generated. We have used a sophisticated syllabification program to syllabify the sentences. For the Assamese sentence ‘ৰাতুল আজি আহিব’, we can get output as shown below. This is not a table but an output.

0.1100	100	SIL (silence)
0.2200	100	ৰা_beg
0.3300	100	তুল_end
0.4400	100	আ_beg
0.5500	100	জি_end
0.6600	100	আ_beg
0.7700	100	হিব_end

_beg means that syllable is at beginning of the word and _end means at end of the word. The first column represents starting time of the syllable. In this example a fixed duration of 0.11 s is taken. It is added up with each next item.

6 Prosodic Phrasing

Prosodic phrasing identifies pause length between two sentences. Even in the same sentence, sometimes we can see two types of prosody or tune if there is a punctuation in between. Human can understand the meaning when he or she reads the text. But for the machine, it is difficult to understand the inherent meaning of the sentence. In this case, proper and correct prosodic phrasing can help a machine. For example ইমান খুনীয়া গাড়ী!, তুমি কিনিবা? It means So much beautiful the car is! you will purchase? There are two prosodic phrasing in the sentence. If the pause between the two sentences is not accurate, meaning will be incorrect. Prosodic phrasing helps in understanding meaning of the whole sentence and work rather well for proper synthesise. For all these, identifying the boundary of the phrasing is important. There are two main methods in festival for prosodic phrasing. One is the by rule and the other is through statistical model.

6.1 *The Rule Method*

In festival CART tree is used for rule method. The rules may be handwritten or trained. First choice in festival is always CART tree [10]. Test is made on every word, the test is done to determine whether the word is at last of prosodic phrase. The CART tree returns B or BB. B means Break and BB means Bigger Break, when there is no need of more pause it is break, so B. On the other hand, BB means the end of Utterance. This procedure is simple and simply adds break (B) after the last token if a punctuation like \, or ;. If there is ?, : or . after the last word, BB is inserted.

6.2 *Phrasing by Statistical Model*

Statistical approach is used in this model. This model tries to find a likelihood of break; it may be after every word or after every sentence. According to this rule it predicts a break if there is another break in close distance. Festival basically supports that model which takes probability of break based on previous. Its implementation lies on the model which is based on parts of speech of the concerned language.

After doing all these observations, we conclude that most important design of speech prosody is design of a good intonation and duration model. Prosodic phrasing comes after that only. Whatever it is, every module should be efficient enough to design a real prosody.

7 Conclusion

This paper is an effort to discuss the important factors when designing a speech prosody. We have discussed how festival builds different acoustic properties for utterances with respect to Assamese language. Properties can be identified properly in digital representation. But the most important thing is, to design an efficient prosody; we must have a rich speech corpus for that language. As our work is Assamese prosody design, we went through whatever available Assamese corpuses are there. Only a few Indian languages have complete annotated corpus. So we had to work with a system which requires minimal prosodic information to function. We are hopeful of getting a rich and informatics prosody with the help of above-discussed factors.

References

1. John, A., Klemm, R.P., Seligmann, D.D.: Method and apparatus for providing a virtual assistant to a communication participant. US Patent 8,107,401 (2012)
2. Shattuck-Hufnagel, S., Turk, A.E.: A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* **25**(2), 56–80 (1996)
3. Sangramsing, K., Monica, M., Jayesh, G.: Hidden Markov model based speech synthesis: a review. *Int. J. Comput. Appl.* (0975–8887) **130**(3) (2014)
4. Kakati, B.: Assamese, its Formation and Development, 5th edn. LBS Publication, Guwahati (2007)
5. Kishore, S.P., Sangal, R., Srinivas, M.: Building Hindi and Telugu voices using Festvox. In: Proceedings of International Conference on Natural Language Processing, pp. 18–21 (2002)
6. Alam, F., Nath, P.K., Khan, M.: Text to speech for Bangla language using festival. In: Proceedings of 1st International Conference on Digital Communications and Computer Applications, Irbid, Jordan (2007)
7. Rao, N., Thomas, S., Nagarajan, T., Murthy, H.A.: Text-to-speech synthesis using syllable like units. In: Proceedings of National Conference on Communication (NCC), pp. 227–280, IIT (2005)
8. Kishore, S.P., Sangal, R., Srinivas, M.: Building Hindi and Telugu voices using Festvox. In: Proceedings of International Conference on Natural Language Processing, pp. 50–85 (2002)
9. Silverman, K., Beckman, M.B., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg J.: ToBI: a standard for labeling english prosody. In: Proceedings of International Conference on Spoken Language Processing, pp. 867–870 (1992)
10. Kishore, S.P., Sangal, R., Srinivas, M.: Building Hindi and Telugu voices using Festvox. In: Proceedings of International Conference on Natural Language Processing, pp. 25–34 (2002)

Location Privacy in Wireless Sensor Networks Using Anonymity

Shriya Kesharwani and Manisha R. Dhage

Abstract A sensor network comprised of multiple devices with sensing abilities and communications technologies. Location privacy is crucial in WSN. Information on the position of the events or location of the sink node can be an initial attention of the enemy. The location privacy aims to cover the position of the sensor that senses the data and makes it tough for an attacker to discover back to the node or location. Location privacy is sorted into two categories, i.e., source node privacy and sink node privacy. Anonymous communication is a core feature of location privacy in WSN because it used to hide the identity of critical sensor nodes. The proposed location privacy schemes provide a solution for node anonymity and location privacy in WSNs.

Keywords Anonymity · Base station · Location privacy · Source node
Wireless sensor networks

1 Introduction

A sensor network consists of multiple nodes called sensors. A sensor device can sense and convert the knowledge in the form of data, prepare it, and next forward it to an aggregation point or a sink for additional practice or analysis. These sensor nodes are capable of monitor the location, sound, temperature, vibration, etc. The sensor node is very bound regarding storage space, calculating power, communication abilities, and battery power [1]. Sensor nodes will be left unattended in most of the applications until the energy of the battery is completely run down. Meanwhile, the sensor node senses the target; it will report the information to the

S. Kesharwani (✉) · M.R. Dhage
Department of Computer Engineering, Sinhgad College of Engineering,
Pune 411041, India
e-mail: sk.trpl18@gmail.com

M.R. Dhage
e-mail: mrdhage.scoe@sinhgad.edu

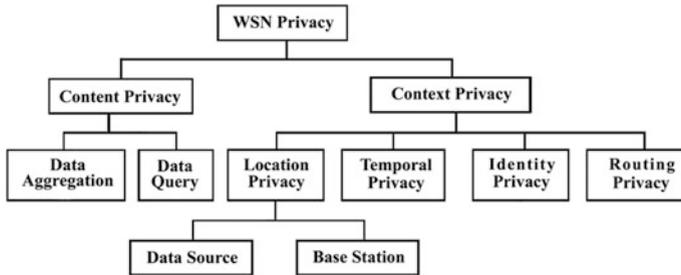


Fig. 1 Classification of WSN privacy

base station (BS) or sink node, either by single-hop or multi-hop communication. A sink is much more powerful than a sensor node regarding storage capacity, power supply, communication capabilities, and computing power.

Privacy in WSN (Fig. 1) is classified into two categories [2]. The first category is content or data privacy, which typically focuses on data. It provides for data confidentiality, integrity, freshness, and other privacy issues. The second category is context or contextual privacy [3]. Location privacy means hiding the location of a source node (SN) and the sink (base station) of the WSN. Here, location does not mean the physical location, it refers the identity of the node.

2 Related Work

Simple anonymity scheme (SAS) is an old anonymity scheme in the WSN field [4]. It uses a large number of pseudonyms ID for the identity of the SN. For anonymity, if a sensor node's genuine ID is not known to different nodes in the system network, they cannot pass judgment on it. Thus, it might be anonymous; sensor node ought to use a pseudonym to identify itself. SN with the BS and pair-wise one-hop neighbors are private and authenticated.

Misra and Xue [5] introduced two designs: (1) Simple Anonymity Scheme (SAS) and (2) Cryptographic Anonymity Scheme (CAS). SAS applies a set of the alias as ID for a node to ensure control of its true identity (ID). Later, nodes in the network shared their individual pseudonym IDs and used to guarantee that the transmission is unacknowledged. The CAS uses a cryptographic OWH (One-Way Hash) function to ensure ID mask.

Ouyang et al. [6] introduced two solutions for ID generation. The first solution is Hashing-based ID Randomization (HIR) which is a solution for CAS in case a sensor node is compromised. The SN communicates with a neighbor applying the hash value (HV) of the neighbor. The second solution is Reverse Hashing Based ID Randomization (RHIR). SN uses the HV in reverse order to interact in the similar style as in HIR. When a hash value is used, it will be dropped out.

Table 1 Comparative analysis

Scheme	Security provided	Disadvantages
SAS	Preserves node Identity and privacy	Not able to add nodes and low-level SLP
SAS + CAS	Low-level security	High computation cost and low-level SLP
HIR and RHIR	Medium-level security to ID	Cannot handle data lost and does not provide location privacy
APR	Efficient anonymity	High computation cost and storage and low-level BLP
EAC	Provide location privacy	Synchronization
FAC	Provide all context privacy	Complex method, high communication cost

Sheu [7] implemented an anonymous path routing (APR) protocol. It protects the identifications of the sensor nodes and rearranges the encryption of every packet at each hop. APR has three designs: One-hop anonymous communication, Anonymous multi-hop path routing, and Anonymous data forwarding.

Fang et al. [8] introduced EAC; An efficient anonymous communication protocol to provide anonymity in the sensor network. EAC only uses symmetric cryptography and hashing function. Before deployment of the node, it is preloaded with some parameters. EAC consists of four system designs: Anonymous data sending, Anonymous data forwarding, Anonymous broadcast, and Anonymous acknowledgment.

Abuzneid [3] present a system called FAC system for WSN. Fortified anonymous communication protocol is a measured structure that gives the source, and base station anonymity. It also provides all type of context privacy. FAC utilized a composite anonymity module where unidentified IDs are utilized instead of the real one. FAC protocol is delay-aware and energy-aware; to give protection to the BS against colluding attacks, FAC presented multilayer encryption.

A comparative study is done for the existing systems, and it is shown in Table 1.

3 Proposed System

Focusing on the monitoring and tracking application, the location of the sensor nodes is crucial. Providing location privacy and maintaining network lifetime are critical issues in WSNs. The location privacy intends to conceal the spot of the sensor nodes and makes it challenging for an enemy to locate the place. Notation used in the framework is shown in Table 2.

Table 2 Notation used in the Framework

Source	Representation	Explanation
Preloaded	ID_i	ID (identity) of the sensor node i
Preloaded	a_i	Random number shared between the node i and BS
Preloaded	b_i	Broadcasting random number for node i
Preloaded	d_i	Acknowledgment random number for node i
Preloaded	$H1$	Hashing function to create anonymous ID
Preloaded	H	Hashing function for message to create digest
Preloaded	K_{i-BS}	Pair-wise key between the node i and BS
Calculated	K_{i-CH}	Pair-wise key between the node i and CH
Calculated	K_{CH-CH}	Pair-wise key between CH
Calculated	$ANID_i$	Anonymous node ID for the node i
Calculated	$BNID_i$	Broadcast node ID for the node i
Calculated	AID_i	Acknowledgment ID for node i
Calculated	a_{i-CH}	Random number shared between the node i and CH
Calculated	Δ_i	Residual energy for the node i
Calculated	D_i	Data sensed by the node i
Operation	\parallel	Concatenation operation
Operation	\oplus	XOR operation

3.1 Pre-deployment Phase

Before genuine arrangement of the sensor nodes in the range of administration, the sensor nodes demand to be examined, enough charged, and preloaded with some parameters. The BS requires knowing the location of every one of the nodes engaging in the wireless network. Besides, all the sensor nodes must be aware of their relative positions to the BS and their acquaintances.

Preloaded parameters are:

$ANID_i$	Anonymous node ID
$BNID_i$	Broadcast node ID
AID_i	Acknowledgment ID
a_{i-CH}	Shared random no.
$K_{i-CH}, K_{CHi-CHj}$	Shared key

3.2 Setup Phase

A. Cluster Formation:

After the sensor node deployed with the preloaded parameters, each node will check if there is any node with the high energy power in its range, if yes then it will

connect to that node and select it as a CH; else, the node itself declare as a CH. After connecting to a CH, all member nodes send their information $M_{\text{discovery}}$ to the head and then CH send the packet to the BS. All cluster heads are either directly connected to the base station or through the neighbor cluster head. Base station stores all the information in the routing table.

$$M_{\text{discovery}} = \text{ID}_i \| K_{i\text{-BS}} \| a_i \| b_i \| d_i \| \Delta_i \quad (1)$$

B. *Creating Anonymous ID:*

The principal intention is to practice anonymous ID instead of real ID for the nodes during communication. Therefore, one anonymous ID per communication is handled. In this fashion, the intruder cannot locate the node using several messages holding the true ID. The method begins with creating anonymous node ID for every sensor node in the network, that is calculated using equation:

$$\text{ANID}_i = H1(\text{ID}_i \oplus a_i) \quad (2)$$

The sensor node i can determine broadcast node ID according to the equation:

$$\text{BNID}_i = H1(\text{ID}_i \oplus b_i) \quad (3)$$

Nodes will calculate the new value with neighbor node according to the equation:

$$a_{i\text{-CH}} = H1(\text{ID}_i \oplus \text{ID}_{\text{CH}}) \quad (4)$$

Between the node i and the cluster head; both sensor node and cluster head will compute a new key, using the equation:

$$K_{i\text{-CH}} = H1(K_{i\text{-BS}} \oplus K_{\text{CH-BS}}) \quad (5)$$

Two connected neighbor cluster head will determine a new key, using:

$$K_{\text{CH}_i\text{-CH}_j} = H1(K_{i\text{-BS}} \oplus K_{j\text{-BS}}) \quad (6)$$

Finally, acknowledgment ID for sensor node will be determined using this equation:

$$\text{AID}_i = H1(\text{ID}_i) \quad (7)$$

3.3 *Communication Phase*

While sensing the event and forwarding a message to the sink performed throughout the transmission phase; several transactions last as far as network lifetime expires.

A. Transmission from Sensor to Cluster Head:

When sensor node senses event, it commits to send a report to the cluster head. The SN_i simply perceives itself by its $ANID_i$, and the sink will identify the origin of the packet by $ANID_i$ as well. Hence, the $ANID_i$ of the source node needs to be carried in the information till the sink gets it. The $ANID$ of a node will be refreshed later for each transportation. The SN_i requires selecting CH to transmit the information to it. Sensor node will create the packet in the format as below:

$$M_{i-CH} = ANID_i \| E_{k(i-CH)}(AID_i \| ANID_i \| E_{K(i-BS)}(D_i) \| H_{K(i-BS)}(ANID_i \| D_i)), \quad (8)$$

where

$$D_i = seq_no \| msg_len \| sensed\ data \quad (9)$$

Formerly SN_i experiences that the information (M_{i-CH}) is transferred to the neighbor that requires settling the current anonymous node ID_i plus assign a different ID for the succeeding communication similarly marked in the equation:

$$ANID_i = H1(ANID_i \oplus a_i) \quad (10)$$

The receiver CH will then reformat the message (M) and again forward to the next CH or the BS. The base station works on the distributed pair-wise key among the nodes and the BS, to decrypt the packet and to prepare the $ANID_i$. $ANID_i$ will be used to notice the source. At this time, base station can refresh the value of $ANID_i$ of node i . Only sink node or BS can decrypt the data (D_i) applying key shared between source and BS.

B. Transmission from CH (Cluster Head) to BS (Base Station):

When SN_i transfers the packet to the CH, later CH needs to reroute the information to either BS if it is directly connected or to the neighbor CH. Simultaneous receiving M_{i-CH} , CH will tally $ANID_i$ in the routing table. If $ANID_i$ do not match, the transmission surely is not referred for CH, and the packet will be dropped instantly. Else, the received message will be decrypted using k_{i-CH} . The packet will be transmitted to next CH or BS; before the packet (M) is recreated:

$$M_{CHi-CHj} = ANID_{CHi} \| E_{k(CHi-CHj)}(AID_i \| ANID_i \| E_{K(i-BS)}(D_i) \| H_{K(i-BS)}(ANID_i \| D_i)) \quad (11)$$

$$M_{CH-BS} = ANID_{CH} \| E_{k(CH-BS)}(AID_i \| ANID_i \| E_{K(i-BS)}(D_i) \| H_{K(i-BS)}(ANID_i \| D_i)) \quad (12)$$

After sending the message when it delivered to the receiver node, sender node will update its $ANID$.

C. *Transmission of Acknowledgment:*

As assumed in the communication, the data packet could be missed or corrupted. In both cases, retransmission of the packet is required. SN changes ANID after every transmission; hence synchronization is required. The SNs should refresh the anonymous information only after assuring that the BS receives the message. Nevertheless, the need for direct contact with the source node and the BS gets little bit problematic manner. The sink cannot convey an acknowledgment to the origin node if it is not directly connected and repetitious hops apart. It has to depend on various acknowledgments on the line connecting the source and the BS. SN requires to determine the acknowledgment ID according to the equation:

$$AID_i = H1 (AID_i \oplus d_i) \quad (13)$$

The packet will be transferred to the sender with the existing state for ANID_{*i*}. Hence, it will reformat M_{i-CH} as it seems in the equation:

$$M_{CH-i} = AID_i \| E_{k(i-CH)}(ANID_i \| E_{K(i-BS)}(D_i)) \| H_{K(i-BS)}(ANID_i \| D_i) \| \text{padding} \quad (14)$$

As in the equation; padding in the message is to assure that all the packets have the identical size; which will prevent size correlation attacks.

D. *Transmission through Broadcasting:*

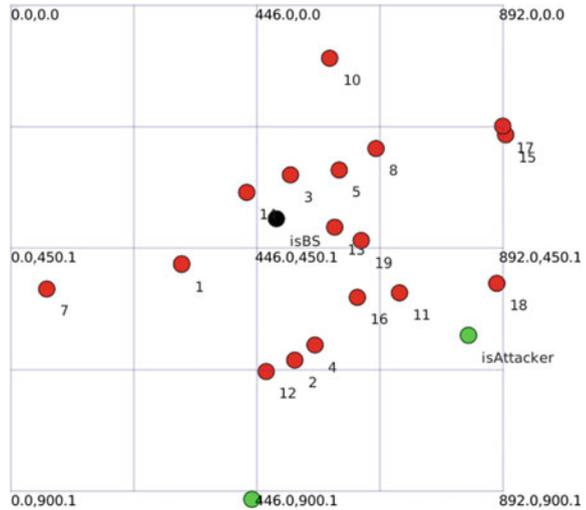
Typically, the base station is needed to broadcast few messages for control and management view. The system requires holding all the messages identical all over the network; hence the whole packet demands to have the identical measurement. Each node is preloaded with a number, i.e., broadcast random number by which anonymous broadcast node ID (BNID) is calculated. The message M sent by the CH (i.e., Broadcast message) is constructed as shown in the below equation:

$$M = BNID_i \| \Delta_i \| \text{padding} \quad (15)$$

4 Performance Evaluation

We run our experiment in network simulator version 3 (ns3) that has shown to produce realistic results. The ns3 simulator runs deployable C++ or Python code; here we use C++ code. We installed ns3.20 version on Fedora13. NetAnim is used to simulate the wireless network. The nodes are positioned by an arbitrary method, and each sensor node has a fixed broadcast range.

Fig. 2 Simulation example



Attack Simulation:

We have manually declared an attacker node (e.g., node 15, 9, and 6), and it acts as an attacker station in WSN. It may collect data packet in the network and analyze the content. As in our simulation (Fig. 2), the attacker receives few packets and tries to find the relation between the nodes; when it relates the IDs of the nodes, he can only get the large pool of the anonymous information; and no idea about the location of the node.

For example (Spoofing Attack/Eavesdropping):

Packet received at Node Id:6 and 9, which is attackers!

Received ANID: 606498310462636032 and 606337781764980736

Then attacker trying to find identity or location of the source/sink node.

Received ANIDs, which are anonymous and no relation between them;

Attacker cannot find the address in the packet!

So, the location of the node is secured.

5 Storage Evaluation

Table 3 relates the computation and storage costs of our proposed scheme with numerous existing methods. (Remark: Here we do not carry the storage cost of data or packet decryption or encryption, the operation by the proposed scheme is equivalent to the existing methods [3, 8].) The capacity of the storage may boost during the figure of the neighbor gains. The gain in storage is linear and almost similar to the different schemes. Whenever we suppose that the random numbers, the keys, the anonymous data, and the hash bit are ‘m’ bits lengthy (in average).

Table 3 Performance comparison

Method	Computation cost	Storage cost (in bits)
SAS	No H.O.	$2mM + 4mM_{ave} + 16$
CAS	Two H.O. and two encryptions	$6m + 7mM_{ave} + 16$
HIR	One H.O.	$2m + 2mM_{ave}$
APR	Six H.O.	$9m + 7mM_{ave} + 2M - 2M_{ave} - 2$
EAC	Four H.O.	$6m + 6mM_{ave} + 2$
FAC	Four H.O. and onion encryptions	$10m + (7m + 16) \times M_{ave}$
Proposed scheme	Three H.O. and onion encryptions	$7m + 7mM_{ave} + 16$

H.O.: Hashing operation, M : total sensor nodes present in the network, M_{ave} : number of neighbors for a node in average

The expected bits for some miscellaneous information entirely are two bytes, then the total storage memory required is: Storage = $7m + 7mM_{ave} + 16$. Abdel-shakour and Abuzneid [3] described the storage for other existing scheme SAS, CAS, HIR APR, EAC, and FAC.

6 Conclusion

The system provides location privacy that includes source node privacy and sink node privacy. The system model used anonymous ID, which changes for every transmission hence the identity of the nodes will secure, which leads to enhancement in location privacy. The acknowledgment and retransmission facility enhance the performance of the system model and maintain synchronization. The proposed system also provides authentication and data integrity. Clustered architecture and the two-level heterogeneous environment increase the network lifetime. Future work in the proposed system is to provide addition and removal of sensor nodes, the addition of fake messages and multi-level heterogeneity.

References

1. Manjesh, B.N., Sudarshan, T.V.: A survey on wireless sensor networks. *IJERT* **4**(4). ISSN 2278-0181 (2015)
2. Ozturk, C.: Source-location privacy in energy-constrained sensor network routing. *ACM Workshop on SASN*, pp. 88–93 (2004)
3. Abuzneid, A.-S.: Fortified anonymous communication protocol for location privacy in WSN: a modular approach. *Sensors* **15**, 5820–5864 (2015)
4. Misra, S., Guoliang, X.: SAS: a simple anonymity scheme for clustered wireless sensor networks. *IEEE ICC* **8**, 3414–3419 (2006)
5. Misra, S., Xue, G.: Efficient anonymity schemes for clustered wireless sensor networks. *Int. J. Sens. Netw.* 50–63 (2006)

6. Ouyang, Y., Le, Z., Xu, Y., Triandopoulos, N., Zhang, S.: Providing anonymity in wireless sensor networks. In: IEEE International Conference on Pervasive Services, pp. 145–148 (2007)
7. Sheu, J.: Anonymous path routing in wireless sensor networks. IEEE Communications Society in the ICC 2008, pp. 2728–2734 (2008)
8. Fang, B., Chen, J., Du, X.: An efficient anonymous communication protocol for wireless sensor networks. *Wireless Commun. Mobile Comput.* (2011)

A Compact CPW-Fed Defected Ground Microstrip Antenna for Ku Band Application

Rashmi Singh, Vinod Kumar Singh and Puneet Khanna

Abstract A simple, compact, wide band microstrip antenna for Ku band application is proposed. The defected ground and CPW feeding is used to achieve wideband antenna. The presented antenna is designed on FR-4 substrate with dielectric constant of 4.4. The overall size of antenna is $25 \times 23 \times 1.6$ mm. This Ku band antenna has been simulated using an Ansoft HFSS, electromagnetic simulation tool. The designed antenna resonates at three frequencies 11.2, 12.5, and 14.3 GHz with the return loss of -27.2191 , -34.2027 , and -35.896 , respectively. It covers a bandwidth of 5.9 GHz extending from 10.5 to 16.4 GHz. It has maximum gain of 1.6 dB. Antenna has stable radiation pattern and group delay.

Keywords Partial ground · Wide band · Ku band · CPW fed

1 Introduction

The frequency range of 12–18 Hz has been chosen as Ku band. Ku band is generally utilized for satellite communication as Ku band provides broadcasting as well as two-way communication [1–3]. Other advantages of Ku band are less signal interference from other communication systems, more reliable, secured connectivity, small antennas, and versatile spectrum.

R. Singh

Department of Electronics & Communication Engineering, S R Group of Institutions, Jhansi, UP, India
e-mail: rashmisingh1311@gmail.com

V.K. Singh (✉)

Department of Electrical Engineering, S R Group of Institutions, Jhansi, UP, India
e-mail: singhvinod34@gmail.com

P. Khanna

Department of Electronics and Communication Engineering, School of Engineering and Technology, IFTM University, Moradabad, UP, India

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_24

231

Microstrip antenna plays very significant role in wireless communication system. Microstrip antennas are used in high performance aircrafts, radar, missiles, and other spacecraft. It has many advantages such as its light weight, simple structure, ease of integration, and less cost. Microstrip antenna requires very less space for installation as these are simple and small in size. The only space these require is the space for the feed line which is placed behind the ground plane. Microstrip antennas are low profile, simple, conformable to planar, and non-planar surfaces, inexpensive to manufacture using modern printed circuit technology [4–9]. Microstrip antenna has gained much importance due to its low profile, simple design, ease of integration, and ease of fabrication.

The designed antenna is fed using CPW feeding mechanism. CPW feeding technique has many advantages over other microstrip feeding techniques. Some of the advantages are low dispersion, low radiation leakage, and the ability to control the characteristic impedance. CPW feeding mechanism can also be integrated with active devices. Microstrip antennas can be designed in various shapes and sizes. However, size of the antenna cannot exceed certain values as these need to be compact in size so that they can be easily installed on the surfaces of aircrafts, radars, or satellites. Smaller size reduces the weight of antenna which is a desirable characteristic for microstrip antenna [10–13]. Microstrip antennas with different shapes such as double U shaped, E shaped, hexagonal shaped have been reported. For Ku band application, triangular slots on patch and elliptical slot cut in the ground plane have been reported [14–16].

In this paper, CPW-fed patch and defected ground plane has been used to increase bandwidth. The designed antenna covers a bandwidth of around 5.9 GHz from 10.5 to 16.4 GHz, which covers most of the Ku band.

2 Antenna Design

The presented antenna operates for three resonant frequencies which is shown in Fig. 1. If there is more than one resonant part available with each operating at its own resonant frequency then the overlapping of multiple resonance leads to broadband applications. The basis of the proposed antenna was a rectangular patch with length L_{p2} and width W_{p2} . Then U shape was obtained by cutting a U-shaped slot from the rectangular patch and finally the proposed design was obtained to enhance the operating bandwidth. Ground plane is same for all the three designs and is made by cutting three small rectangular slots from the main rectangular ground plane.

The patch has been designed on FR-4 substrate with dielectric constant equal to 4.4 and height of the substrate is equal to 1.6 mm. The overall size is $25 \times 23 \times 1.6$ mm which is a very compact size and hence suitable for satellite applications. 50Ω characteristic impedance is achieved on fixing the feed length at 3 mm. The design parameters of anticipated antenna is shown in Table 1.

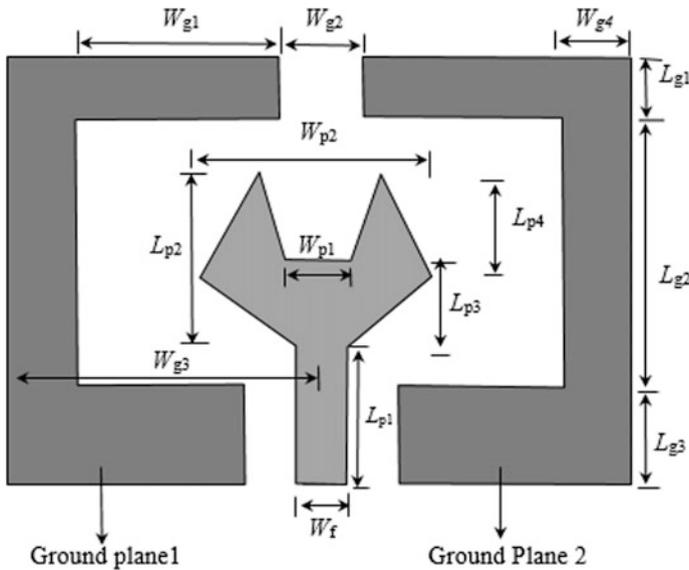


Fig. 1 Configuration of the proposed microstrip antenna

Table 1 Design parameters of the proposed CPW-fed shaped monopole antenna

Parameters	L_{p1}	L_{p2}	L_{p3}	L_{p4}	L_{g1}	L_{g2}	L_{g3}
Unit (mm)	8.5	11	5	6	2.5	15	7.5
Parameters	W_{p1}	W_{p2}	W_{g1}	W_{g2}	W_{g3}	W_{g4}	W_f
Unit (mm)	4	15	7	4	12.5	2.5	3

3 Optimization of Proposed Microstrip Antenna

The Ansoft HFSS-2010 has been used to design and optimize the results produced by the antenna. Initially rectangular patch was used to achieve the desired bandwidth, then the rectangular patch was modified to U-shaped patch and finally the proposed shaped patch was obtained to optimize the results. The return loss for rectangular, U shape, and proposed shape has been denoted by black, red, and blue colored curves respectively in Fig. 2. It can be clearly seen that there are three resonant bands at 11.2, 12.5, and 14.3 GHz, respectively.

3.1 Variation of Patch Parameters

Figure 3 shows the results of the proposed antenna with variation of length L_{p2} . It is seen that the bandwidth for the return loss less than 10 dB of the antenna remains almost constant. However, there is mismatch of impedance of radiating patch and input impedance at frequencies for $L_{p2} = 10$ and 12 mm. Thus $L_{p2} = 11$ mm has

Fig. 2 Simulated return loss against frequency for the proposed microstrip antenna, rectangular antenna, and U-shaped antenna for the same ground plane

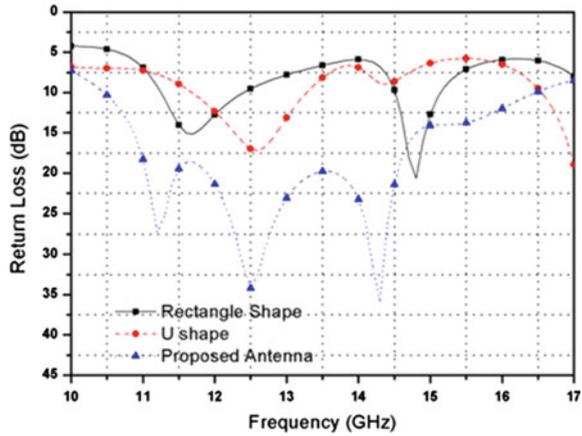
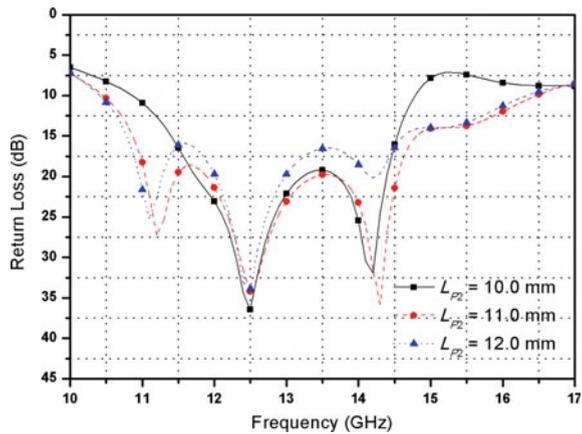


Fig. 3 Simulated return loss against frequency for the microstrip antenna with various L_{p2}



been chosen as the optimum parameter. The simulated results of the proposed antenna with L_{p3} from 4 to 6 mm have been shown in Fig. 4. It is observed bandwidth again remains constant, but better return loss has been obtained for the same resonant frequencies. Feed width was also varied from 2 to 4 mm and it was found that optimum bandwidth was obtained at $W_f = 3$ mm with more number of resonant modes which is shown in Fig. 5.

3.2 Variation of Ground Parameters

Figure 6 shows the simulated results of the proposed antenna with W_{g2} from 3 to 5 mm. It is observed that for $W_{g2} = 4$ mm the bandwidth for the return loss less than 10 dB shifts slightly toward lower frequency band. Hence $W_{g2} = 4$ mm is chosen as the optimum length of W_{g2} .

Fig. 4 Simulated return loss against frequency for proposed microstrip antenna with various L_{p3}

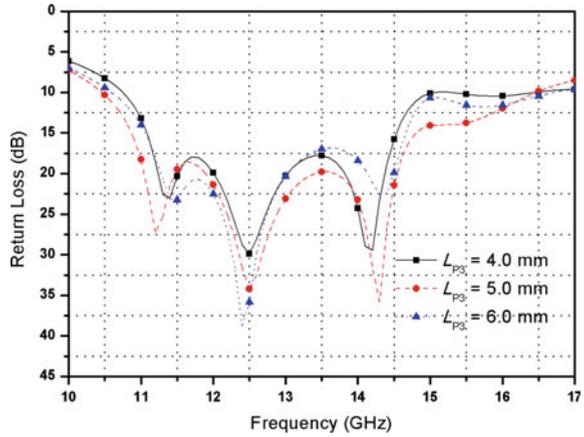


Fig. 5 Simulated return loss against frequency for proposed microstrip antenna with various W_f

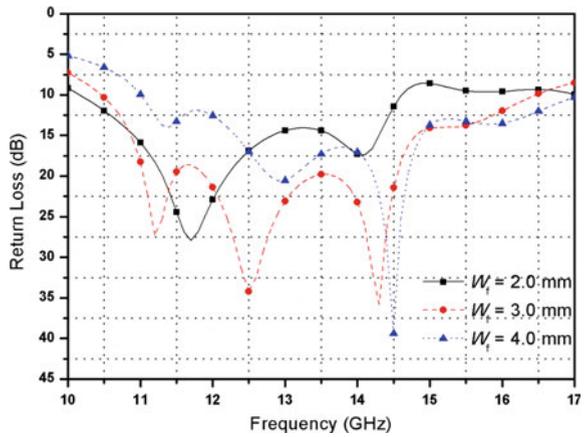
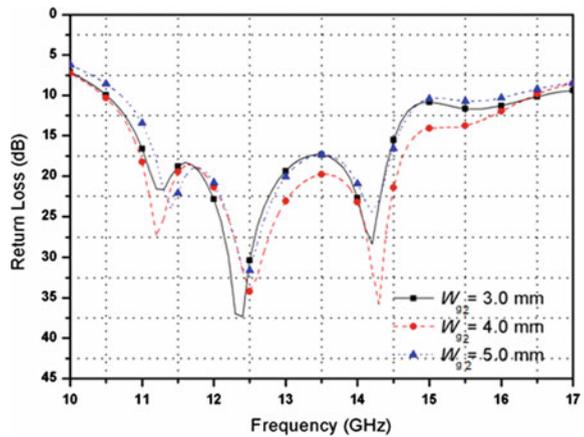


Fig. 6 Simulated return loss against frequency for proposed microstrip antenna with various W_{g2}



4 Conclusions

A CPW-fed microstrip antenna is proposed for Ku band application. The overall size ($25 \times 23 \times 1.6$ mm) is compacting thus suitable for installation on heavy satellites. The CPW feeding mechanism manages to minimize the dispersion losses as well as reduces the radiation leakage. The results are obtained using electromagnetic Ansoft HFSS-2010 solver. The antenna has maximum gain of about 1.6 dB. The antenna has stable radiation pattern and almost constant group delay over the entire 5.9 GHz bandwidth that extends from 10.5 to 16.4 GHz, thus making the antenna suitable for Ku band application.

References

1. Bhadouria, A.S., Kumar, M.: Wide Ku-band microstrip patch antenna using defected patch and ground. In: IEEE International Conference on Advances in Engineering and Technology Research (2015)
2. Gautam, A.K., Chandel, R., Kanaujia, B.K.: A CPW fed hexagonal shape monopole like UWB antenna. *Microw. Opt. Technol. Lett.* **55**(11), 2582–2587 (2013)
3. Gautam, A.K., Yadav, S., Kanaujia, B.K.: A CPW fed compact inverted L-strip UWB microstrip antenna. *Microw. Opt. Technol. Lett.* **55**(7), 1584–1589 (2013)
4. Sreenath, S., Ashkarali, P., Thomas, P., Dinesh, R., Anandan, C.K.: CPW fed compact bent monopole antenna for UWB applications. *Microw. Opt. Technol. Lett.* **55**(1), 56–58 (2013)
5. Singh, N., Singh, A.K., Singh, V.K.: Design & performance of wearable ultra wide band textile antenna for medical applications. *Microw. Opt. Technol. Lett.* **57**(7), 1553–1557 (2015)
6. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: A wide band compact microstrip antenna for GPS/DCS/PCS/WLAN applications. In: *Intelligent Computing, Networking, and Informatics*, vol. 243, pp. 1107–1113. Springer, Berlin (2014)
7. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: Bandwidth optimization of compact microstrip antenna for PCS/DCS/bluetooth application. *Central Eur. J. Eng.* **4**(3), 281–286 (2014)
8. Mandal, T., Das, S.: Ultra wideband printed hexagonal monopole antennas with WLAN band rejection. *Microw. Opt. Technol. Lett.* **54**(6), 1520–1525 (2016)
9. Gupta, N., Singh, V.K., Ali, Z., Ahirwar, J.: Stacked textile antenna for multi band application using foam substrate. In: *International Conference on Computational Modelling and Security (CMS 2016)*. *Procedia Computer Science*, vol. 85, pp. 871–877 (2016)
10. Singh, N.K., Singh, V.K., Naresh, B.: Textile antenna for microwave wireless power transmission. In: *International Conference on Computational Modelling and Security (CMS 2016)*. *Procedia Computer Science*, vol. 85, pp. 856 – 861 (2016)
11. Srivastava, R., Ayub, S., Singh, V.K.: Dual band rectangular and circular slot loaded microstrip antenna for WLAN/GPS/WiMax applications. In: *International Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 45–48 (2014)
12. Dhupkariya, S., Singh, V.K.: Textile antenna for C-band satellite communication application. *J. Telecommun. Switching Syst. Netw.* **2**(2), 20–25 (2015)
13. Singh, V.K., Singh, N.K.: Compact circular slotted microstrip antenna for wireless communication systems. *J. Microw. Eng. Technol.* **1**(1), 07–14 (2015)
14. Srilakshmi, A., Koteswararao, N.V., Srinivasarao, D.: X band printed microstrip compact antenna with slots in ground plane and patch. In: *IEEE Conference on Recent Advances in Intelligent Computational Systems*, pp. 851–855 (2011)

15. Yang, F., Zhang, X.X., Ye, X., Rahmat Samii, Y.: Wide band E-shaped patch antennas for wireless communications. *IEEE Trans. Antennas Propag.* **49**(7), 1094–1100 (2001)
16. Samsuzzaman, M., Misran, N., Islam, M.T., Mandeep, J.S.: Dual frequency new shaped microstrip patch antenna for Ku-band applications. In: *IEEE Student Conference on research and development* (2012)

A Study on Different Normalization Approaches of Word

N. Chitrapriya, Md. Ruhul Islam, Minakshi Roy and Sujala Pradhan

Abstract With the advancing of social media communication, people communicate with each other through SMS (Short Message Service), tweets, and chats messages. But the texts used in such medium are quite different from the standard text such as in limitation of character length, misspelling, and some in abbreviated form also called as Non-Standard Form (NSF) which will not be found on dictionaries. The aim of this paper is to study the different existing approaches used for normalizing such kind of texts.

Keywords Normalization · Non-standard form · Micro-blog text
SMS · SMS text · Text normalization

1 Introduction

It is becoming a trend nowadays, where people tend to post their lives happening on social networking sites, gives their opinions or interests, they give rating on products or even a recent movies released in theaters. Mostly, users' post were either written in some abbreviated way like "pliz" instead of "please", "gud" instead of "good", "M" as "I am" etc. Such style of writing a text can be referred to as a "microtext". Here, words are either misspelled or can be in some Non-Standard Form (NSF) which are written intentionally considering a phonetic or while

N. Chitrapriya · Md. Ruhul Islam (✉) · M. Roy · S. Pradhan
Department of Computer Science and Engineering,
Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Majitar, Sikkim, India
e-mail: ruhulislam786@gmail.com

N. Chitrapriya
e-mail: cpriya.work@gmail.com

M. Roy
e-mail: mr9314@gmail.com

S. Pradhan
e-mail: sujala_1207@yahoo.co.in

empathizing something like a message with texts showing more emotions as “soooooo bad...”. Moreover, formation microtext could be due to a very limited number of characters (140 characters for Tweets, and 160 characters for SMS messages). So, it is difficult to use contextual or grammatical information from these texts. Non-standard words (NSW) are different from standard words in a number of important respects. First of all, the way of pronouncing it almost different from pronouncing a normal word [1].

Text normalization can be defined as the way of converting a text into a distinct form called as canonical [2]. The traditional way of text normalizations is the normal spelling checking as its occurrence of error is much higher than any normal texts. Spelling checking models are basically based on edit distance [3, 4]. There are many algorithms and tools (like Natural Language Processing, NLP) which can generate or understand an already well-formed words, but it has been a difficult job to analyze and applied the same tools for a microtext or SMS text due to its abbreviated words, misspellings, and noises. So, to retrieve the data from a microtext or phone text, it is very much essential to normalize. The main purpose of normalizing a SMS/social media text is for making word more readable for machines as well as for a human. Once the text has been normalized, it can then use or ready for any NLP-tools [5].

Normalizing an informal text is important for various languages processing work. It is helpful in area like information retrieval, summarization, and keyword, topic, and sentiment and emotion detection. In the next section, we would be discussing some of the various existing model used for normalizing a microtext or SMS words. In the next section, we would be discussing some of the various existing model used for normalizing a microtext or SMS words.

2 SMS (Short Message Service) Text

It is a common method of sending a message through mobile phones. The SMS language can be considered as a different form of English language with some inference in grammar and vocabulary. SMS allows for text messages of 160 or 70 characters in length [6]. But the chat style texts contains spelling errors intentionally or typo errors, sometimes ungrammar, shorten of words or abbreviated forms or words self-created which are out of vocabulary (OOV) or insertions of emotions. The main reason could be the limited length of a message character. So we can categorize a SMS text into three parts as shown in the Table 1.

Table 1 Common trends of SMS text

Spelling	Ungrammar	Non-standard words
Typo error	Substitution of shorter words	Use of emotions
Deletion of whitespace between two words.	Deletion of pronouns (especially subject)	Abbreviated words
Use of number in words		Features from spoken languages

2.1 SMS Normalization

Generally, text normalization deals with Non-standard Words (NSWs) which is used in text-to-speech (TTS) [5]. Normalization a text is an important and yet the first step for any system of TTS [7]. It deals with tokens which includes digit, some abbreviations, and symbols while normalization an SMS text needs to deal with language like “n8” (night) or “clg” (college). Such texts are typically created by users and are not yet formalized in linguistics. Text message normalization basically consists of three well-known metaphors of NLP such as spelling correction, machine translation, and automatic speech recognition [5]. Normalizing system proposed by Pennell and Liu [8] used the deletion-based of abbreviation from an informal words. Reference [9] gives various way to find out the problem of SMS normalization with respect to a NLP or using other existing tools.

(a) **Spelling Correction:**

SMS text normalization has many common with traditional way of spelling correction. Conventionally, spelling correction can be subdivided into branches. It is concerned in identifying errors either a nonword or errors for a real word. Correcting errors for nonword basically concentrate on creating and grading a record of feasible spelling corrections for any word not found in a spelling dictionary. On the other hand, real-word spelling correction or a context sensitive spelling correction concerned with detecting incorrect operation of correct words in certain contexts [5].

(b) **Machine Translation (MT)**

Statistical Machine Translation (SMT) is model which produces translations based on statistical models. This model requires a lot of serious training data and they are not available very certainly. Also, it requires lot of work for creating an annotate corpus which will hide all poorly formed words and condition, with suitable solution sufficiently [10].

(c) **Automatic Speech Recognition (ASR)**

This model is used to handle the SMS normalization within the lattice of weighted phonetic sequences [10]. It is considered by far the best word sequence is discovered by this system.

3 Different Normalization Approach for SMS Text

Because of the presence of abbreviations, typing mistakes, nongrammatical, phonetic or emotion words in the SMS causes difficulties for processing using any tools. To generate more accepted form (i.e., lexical) of such messages, some approaches were used to preprocess them before it has been used [10]. Many extensive works were done on normalizing SMS texts or social text. In this paper, we have studied some of the approaches used in normalizing SMS text,

such as *phrase-based statistical* [11], *character-Level machine translation* [12], *unsupervised noisy channel model* [13] and *Character- and Phone-based Machine Translation Approach* [14].

3.1 Phrase-Based Approach

For a better affinity towards the lexical and linguistics equivalent an updated model phrase-based model was proposed by Aw et al. [11]. The idea is to normalize numerous words to numerous alignments of words, letting a series of SMS words to be normalized to a series of words in English. A novel system was proposed by Kaufmann [15], which can accurately correspond to a message from twitter media to a syntactically correct sentence of an English language. This approach used the idea of combining STM software with a preprocessor.

Data Collection and Categorization

All corpuses consist of numerous messages that are composed from two different sources, one from a SMS chatting group and other chats among students. These text behaviors are categorized into: (1) Orthographic Variation (which usage non-standard or abbreviated self-created word) and (2) Grammar Variation (which deletes articles and subject pronoun).

Normalization Phase

Let us consider e as an English sentence, SMS sentence as s . Let us assume that this sentence, e can be divided into different phrases with a T segmentation. Considering that every phrases in e sentence can match with one of the phrase in sentence s [11].

In a phrase-based normalization of SMS, it uses a channel model to calculate the maximum approximation value for the sum over all phrase segmentations T :

$$P(S_1^m | e_1^n) = \sum_A P(S_1^m, T | e_1^n) \quad (1)$$

$$\approx \max_T \left\{ P(T | e_1^n) \cdot P(\tilde{S}_1^k, T | \tilde{e}_1^k) \right\} \quad (2)$$

So, the model has got three transformations (insertion, deletion, and substitution) through the normalizations pair $(\tilde{S}_k | \tilde{e}_{ak})$, with the mapping probability $P(\tilde{S}_k | \tilde{e}_k)$. The mapping probability is estimated via relative frequencies, $N(\tilde{S}_k | \tilde{e}_k)$ as shown below.

$$P(\tilde{S}_k | \tilde{e}_k) = \frac{N(\tilde{S}_k | \tilde{e}_k)}{\sum_{\tilde{S}_k} N(\tilde{S}_k | \tilde{e}_{ak})} \quad (3)$$

The search criterion is done using a language model (bigram) and considering Bayes decision rule.

$$\tilde{e}_1^N = \arg \max_{\tilde{e}_1^N} P(e_1^N) \cdot P(S_1^M | e_1^N) \quad (4)$$

This model for normalization consists of two sub-models; first, $P(e_n | e_{n-1})$ which is a word-based language model (LM) and second, $P(\tilde{S}_k | \tilde{e}_k)$ which is a phrase-based lexical mapping model (channel model).

Training Phase

In a phrase-based model, training is done by first making the SMS sentence collection aligned at the phrase level. In this phase, an assumption is made on the alignment unit that every SMS word can be linked to a series of adjacent word of English, which is not possible in the other way round. In the training phase, an EM (Expectation-Maximization) algorithm is used to find the maximum likelihood. A search called Viterbi was used to deduce the alignment. It also uses a filtering process so as to remove manually all the low-frequency noisy alignment pairs. Bigram language model along with a lexical mapping based on phrases was used to conduct the training phase. The result after the analysis of the model shows that the efficacy lies in its capability to validate the mapping as incases like mapping the words “2” or “to” or “two” or “too” same as in “4” to “for” or “four”. While analyzing the errors, the results shows that the model lies in inserting a proper pronoun and including verb or an auxiliary giving meaningful information about the main verb which provides significant means of understanding the context.

Although, this approach was conducted on English-Chinese translation systems [16] which produces three translations output sets: (a) normalized messages with simple dictionary to look-up, (b) raw SMS and (c) normalized messages. The experiment carried out on this model, shows a good sign on the performing SMS text normalization. It is found out that using off-the-shell STM system, a very satisfactory Word-Error-Rate result can be achieved; together with this system with a system based on an analogy with the speech recognition problem gives a better result.

3.2 Unsupervised Noisy Channel Approach

This approach of text normalization was proposed by Cook and Suzanne [13]. It uses huge group of SMS text and its corresponding standard forms. The main objective of using an unsupervised system is that has ability to adapt any new word

form for normalization of text in any genres it could be in Internet discussion forums without making a big training corpus. Normalization is also important for such genres when given the need for applications like question and answering or translation. Instead of employing general error model to get all new words, rather a hybrid model was used, where each process of forming a word is modeled directly in relation to linguistic observations particular to that formation [13]. Also, text normalization can be done using unified tagging approach based on Conditional Random field proposed by Zhu et al. [17]. Different types of noises can occur in SMS words or also in any data on internet. Some of limitations which can be associated with these types like an unnecessary substitutions or mismatching a context [18]. A new error model for noisy channel was proposed by Brill and Robert [19] for correcting the spelling of a text based on generic string, i.e., automatically training a system to correct spelling errors.

Data Collection and Categorization

This approach uses data provided by Choudhury et al. [12] which consist of texting form obtain from a group of text messages. The messages then manually decide their respective standard form. Words are categorized based on the most frequently standard form, but in situation of draw, then frequency is to be chosen based on the choice. Various iterations of categorization were performed in sequence so as to calculate suitable categories. It has to make sure that all categories were used constantly. The word formation type can be categorized as Stylistic variation (exhibit non-standard spelling), Subsequent abbreviation (often omitting vowels), Prefix/suffix/H-clipping (omit just a g or h), syllabic (syll.) letter/digit (single letter or digit), Phonetic abbreviations (sound), Spelling error, Punctuation, and error (standard form which are not correct) or last category is unclear (words that did not fall within any of categories).

No texting forms data matched to multiple standard form words, e.g., “**gonna**” instead of “**going to**”. The way of texting corresponds to only one standard form of word. And it is also necessary to consider certain text forms that can have properties of more than one category which could be considered a subsequence abbreviation or stylistic variation.

Normalization

Consider S to be a sentence consisting standard forms $s_1s_2s_3\dots s_n$. Let a sequence of texting, T of form t as $t_1t_2t_3\dots t_n$ and which maybe different from the standard forms. For a given T , i.e., the texting form, the aim is to calculate its corresponding standard forms (S).

The text message normalization is model using a noisy channel given by equation: $\arg \max_S P(S|T)$ can written as $\arg \max_S P(T|S)P(S)$ which can further be extended as

$$= \arg \max_S \left(\prod_i P(t_i | s_i) \right) P(S) \quad (5)$$

$$= \arg \max_{s_i} \sum_{wf} P(t_i | e_i, wf) P(wf) P(s_i), \quad (6)$$

where $P(T|S)$ is the product of probabilities. An assumption is made considering that each text form (t_i) will depend only on sentence standard form (s_i) but not on the context in which it took place. A **word model** is given by $P(t_i | e_i, wf)$, **word formation prior** represented by $P(wf)$ and $P(s_i)$ as a **language model**.

Word Models

Words are modeled considering the formation type like stylistic variation, prefix clipping, and subsequence abbreviations due to their frequency in the text data. While g-clippings and h-clippings are moderately used, word model does not consist of any syllable, digits, or even punctuation directly. Instead, it simply replaces digits with a phoneme representation. For example, replacing the same case “to” by a digit “2”. Before applying to the model, any punctuation is to be removed.

Word Formation Prior

With respect to an unsupervised model, word formation prior represented as $P(wf)$ is calculated using uniform distribution and maximum likelihood estimates (MLEs). But this method is somewhat supervised due to its requirement in determining only those frequency which are similar to the word formation types, and not labeled texting form standard form pairs.

Language Model

A unigram language model for $P(s_i)$ was considered. Other model of language like a higher order of n-gram models could be employ simply instead of using a language model unigram.

So, this approach was to create multiple text models which will correspond to any frequently text language generation procedure and an unsupervised method for parameter estimation. The approach intended to improve word model by associating with additional linguistic knowledge and also to experiment on n -gram language models.

3.3 Character-Level Machine Translation Approach

This approach was proposed by Pennell et al. [20] which describes two methods for extending any abbreviations present as an informal text such as text messages on mobile, conversations on a chat room, etc. Training of the system is basically done at the character level which is divided into two phases. First phase method uses a

character-level MT system to generate feasible assumption for every abbreviation while the second one uses a Language Model to select premises with according to the situation and decode it.

Data collection and Annotation

Data was collected from twitter.com especially the status updates. A meta-data of Twitter's were used to collect messages sent through SMS only.

e.g., *“In moibrthdy party wdhmoi lo Lvngsista.”*

A message can give a score based on the word count index, perplexity, OOV count (out-of-vocabulary), percentages or frequency scores. A sentence will have score with a weighted average of its position in every list. More weight is given to the non-duplicate OOV percentage and less weight is given to the OOV frequency scores. The sentences are arranged in a list based on the final score calculated. Then, after that a processing step will iterate across the list by checking if any new words (OOV) were made comparison to a higher ranked sentence. Finally, those messages were annotated in the order of rank.

Normalization Phase

Normalization is done by using two-phase approach for any SMS text. Generally, the first phase will work on a character-level system machine translation mainly means to generate all the possible hypotheses for every abbreviated text. The later phase performed on a language model which is responsible to choose a hypothesis in context to the sentence used [20].

The MT system performed many-to-many mapping. The system is trained at character level where mapping is done between character to character instead of learning mappings between words and phrases.

For example, the ending “-er” is abbreviated by a single character like in “brother” as “broda”. All the characters were also mapped to the symbol like “@” for “at” or used of number in the text as “8” instead of “ate” etc. [20].

The translation system can be represented as

$$\hat{w} = \arg \max p(w|a) \tag{7}$$

$$= \arg \max p(c1(w), \dots, cn(w)) \tag{8}$$

$$\times p(c1(a), \dots, cm(a)|c1(w), \dots, cn(w)),$$

where $c_i(w)$ is a character in the English word w , a is for abbreviation, $p(c1(w), \dots, cn(w))$ is obtained using a character LM, and $p(c1(a), \dots, cm(a)|c1(w), \dots, cn(w))$ is based on the learned phrase translation table.

Training Phase

First Phase:

Training is done by picking pairs of words from the data which are already annotated. The first item in a pair is the actual token that can be or cannot be in an abbreviation form and second item will be its corresponding English word. All those tokens which were unable to translate and also those type of words which has a sound effects like arghh, ahhhhh, zzzzz, uff, eww, etc., were removed. Then from the remaining data, any punctuations were removed, that excludes emotions which user often treat as words. To facilitate character-level training, between each character a space is inserted and for any spaces replaces with an underscore symbol. At this point, the words are normalized as a single abbreviation without context.

Second-Phase:

The second-phase used a word-level language model to remove uncertainty in hypotheses when context is available [20]. A traditional noisy channel model was used for recognizing a speech or for decoding a MT. Determining the Standard English sentence, $W = w_1w_2w_3\dots w_n$ from a given informal sentence given by, $A = a_1a_2a_3\dots a_n$ can formally be represented as

$$\hat{W} = \arg \max P(W|A) \quad (9)$$

$$\begin{aligned} &= \arg \max P(W)P(A|W) \\ &\approx \arg \max \prod P(w_i|w_{i-n+1}\dots w_{i-1}) \times \prod P(a_i|w_i), \end{aligned} \quad (10)$$

where the approximation is based on the assumption that each abbreviation depends only on the corresponding word, and a word is dependent on its previous $(n - 1)$ words.

$$\hat{W} = \arg \max (\log P(w_i|w_{i-n+1}\dots w_{i-1}) + \sum \log(P(a_i|w_i))) \quad (11)$$

The abbreviation score $P(a_i|w_i)$ represents the likelihood that abbreviation a_i is derived from word w_i , and can be obtained from

$$P(a_i|w_i) \propto \frac{P(w_i|a_i)}{P(w_i)}, \quad (12)$$

where $P(w_i|a_i)$ is the Abbreviation Model (AM) from the character-level MT system, and $P(w_i)$ is from the character LM used in MT decoding.

In this approach, the result from the translation system is used as a likelihood score, without dividing by the character-level LM contribution. To train this system, a cross-validation using the annotations was setup. By removing multi-word abbreviations and also words which are already in proper form generally increases accuracy during training phase.

3.4 Character and Phone-Based Machine Translation Approach

This approach [14] is the extension of the previous approach, i.e., character-based machine translation approach by [20], which uses two-stage translation methods to get advantage of phonetic information by translating NSW (Non-Standard Words) to feasible pronunciations, which are later translated into some proper words. These words are then combined with character-based translation module as single step. This approach gives a better result compared to the previous approach by Chen and Yang discussed in Sect. 3.3.

Normalization Phase and Training Phase

First, the system **A** shown in Fig. 1 is the character-based machine translation model [20]. It is similar to MT (machine translation). For a given sequence of word this model main objective is to translate a non-standard word into character sequence.

Formally, a NSW represented by $A = a_1a_2a_3\dots a_n$ and the SW (standard word) represented by $S = s_1s_2s_3\dots s_n$. Let s_i and a_i be the characters in the words. So,

$$S = \arg \max p(S|A) \tag{13}$$

$$= \arg \max p(A|S)p(S) \tag{14}$$

$$= \arg \max p(a_1a_2a_3\dots a_n|s_1s_2s_3\dots s_n)p(s_1s_2s_3\dots s_n) \tag{15}$$

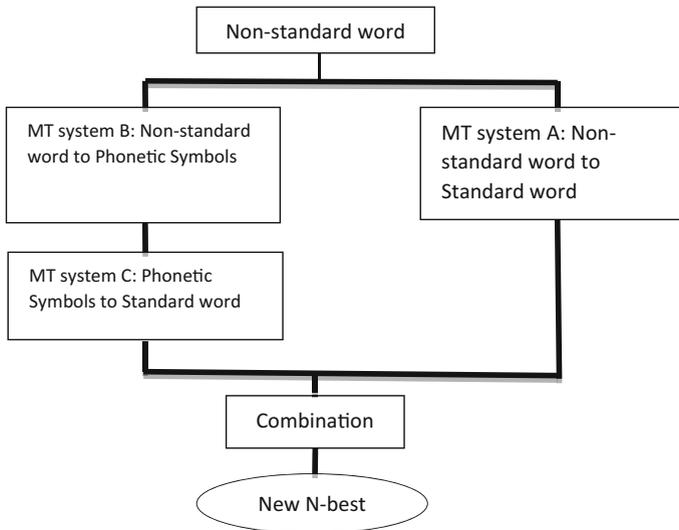


Fig. 1 Machine translation approach for normalization [14]

This $p(a_1a_2a_3\dots a_n|s_1s_2s_3\dots s_n)$ and $p(s_1s_2s_3\dots s_n)$ are the MT model and character-level model. This MT model ($a_1a_2a_3\dots a_n|s_1s_2s_3\dots s_n$) is prepared by using a corpus containing pair of standard words and a non-standard words. On the other hand, the model $p(s_1s_2s_3\dots s_n)$ representing character model is prepared by using dictionary of English language. It eliminates those candidates that are not present in the dictionary and m N -best candidates.

As, the above one-step translation model uses only the outer concept, i.e., the character information and due to its inefficiency in modeling the NSW. A new model in addition to the one-step was used called as the two-stage approach. The second model known as the two-stage approach uses a phonetic information. System *B* of Fig. 1, translates a non-standard into its probable pronunciation, i.e., a sequence of character and into a phonetic sequence; while the output of this system acts as the input of system *C* which will translate the phonetic sequence into a standard word. Also, to be noted the input of system *C* might be in dictionary itself that include the vocabulary words given by maximum confidence value. So, the two-stage phase gives the result of words in standard form by filtering the system *C* hypotheses and then finally kept the N -best candidates. The two systems, i.e., the one-step and two-step translation methods output when combined can give a better result in performance as the strength of each system can be used as complementary information. The output of these two systems will give list of N -best candidates' word, which requires a heuristic rules to combine the outputs. The candidate list of single-step module contains candidate with confidence value. All words whose confidence scores are lower than a predefined threshold were removed from the list. Then it is merged into a new list along with two-stage system. The first candidate from the one-step methods is chosen, then the another first candidate from the two-stage translation result, so on till list is empty and the final list is without redundant candidates.

4 Conclusion

This paper aims to study some of the existing SMS text normalizations. In Sect. 2, it briefly discusses what SMS means and why normalizing a text is important. Further, in Sect. 3, four different models were studied: first was a model proposed by [11] which states that any normalized SMS messages in English along with a SMT model with phrases, gives result notably in a precise translation whenever the output of a system was passed through commercialized systems of Machine Translation [21, 22]. Second, Cook and Stevenson [13] uses the concept of a model based on noisy channel and the main objective is to obtain the most possible text sequence from the observed message which is actually noisy. This approach divides any non-standard token into various categories. Third model discussed was proposed by Pennell et al. [20], whose basic idea to develop a Machine Translation system based on character level. This approach is considered to be robust to a new abbreviation. It is employed by two-phased system. First phase is a character level

trained MT model; the result is to produce word hypotheses. Second phase is a tri-gram LM for choosing a hypothesis that fits into the input context. Finally, the fourth model [14] is the extension of the character-based machine translation that combined the two systems which acts a complementary source of information for a better performance. As a future work, we aim to compare and analyze the better approach or the impact of these models in today's SMS text normalizing.

References

1. Sproat, R., Black, A., Chen, S., Kumar, S., Ostendorf, M., Richards, C.: Normalization of non-standard words. *Comput. Speech Lang.* **15**(3), 287–333 (2001)
2. https://en.wikipedia.org/wiki/Text_normalization, 9:52 PM, 15/06/16
3. Damerau, F.J.: A technique for computer detection and correction of spelling errors. *Commun. ACM* **7**, 171–176 (1964)
4. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. *Sov. Phys. Doklady* **10**, 707 (1966)
5. Xue, Z., Yin, D., Davison, B.D.: Normalizing microtext. *Analyzing Microtext. AAAI Workshop (WS-11-05)* (2011)
6. <http://www.androidauthority.com/what-is-sms-280988/>
7. Pennell, D.L., Liu, Y.: Normalization of text messages for text-to-speech. *ICASSP, 978-1-4244-4296-6/10/\$25*, IEEE (2010)
8. Pennell, D., Liu, Y.: Toward text message normalization: modeling abbreviation generation. In: *ICASSP, Prague, Czech Republic* (2011)
9. Kobus, C., et al.: Normalizing SMS: are two metaphors better than one? In: *Proceedings of the 22nd international conference on computational linguistics*, pp. 441–448, Manchester (2008)
10. Khanuja, G.S., Yadav, S.: Normalisation of SMS text. Department of Computer Science and Engineering, IIT Kanpur, India (2013)
11. Aw, A.T., Zhang, M., Xiao, J., Su, J.: A phrase-based statistical model for SMS text normalization. In: *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 3340, Sydney, Australia (2006)
12. Choudhury, M., Saraf, R., Jain, V., Mukherjee, A., Sarkar, S., Basu, A.: Investigation and modeling of the structure of texting language. *Int. J. Doc. Anal. Recogn.* **10**:157–174 (2007)
13. Cook, P., Stevenson, S.: An unsupervised model for text message normalization. In: *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pp. 71–78 (2009)
14. Li, C., Liu, Y.: Normalization of text messages using character- and phone-based machine translation approaches. Computer Science Department, The University of Texas at Dallas, Richardson, TX, USA (2012)
15. Kaufmann, M.: Syntactic normalization of twitter messages. In: *The 8th International Conference on Natural Language Processing* (2010)
16. Aw, A.T., Zhang, M., Fan, Z.Z., Yeo, P.K., Su, J.: Input normalization for an English-to-Chinese SMS translation system. *MT Summit* (2005)
17. Zhu, C., et al.: A unified tagging approach to text normalization. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 688–695, Prague, Czech Republic (2007)
18. Desai, N., Narvekar, M.: Normalization of noisy text data. In: *International Conference on Advanced Computing Technologies and Applications, ICACTA* (2015)

19. Brill, E., Moore, R.C.: An improved error model for noisy channel spelling correction. In Proceedings of ACL Hong Kong, pp. 286–293 (2000)
20. Pennell, D., Liu, Y.: A character level machine translation approach for normalization of SMS abbreviations. In: Fifth International Joint Conference on Natural Language Processing, pp. 974–982 (2011)
21. Clark, E., Arakia, K.: Text normalization in social media: progress, problems and applications for a pre-processing system of casual English. Pacific Association for Computational Linguistics (2011)
22. Raghunathan, K., Krawczyk, S., Manning, C.: SMS text normalisation using SMT. NLP Department of Stanford University, Stanford Projects (2012)

Duplicate Resource Detection in RDF Datasets Using Hadoop and MapReduce

Kumar Sharma, Ujjal Marjit and Utpal Biswas

Abstract In the Semantic Web community many approaches have been evolved for generating RDF (Resource Description Framework) resources. However, they often capture duplicate resources, that are stored without elimination. In consequence, duplicate resources reduce the data quality as well as increase unnecessary size of the dataset. We propose an approach for detecting duplicate resources in RDF datasets using Hadoop and MapReduce framework. RDF resources are compared using similarity metrics defined at resource level, RDF statement level as well as object level. The performance is evaluated with the evaluation metrics and the experimental evaluation showed the accuracy, effectiveness, and efficiency of the proposed approach.

Keywords Duplicate data · Semantic Web · RDF · Hadoop · MapReduce

1 Introduction

Duplication is the most common factor for affecting the data quality. It occupies more space than needed and ingests more time during accessing. It causes irrelevant observations, as users often depend on valuable information for drawing some conclusions and disclosing new insights. Duplicate data reduces the data quality and brings interoperability-related issues. Prior to retaining the data quality and

K. Sharma (✉) · U. Biswas

Department of Computer Science and Engineering, University of Kalyani,
Kalyani, West Bengal, India
e-mail: kumar.asom@gmail.com

U. Biswas

e-mail: utpal01in@yahoo.com

U. Marjit

Centre for Information Resource Management (CIRM), University of Kalyani,
Kalyani, West Bengal, India
e-mail: marjitujjal@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_26

253

reducing the storage area, it is necessary to detect duplicate information and to decide whether a given dataset maintains its data quality. Duplicate data detection [1] in large datasets is a very difficult job. As it requires distinct pairwise comparison of the resources in order to compare and determine the amount of duplicity. For large datasets, it takes more computations and time complexity increases as data size grows on. In order to determine duplicate information for large datasets in a very short period of time, we need a system that can perform multiple jobs in parallel. Only distributed systems have this facility and one such system is Hadoop and MapReduce framework [2–4].

Given an RDF dataset, the proposed approach splits the dataset into multiple files based on resource type. Each split files are supplied to the resource comparison job for comparing individual resources in parallel. This paper is organized as follows: Sect. 2 presents the related work and Sect. 3 discusses the background information. Section 4 presents the proposed approach and Sect. 5 shows the experimental results and finally Sect. 6 concludes the paper.

2 Related Work

Duplicate detection is the process of determining resources having similar identities. This problem is also known as co-reference resolution, identity uncertainty, record matching, and instance identification [1]. It has been vastly researched in the field of relational databases and XMLs [5–8]. In Semantic Web community, this problem is also known as near-duplicate resource detection [9], entity co-reference [10], and record linkage [11]. The record linkage matches two records situated at two different datasets based on certain criteria. This is especially used to establish relation between multiple RDF datasets. Such type of work is found in [12].

To the best of our knowledge, duplicate resource (RDF resource) detection problem has not been solved using Hadoop and MapReduce framework. However, there are few research works in the literature for performing similar kind of tasks. Near-duplicate detection presented by Ioannou et al. [9] uses similarity metrics for detecting near-duplicate RDF resources. They detect near-duplicate resources using similarity metrics at semantic level and apply blocking techniques using Locality Sensitive Hashing (LSH) to reduce pairwise resource comparison. In [13] the duplicate resource detection has been performed using similarity estimation where they detect the resources belonging to multiple data sources and different schemas. Jin et al. [14] proposes K-Radius sub-graph comparison approach for detecting duplicate data in RDF. However, these approaches still take huge amount of time to process large datasets as they perform on a single node environment. Our method is based on existing RDF dataset, generated by the third-party applications. The methodology we use is suitable for large size datasets as the resources are compared in parallel on distributed environment.

3 Concepts and Background Information

Semantic Web [15] is an emerging technology that provides a common way to share and reuse data across the web. The most common format to represent data in the Semantic Web is Resource Description Framework (RDF) [16]. RDF describes resources in the form of triple structure consisting of Subject (S), Predicate (P), and Object (O). Each such triple states that the subject and object are in some kind of relationship joined by the predicate. The subject is a resource identified by a URI, predicate is the resource's property which is also identified by a URI and the object represents the property's value which can either be a literal or another resource.

Hadoop [4] is an open-source software package designed for computing large volumes of data in parallel using commodity hardware. Required number of nodes is connected to form a distributed system and the architecture is based on the master/slave model. It consists of mainly two components: Hadoop Distributed File Systems (HDFS) for storing data and MapReduce (MR) for processing stored data. HDFS and MR are designed to run over commodity hardware.

MapReduce is the heart of Hadoop framework. It is a programming model used by Hadoop for processing and analyzing data. A MapReduce job is consists of two main procedures—*Map* and *Reduce*. *Map* takes a set of data as input and converts it into another set of data as output in the form of *key-value* pair. It reads data from an input file line by line and splits line into a set of *key-value* pairs. *Reduce* takes output from a *Map* task as input and produces another set of data in reduced form.

4 Proposed Approach

We propose a MapReduce-based duplicate resource detection in a Hadoop distributed environment. Hadoop-based environment scales well to the distributed nodes to carry out the data processing and computations. The distributed nodes along with master node perform same kind of job in parallel, where data is also distributed among them. The architecture is shown in Fig. 1.

The proposed architecture starts with an RDF dataset whose duplicity is to be computed. Generally, the RDF data can be expressed in various formats. Some of these formats are RDF/XML, N3, and N-Triples. In this work N-Triples file format has been used to provide to the Map function as input file. N-Triples is a plain text serialization data format, which represents RDF data line by line. Each line contains an RDF statement in the form of $\langle S P O \rangle$ terminating with a full stop. Since Hadoop mapper reads the input file line by line, hence data preprocessing stage converts the input file (RDF dataset) into N-Triples format so that it is acceptable to the Hadoop mapper. In the following, we describe the process of duplicate resource detection in detail.

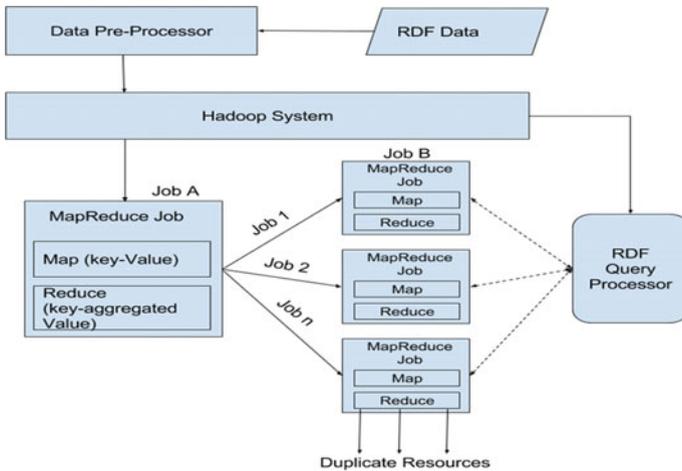


Fig. 1 Proposed architecture

4.1 Duplicate Resource Detection

The job for performing duplicate data detection begins with an RDF dataset. Here, we need two MapReduce jobs: one for processing individual clusters and other for processing RDF resources under each cluster. We call these jobs as *Job-A* (*Map-A*, *Reduce-A*) and *Job-B* (*Map-B*, *Reduce-B*). RDF resources are grouped based on their instance type using *rdf:type* attribute. Individual resources are extracted using *rdf:type* attribute. The extracted resources are placed to their respective files named by the cluster name. Hence, the data in input file will be in a format as follows: *Cluster name* followed by an arrow followed by the set of RDF resources in N-Triples format (enclosed under brackets) followed by new line character (placing whole cluster along with its data in one line). This makes the key-value combination of cluster and its data as (*Cluster Name*→(*S*→{*P O*})). This way map function of the *Job-A* reads one cluster line by line. The function *Map-A* is shown in the Algorithm 1.

Now, *Reduce-A* function needs processing individual cluster (cluster name as key part) and its value as a set of RDF resources in such a way that it can allocate a new job (*Job-B*) for processing individual resources and doing pairwise resource comparison. The *Reduce-A* dynamically creates an input file for *Job-B*, which puts a set of RDF resources (value of *Reduce-A*) and saves into a file named as cluster name. Also, we pass parent key name as the cluster name into the configuration settings of each *Job-B* so that the *Reduce-B* can determine the set of RDF data for querying. The pseudo-code of *Reduce-A* is shown in Algorithm 2.

Algorithm 1: Function Map-A

```

1: function MAP(key, value)
2:   variable cluster[] = value.split("→");
3:   variable mapKey = cluster[0];
4:   emit(mapKey, cluster[1]);
5: end function

```

The *Job-B* starts with another file (N-Triples format file). The *Map-B* (Algorithm 3) receives the line of text consisting of $\langle S P O \rangle$. Each subject is treated as key and the combination of $\langle P O \rangle$ as value. This key-value combination ($S \rightarrow \{P O\}$) is passed to the *Reduce-B* where it reduces the map part for common keys. For each key it accumulates common values—set of $\langle P O \rangle$ and then for each $\langle P O \rangle$ it queries the RDF dataset, which is obtained using parent key (cluster name) supplied from *Reduce-A*. *Reduce-B* (Algorithm 4) queries resources that contain each $\langle P O \rangle$ compares current key or resource with all collected resources and computes its similarity measure. This way the similarity measure is computed for every resource and the measure is written in the output file along with resource URI. We now briefly describe how the RDF resource similarity is measured in the following sub-sections:

5 Comparing RDF Resources

We describe similarity metrics built for RDF data. In the following, we describe the similarity function for comparing string data, statements and resources.

Algorithm 2: FunctionReduce-A

```

1: function REDUCE(key, values)
2:   variable count = 0;
3:   variable clusterData = "";
4:   for each data in values Do
5:     clusterData.append(data);
6:   end For
7:   variable file = createFileWithFileName(key);
8:   file.write(clusterData);
9: variable conf = new Configuration();
10: conf.set("ParentKey", key);
11: job = createNewJob(conf);
12: job.setInputPath(file.getPath());
13: emit(key, file.getPath());
14: end function

```

Algorithm 3: Function Map-B

```

1: function MAP(key, value, context)
2:   variable conf = context.getConfiguration();
3:   variable data[] = value.split("→");
4:   variable mapKey = data[0]; //cluster name
5:   variable value = "";
6:   value.append(data[1]);
7:   value.append(conf.get("parentKey"));
8:   emit(mapKey, value); //value = RDF resources in N-Triples format
9: end function

```

String Similarity Measure

String similarity measure uses cosine similarity metric for matching string data. The similarity between two texts or strings is defined as the cosine angle between vector representations of the two strings. It is computed as follows:

$$\text{Sim_Str}(A, B) = \cos(\theta) = A \cdot B / |A| \cdot |B| \quad (1)$$

Here, cosine similarity metric has been used because of its simplicity, efficient to evaluate, and always gives the value between 0 and 1.

Similarity Measure for RDF Statement

RDF statement measure is a measure of similarity between two RDF statements considering the similarity pairs of predicate and object. It is defined as follows.

Definition 1 Given two RDF statements S_1 and S_2 a similarity function sim_stmt , and a similarity threshold th_{stmt} , then S_1 and S_2 are similar if $\text{sim_stmt}(S_1, S_2) > = \text{th}_{\text{stmt}}$.

For the two RDF statements S_1 and S_2 , the RDF statement similarity measure is calculated as:

$$\text{Sim_Stmt}(S_1, S_2) = \text{Eq}(S_{1P}, S_{2P}) * \text{Sim_Obj}(S_{1Obj}, S_{2Obj}), \quad (2)$$

where, Eq is a function, which tells whether two predicates are equal, or not. Two predicates are said to be equal if they belong to same domain and range as well as their attribute name is same. Sim_Obj is another similarity function, which determines the similarity measure between two objects with the help of string similarity function.

Algorithm 4: Function Reduce-B

```

1: function REDUCE(key, values)
2:   variable keyValue = "";
3:   for each value in values Do
4:     keyValue = value;
5:   end for
6:   value keyPart[] = keyValue.split("#parentKey#")

```

(continued)

(continued)

```

7: variable modelPath = getInputFilePath()+keyPart[1]+ ".rdf"
8: variable masterModel = FileManager.loadModel(modelPath);
9: variable keyResource = masterModel.getResource(key);
10:variable totalRSW = 0;
11: for each r in masterModelDo
12:   If keyResource != rThen
13:     totalRSW += ResourceSimilarityWeight(keyResource, r);
14:   End If
15: end for
16:variable value = "";
17: value.append(totalRSW);
18: emit(key, value);
19: end function

```

Similarity Measure for RDF Resources

RDF resource similarity is calculated using Jaccard similarity measure, which measures the similarity between two data sets and is defined as the number of features common to all divided by the total number of features [17].

Definition 2 *Given two RDF resources R_1 and R_2 a similarity function sim_res , and a similarity threshold th_{res} , then these two resources are said to be similar if $\text{sim_res}(R_1, R_2) > = \text{th}_{\text{res}}$*

Given two RDF resources R_1 and R_2 , the resource similarity measure is calculated as

$$\text{Sim_Res}(R_1, R_2) = |R_1 \cap R_2| / |R_1 \cup R_2| \quad (3)$$

where R_1 and R_2 contains finite sets of RDF statements.

6 Performance Evaluation

The performance is evaluated with the evaluation metrics for ensuring the accuracy, effectiveness, and efficiency of the proposed approach. Java 1.8 and Jena Semantic Web framework have been used for processing RDF data. Hadoop 2.7.2 is used for setting up distributed environment having two distributed nodes on Mac OS X 10.10.2 systems each having 4 GB of RAM and Core i7 processors. RDF datasets of variant sizes have been executed using both Hadoop and non-Hadoop environments. As shown in Fig. 2, non-Hadoop approaches take more time than the Hadoop-based approaches since they require performing pairwise resource comparisons. The processing time reduces while we increase the number of slave nodes on Hadoop-based environment.

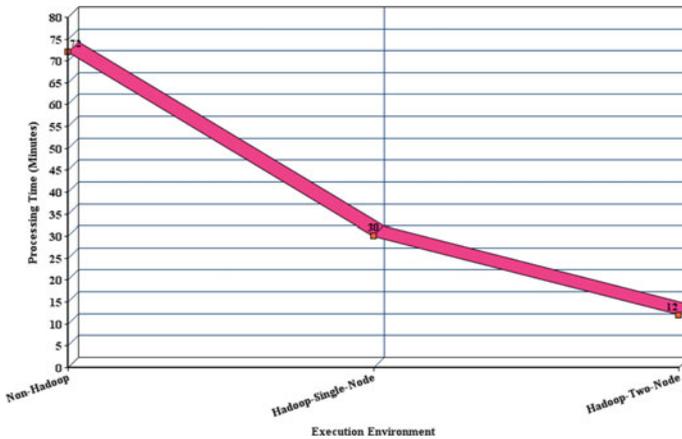


Fig. 2 Comparative analysis

7 Conclusion

In this paper, we present an approach for detecting duplicate resources in RDF datasets using Hadoop and MapReduce framework. The presence of duplicate resources gives an insight about the quality of the dataset in terms of duplicate data as well as tells about whether a dataset should undergo factorization process in order to efficiently store it. The duplicate resources help in identifying which parts of the resources of the dataset contain majority of duplicate information, thereby helping in eliminating those resources. The result shows the accuracy, effectiveness, and efficiency of the proposed system as Hadoop performs far better when size of the input dataset is very large. Because of its ability to process similar jobs in parallel over a distributed system, the time taken is very less as compared to other approaches using a single processor.

References

1. Elmagarmid, A.K., Ipeirotis, P.G., Verykios, V.S.: Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.* **19**(1), 1–16 (2007)
2. Holmes, A.: *Hadoop in practice*. Manning Publications Co. (2012)
3. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Commun. ACM* **51**, 107–113 (2008)
4. Zhou, P., Lei, J., Ye, W.: Large-scale data sets clustering based on MapReduce and Hadoop. *J. Comput. Inform. Syst.* **7**(16), 5956–5963 (2011)
5. Kelkar, B.A., Manwade, K.B., Patil, G.A.: Near duplicate detection in relational database. *Int. J. Eng. Res. Technol.* **2**(3), (2013) (ESRSA Publications)

6. Achimugu, P., Soriyan, A., Oluwagbemi, O., Ajayi, A.: Record Linkage system in a complex relational database-MINPHIS example. *Stud. Health Technol. Inform.* **160**(Pt 2), 1127–1130 (2009)
7. Weis, M., Naumann, F.: Detecting duplicate objects in XML documents. *Proceedings of the 2004 International Workshop on Information Quality in Information Systems*, pp. 10–19. ACM (2004)
8. Weis, M., Naumann, F.: Detecting duplicates in complex XML data. *Data Engineering (ICDE'06)*, IEEE, pp. 109–111 (2006)
9. Ioannou, E., Papapetrou, O., Skoutas, D., Nejdl, W.: Efficient Semantic-Aware Detection of Near Duplicate Resources. *The Semantic Web: Research and Applications*, pp. 136–150. Springer, Berlin (2010)
10. Song, D., Heflin J.: Domain-independent entity coreference in RDF graphs. *Proceedings of the 19th ACM International Conference on INFORMATION and Knowledge Management*, ACM, pp. 1821–1824 (2010)
11. Ioannou, E., Niederée, C., Nejdl, W.: Probabilistic entity linkage for heterogeneous information spaces. *Advanced Information Systems Engineering*, pp. 556–570. Springer, Berlin (2008)
12. Volz, J., Bizer, C., Gaedke, M., Kobilarov, G.: *Discovering and maintaining Links on the Web of Data*, pp. 650–665. Springer, Berlin (2009)
13. Li, M., Wang, H., Li, J., Gao, H.: Efficient Duplicate Record Detection Based on Similarity Estimation. *International Conference on Web-Age Information Management*, pp. 595–607. Springer, Berlin (2010)
14. Jin, H., Huang, L., Yuan, P.: K-radius Subgraph Comparison for RDF Data Cleansing. *International Conference on Web-Age Information Management*, pp. 309–320. Springer, Berlin (2010)
15. Yadagiri, N., Ramesh, P.: Semantic web and the libraries: An overview. *Int. J. Library Sci.* **7**(1), 80–94 (2013)
16. Faye, D.C., Curé, O., Blin, G.A.: A survey of RDF storage approaches, pp. 11–35 (2012)
17. Niwattanakul, S., Singthongchai, J., Naenudorn, E., Wanapu, S.: Using of Jaccard coefficient for keywords similarity. *Proc. Int. MultiConference Eng. Comput. Scientists* **1**, 13–15 (2013)

TEECS: A Time-Based Energy Efficient Clustering Scheme in Wireless Sensor Networks

Nirnaya Pradhan, Kalpana Sharma, Vikash Kumar Singh
and Nima Donka Tamang

Abstract The main challenge of wireless sensor networks (WSNs) is to conserve energy of sensor nodes and to increase network's lifetime. The energy efficient protocol is one of the most focused research issues in WSNs. Network layer routing protocols are proven to be more energy efficient when it uses clustering schemes. In this paper, with an objective to prolong the network's lifetime we have proposed a timer-based, distributed, energy efficient clustering scheme for wireless sensor network and we have also presented the implementation details and comparison of the TEECS protocol with existing LEACH protocol. In the results, we found out that, the TEECS protocol was more energy efficient than the LEACH protocol.

Keywords WSN · Cluster · LEACH · Timer · RSSI · TEECS · Energy efficiency

1 Introduction

WSNs comprises of self-organizing multiple nodes which are used for monitoring environmental and physical conditions like sound, fire, pressure, temperature, humidity, etc. The basic constituents of a node are sensing, data processing,

N. Pradhan (✉)

Department of Computer Science and Engineering, Advanced Technical Training Centre,
Bardang, Sikkim, India

e-mail: nirmay.pradhan@gmail.com

K. Sharma · V.K. Singh

Department of Computer Science and Engineering, Sikkim Manipal University,
Majhitar, Sikkim, India

e-mail: headcs.smit@gmail.com

V.K. Singh

e-mail: vikashsmit2009@gmail.com

N.D. Tamang

Department of Electrical and Electronics, Centre for Computers and Communication
Technology, Nandugaon, Sikkim, India

e-mail: nimadonka18@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication
and Computing*, Lecture Notes in Electrical Engineering 443,

https://doi.org/10.1007/978-981-10-4765-7_27

communication, and power unit. In most application, the sensor nodes are coupled with irreplaceable and limited energy source. Since energy plays a vital role in WSN, we must consider the energy efficient routing protocol to preserve the consumption of energy of each node and hence prolong the network's lifetime.

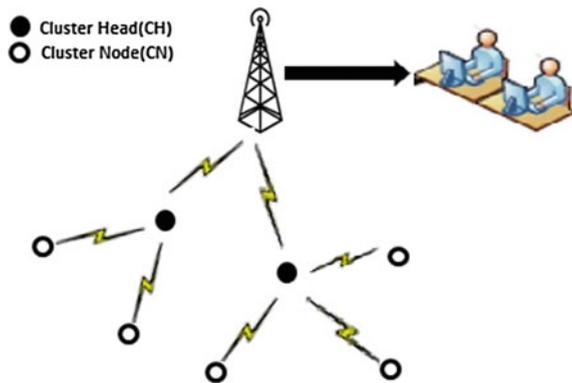
Clustering scheme is most effective means to save energy consumption and for managing enormous sensor nodes in WSN [1]. It minimizes the amount of nodes participating in the transmission of data to the Base Station (BS) [2].

The sensor nodes are grouped into distinct clusters using clustering process. One coordinator node is selected from the individual cluster known as a cluster head (CH). The non-CH sensor nodes inside each cluster act as cluster members (CMs). The CM senses its environment to collect data and send it to their corresponding CH. In this entire process, CH acts as a router. CH aggregates the received data and forwards this data to the BS using single hop communication.

In most CH selection techniques, CHs are selected randomly. In such a technique, if the communication range to the BS is more and if the selected CH has a lesser amount of energy, then CH drains its energy more quickly. Also, if the CHs are not selected properly then intra-cluster communication is not minimal and hence sensor nodes consume more energy. Therefore, this leads to imbalance energy utilization among sensor nodes in a network. The architecture of generic WSN is shown in Fig. 1.

We have proposed a new time-based energy efficient clustering scheme, called TEECS for a cluster-based WSN. Initially, the node gets deployed in the sensing regions, and then the Base Station (BS) broadcasts the hello message. After that, all the non-CH nodes calculate the signal quality from received signal strength indicator (RSSI) of the hello message. The node then calculates the time which is based on residual energy and the quality of a signal. The rest of the paper is organized as follows. In Sect. 2, review of the related work is presented. Radio WSN model and energy model used in this study in Sect. 3. The proposed approach is explained in Sect. 4. Results and analysis are in Sect. 5 and finally, the conclusion in Sect. 6.

Fig. 1 General sensor network architecture



2 Related Work

The first hierarchical cluster based protocol was proposed by Heinzelman et al. known as LEACH [2]. It uses a clustering process to group the sensor nodes, by local coordination between sensor nodes. Each sensor nodes makes an autonomous decision without any centralized control to create clusters.

In LEACH, the CHs are selected by using a random selection process. This process is carried out by rotating the CH ownership. To be elected as a cluster head all sensor nodes gates an equal opportunity. This avoids the depletion of an energy of an individual sensor node.

The working principle of LEACH is categorized into two phases. (i) Setup and (ii) Steady state phase.

In LEACH, sensor nodes are structured into the cluster in the setup phase. First, each node calculates the threshold $T(i)$. Then each sensor node i select a random number from 0 to 1. The node i becomes a CH if the threshold $T(i)$ value is greater than random number.

$$T(n) = \begin{cases} \frac{P}{1-P(r \bmod \frac{1}{P})} & \text{if } n \in G \\ 0 & \text{Otherwise} \end{cases}, \quad (1)$$

where

n number of nodes

P the desired percentage of CH

r round number, and

G is the set of sensor nodes that has become CH in the last $1/P$ rounds

In round zero, if the sensor node has become a CH then for the next $1/p$ rounds a sensor node cannot become a CH. At the initial stage (round zero) the probability P of becoming a cluster head is equal for each sensor node. Cluster heads broadcast its status using CSMA MAC protocol. Among all the received broadcast messages from the cluster heads, a non-CH nodes select the CH which has greater signal strength. The non-CH node then tells the recently selected CH about its decision by sending join message using CSMA MAC. After clusters get formed in a network, a CH calculates time slot schedule for its members using TDMA technique. This is done to indicate when a member node should be permitted to transmit its sensed data. The original data is aggregated by CH after the reception of the sensed data from its member nodes. This data holds only the meaningful information which is to be transmitted to the sink node.

In LEACH, sensor nodes transmit their sensed data to its coordinator node and CH transmits data directly to the BS. In intra-cluster communication, sensor nodes dissipate their energy slowly, and CH dissipates their energy fast when they are located far away from the BS. Selecting a CH randomly is the main drawback of LEACH. A better approach to wisely select a CH is by considering the energy level of a sensor node and its location information.

LEACH-C [6] (LEACH Centralized) is application specific protocol proposed by Heinzelman et al. The centralized clustering algorithm is used to improve cluster formation in WSN. At the start of each round, BS collects energy level and location information from all sensor nodes deployed to perform the specific task. To collect location information we require GPS, the main drawback of GPS is it consumes more energy.

TB-LEACH improves the cluster partition by modifying the CH selection algorithm of LEACH [8]. In TB-LEACH a random time interval is assigned to each node. The node which has small time interval expires first and that becomes a cluster head. The counter is set for each node to maintain a constant number of CH. The limitation of TB-LEACH is that it assigns random time interval for each node without checking its remaining energy level and the location information. Due to this limitation, the selected CHs are not optimal.

EECS (Energy Efficient Clustering Scheme) [11] balances the load among the CH. It starts with CH selection phase where each sensor node competes for the CH, based on their residual energy. EECS maintains a uniform distribution of CHs over the network and decreases intra-cluster communication distance. But, EECS exchange large number of messages between its neighbors which introduces extra communication overhead.

3 WSN Model and Assumption

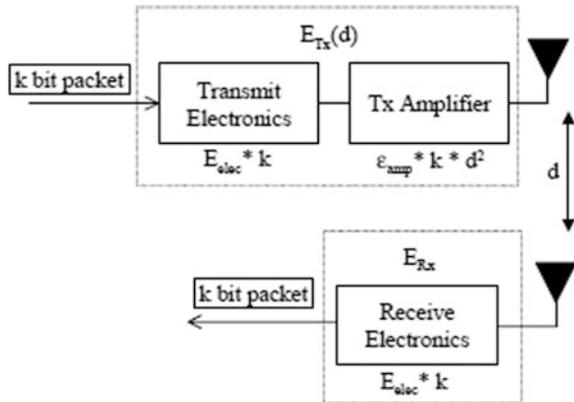
In this model, we assume that all sensor nodes are homogenous and are deployed randomly in the network field for the execution of the specific task. After deployment all sensor nodes and base station becomes static. Initially, we assume that equal amount of energy is provided to all the sensor nodes. Another assumption is that for intra-cluster communication power level of transmitting sensor node can be adjusted dynamically. In order to save more energy node's transceiver can go into sleep mode.

3.1 Energy Model

For studying the energy consumption analysis the first-order wireless model [2] has been used. We assume that for each sensor nodes energy depletion is due to transmission and reception of data. To compute the energy consumption of a message of size k bit over a distance d is given by the formula in Eq. 2.

$$E_{Tx}(d) = (\epsilon d^\alpha + E_{elec})k, \quad (2)$$

Fig. 2 Energy consumption model



where E_{elec} represent the energy consumed by the electronic circuit, $\epsilon \in \{\epsilon_{fs}\}$ is the transmitter amplifier in the free space, and α is the path loss exponent, $2 \leq \alpha \leq 4$. The consumption of the energy by the receiver for reception of the message is given by

$$E_{Rx} = kE_{elec} \tag{3}$$

Therefore, the overall energy consumed by the sensor node while receiving and forwarding a message over a distance d is given by Eq. 4.

$$E_{tot}(d) = (\epsilon d^\alpha + 2E_{elec})k \tag{4}$$

The energy consumption model is shown on Fig. 2.

4 Proposed Approach

In TEECS, sensor network is partitioned into distinct clusters and one coordinator is selected from each cluster known as Cluster Head. Single-hop communication is used to communicate between the CH and the sink node. After deployment, sensor nodes enter into receiving mode by turning its radio ON in order to receive the message transmitted by the BS. The BS then broadcasts hello message by using maximum power level. After receiving the hello message sensor nodes compute the signal quality using RSSI. By doing this each node dynamically sets the appropriate power level to transfer data with the BS. A novel approach is used in the selection process of CH, whereas the cluster construction phase is similar to the LEACH. Detailed description of these two phases is in the following subsections.

4.1 Cluster Head Selection

In cluster head selection process, individual timer is computed for each node before the campaign for CH. Timer for the individual node is calculated by using the following formula.

$$\text{timer}(i) = \frac{E_m(i) - E_r(i)}{E_m(i)} \times \frac{1}{q}, \quad (5)$$

where q is the signal quality, $E_r(i)$ is the remaining energy level and $E_m(i)$ is the initial energy level of the sensor node i . A sensor node having more signal quality and high residual energy expires first since it generates a shorter time interval. When the timer expires, sensor node i is declared as a CH and it broadcasts advertisement message. After receiving the advertisement message, receiver cancels its timer to withdraw CH candidature and hence becomes the non-cluster head for that round.

4.2 Cluster Formation

For cluster formation, each non-CH nodes will sort the neighbor CH list using RSSI values. The non-cluster node sends join request to the CH which has maximum RSSI value. It then computes the required power level to communicate with the CH. By doing this non-cluster node can conserve the energy while transmitting the data to the CH (Table 1).

Table 1 Parameter used in simulation

Sl. No.	Parameters	Values
1	Routing protocol	LEACH, TEECS
2	Network size	50×50 m
3	Data packet size	500 bits
4	Initial energy per node	10 J
5	Tx and Rx energy	50 nJ/bit
6	Amplifier energy (ϵ_{fs})	10 pJ/bit/m ²
7	Amplifier energy (ϵ_{mp})	0.0013 pJ/bit/m ⁴
8	Control message size	100 bits
9	Simulation time limit	300 s

Algorithm: Cluster Head selection algorithm

```

    /*Cluster Head Selection*/
1.  if (sink node) then
2.      Broadcast (hello message) in the range Maximum;
3.  end if
4.  if(!sink node) then
5.      for each node i
6.          if( node i receives Hello message) then
7.              calculate RSS;
8.              calculate required power level to communicate with Base Station;
9.              calculate signal_quality
10.              $T(i) = ((E_m(i) - E_r(i)) / E_m(i)) \times (1 / \text{signal\_quality})$ 
11.         end if
12.     end for
13.     if (t(i) == 0) then
14.         node i broadcast(CH advertisement) in the range Maximum;
15.     end if
16.     if (node j receives CH advertisement message) then
17.         calculate RSS;
18.         calculate required power level to communicate with CH;
19.         node j cancel its CH candidature and becomes non CH node;
20.         update Neighbor CH(j);
21.     end if
22.     if (node j is non CH node and receives a CH advertisement message) then
23.         calculate RSS;
24.         update Neighbor CH(j);
25.     end if
    /*Cluster Formation*/
26.     for each non CH node j
27.         sort neighbor CH(j) base on received signal strength;
28.         select CH having largest received signal strength;
29.         calculate required power level to communicate with CH;
30.         Adjust power level to communicate with Cluster Head;
31.         node j send join message to CH;
32.     end for
33. end if

```

5 Results and Analysis

We deploy 50 sensor nodes using random uniform distribution into the network of size 50×50 m. For evaluating the performance of TEECS we use Castalia simulator version 3.2 to simulate TEECS and LEACH.

Figure 3 shows the comparison of energy consumed by TEECS and LEACH. Here we can see as the time increases the performance of TEECS becomes better, i.e., it consumes less energy as compared to the LEACH.

Figure 4 shows the initial stage of LEACH and TEECS. Initially, almost equal amount of energy was consumed by all the nodes but as the simulation time increases the energy consumption varies between these two schemes.

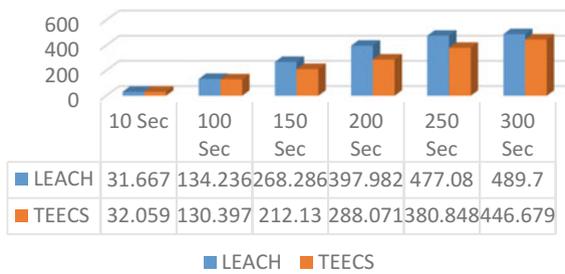


Fig. 3 Energy consumed by LEACH and TEECS

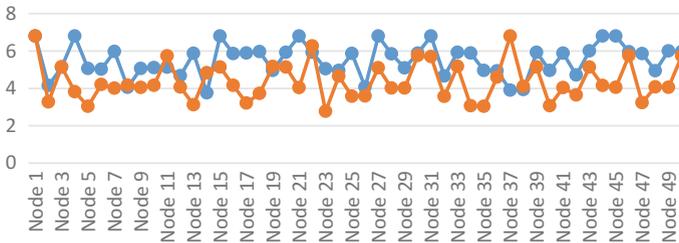


Fig. 4 Energy consumed by each node using LEACH and TEECS (simulation time 10 s)

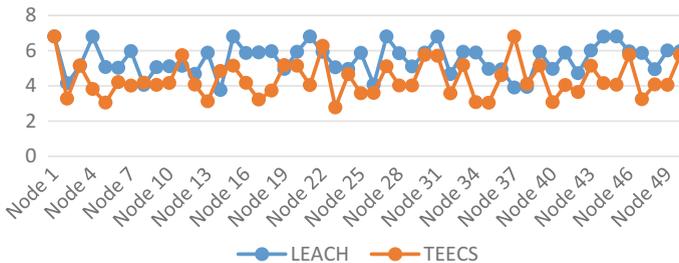


Fig. 5 Energy consumed by each node using LEACH and TEECS (simulation time 150 s)

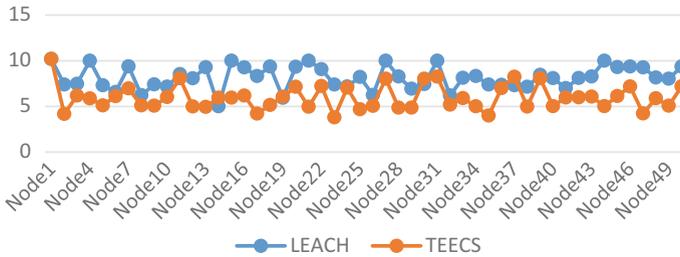


Fig. 6 Energy consumed by each node using LEACH and TEECS (simulation time 200 s)

The result of the energy consumed by each node captured at 150 s is shown in Fig. 5. The nodes which use TEECS consume less energy as compared to LEACH.

The result of the energy consumed by each node captured at 200 s is shown in Fig. 6. As compared to the result captured at 150 s the TEECS outperforms at 200 s, i.e., it becomes more energy efficient. Therefore from the above analysis, we can conclude that as the simulation time increases the performance of TEECS increases.

6 Conclusion

The main challenge in the development of WSN is the conservation of energy of a sensor node. In this paper, a novel technique is presented towards time-based energy efficient clustering for WSN called TEECS. It selects a node with more energy level and lesser distance to the BS. It also distributes the CH uniformly across the network field while selecting a cluster head. The transmitting power level is selected dynamically based on the RSSI values. Hence, this process conserves more energy, i.e., it is more energy efficient. Our work mainly focuses on the cluster head selection process and setup stage.

If the network field increases then multi-hop communication can be used for more energy saving. So, we can adopt hierarchical multilayer clustering techniques to implement multi-hop communication in future works.

References

1. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* **30**(14), 2826–2841 (2007)
2. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of 33rd Hawaii International Conference System Science* (2000)

3. Heinzelman, W.B.: Application-specific protocol architectures for wireless network. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2000)
4. Junping, H., Yuhui, J., Liang D.: A time-based cluster-head selection algorithm for LEACH. IEEE Symposium on Computers and Communications, pp. 1172–1176 (2008)
5. Ye, M., Li, C., Chen, G., Wu, J.: EECS: an energy efficient clustering scheme in wireless sensor networks. IEEE International Performance Computing and Communication Conference, pp. 535–40 (2005)

Packet Sniffing and Network Traffic Analysis Using TCP—A New Approach

Aishwarya Bhandari, Samala Gautam, Tawal K. Koirala
and Md. Ruhul Islam

Abstract Due to the rapid development of Internet the size of computer network are continue to grow in size. Thus the network traffic are increasing drastically. To keep the data transfer smooth and faster in the network the monitoring, and management of network traffic is an important factor in field of computer network. The data, information are communicated via TCP or UDP Internet protocol between nodes in the network. These transport layer protocol carry the packet or datagram from source-to-destination site. The packet analysis or packet sniffing is process of capturing the packet and analyze the log traffic passes over the network or a part of network. Packet sniffer is used for wired or wireless network. A tool is developed to accomplish the monitoring task that removes deficiency of existing tool. The packet sniffer is used to capture the traffic and analyzed those captured traffic. The report is generated based on analyzed captured traffic. Various transport layer protocol like TCP, UDP, etc., can be implemented for analysis and filtering can be done based on protocol used. A new approach of packet sniffing and network traffic analysis over TCP connection oriented network using Wireshark software has been discussed in this paper.

Keywords Packet sniffing · Packet capture · Wireshark · Network · TCP

A. Bhandari (✉) · S. Gautam (✉) · T.K. Koirala (✉) · Md. Ruhul Islam (✉)
Department of Computer Science and Engineering, Sikkim Manipal Institute
of Technology, Sikkim Manipal University, Majhitar, India
e-mail: bhandariaishwarya@gmail.com

S. Gautam
e-mail: sam1109fb@gmail.com

T.K. Koirala
e-mail: koirala01@gmail.com

Md. Ruhul Islam
e-mail: ruhulislam786@gmail.com

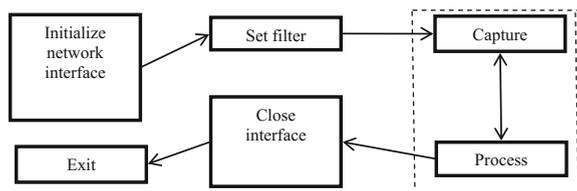
1 Introduction

In age of computer internet is gaining popularity worldwide. The data communication take place via TCP, UDP (Transport layer protocol) through packet. TCP is connection oriented and UDP is connectionless protocol [1]. As the size of computer network increasing day by day, the network traffic also equally increasing. To monitor the network traffic Packet Sniffer is used to diagnose the network related problem. Wireshark is a open-source software for network packet analyzer that is available freely through Internet [2]. It has many application field such as education, troubleshoot network, software and communication protocol development, network packet analysis, etc. [3, 4]. To analyze traffic in the morning, afternoon, and evening sessions using Wireshark for transport layer protocols (TCP) is used in this paper. Analysis is done over the different TCP parameters such as throughput, round Trip Time, Time sequence graph. To find out the cause of varying traffic in the different time slots.

2 Network Traffic Analysis—A New Approach

Broadcast storm is one of most serious network problem faced by network. The broadcast storm occurs due to improper network configuration, poor-quality technology used or by defective network device used in implementation for these reason the traffic generated in network floods like storm. These storm increases in the network and try to deactivate the network completely which destroys network activity. Malware threat is another common infection in network that aims to send a barrage of e-mail or duplicate the computers on the LAN or over the Internet. A corrupted computer can slow down the activity of Internet traffic, also the computer is on bad situation with your ISP. In certain point of time, a single user can occupy and use so much bandwidth such that other network user affected extremely with their work. So to find the cause of the problem we analyze the network traffic using Wireshark and compare the traffic and find out what is the reason for these problems in a network [6-8] (Fig. 1).

Fig. 1 How the data will be filtered and captured



3 Proposed Solution Strategy

The above figure shows us how the data is being captured and filtered. First while installing this software the data NIC will be automatically initialized. While receiving the packet by NIC it first check the MAC address of the packet with its own, if the MAC address found as similar it accept the packet otherwise filter the packet. The set filter in the above figure denotes the types of filter for different protocols that are through from the NIC. Then after the protocols are captured and the captured packets will be processed which means only the selected packet will be shown. Then we can determine which protocols should be filtered and to analyze. The protocols that are being analyzed are as follows: (Fig. 2).

- I. The first analysis is on TCP
 - TCP time sequence graph
 - TCP Throughput graph
 - TCP round trip time graph.

4 Design Strategy for the Solution

The above figure shows the existing network topology of SMIT for academic block (A, B, C, D, E). The model that is being followed in this network is hierarchical network model. To minimize the network cost hierarchical model is very helpful as the appropriate networking parts or devices of each layer can be easily purchase either online or from market. The specialty of hierarchical design model is its modularity that enables you to effectively and accurately design network capacity within each hierarchy layer. Thus, wasted bandwidth within the network can be

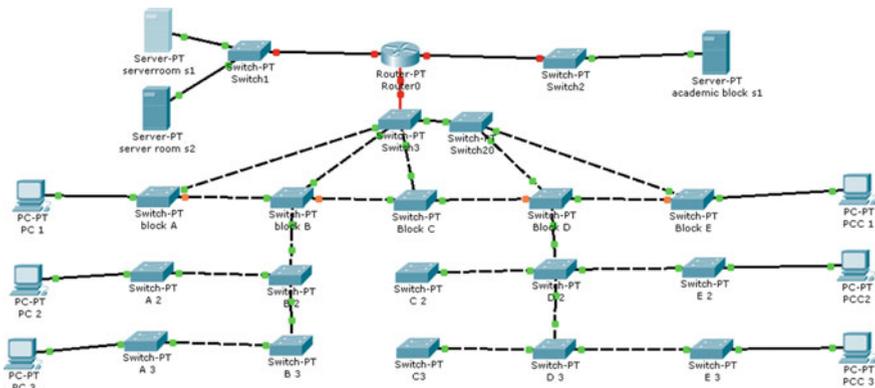


Fig. 2 Existing network topology of SMIT for academic blocks (A, B, C, D, E)

easily reduced. The great advantage of using hierarchical design is that it mitigate changes in the network environment. SMIT network is interconnected to each other via switches which uses hierarchical topology. The above figure shows the network of the academic block where all the systems are given their own IP addresses.

5 Implementation and Result

To capture and analyze the packet the program that has been used is Wireshark. The main responsibility of sniffer program is to make network interface card on the machine for entering in promiscuous mode. The NIC within the LAN Ethernet is configured with a “filter” such that it rejects all traffic that does not belong to it. As NIC first check the MAC address of every incoming packet, whenever a packet or frames whose destination MAC address mismatches with its own then such packets or frames are overlooked by NIC and ignores all such frames. A sniffer turns off the filter through NIC driver thus sniffer puts the NIC into promiscuous mode. The PCs and workstations used the NICs in promiscuous mode to analyze network traffic. Nowadays, it is possible to modify the MAC address on many NICs; so a proper equipment required to observe all the traffic ongoing in the network and thus be promiscuous. The packet sniffer has two basic components, packet analyzer, and packet capture (pcap). The received packet from physical layer, data link layer, IP, and transport layer are captured through pcap and application layer packets are analyzed by packet analyzer part. Packet analyzer often interacts with the pcap to further captures various packets from applications running on the network. Through the filter packets are rejected or accepted. Continuously the log is maintained for accepted packets and processes the captured packets.

5.1 TCP-Based Analysis

Out of many Internet protocol TCP is core protocol in IP suite. TCP is connection oriented protocol which means TCP establishes connection using three way handshake, and also has error and flow control mechanisms. TCP is stream-oriented, reliable protocol that provide ordered and error check delivery of stream of bytes between various applications that is running on host which are communicating over an IP network. TCP uses application such as www, e-mail, remote administration, file transfer, etc., the graph below shows the TCP broadcast over SMIT network (Fig. 3).

The above figure shows the flow of TCP. The red bar denotes the flow of TCP in the graph. The graph is of 1 h, i.e., 3500 s. From this graph, we learn that the TCP packets are used more from 2500 to 3000 s.

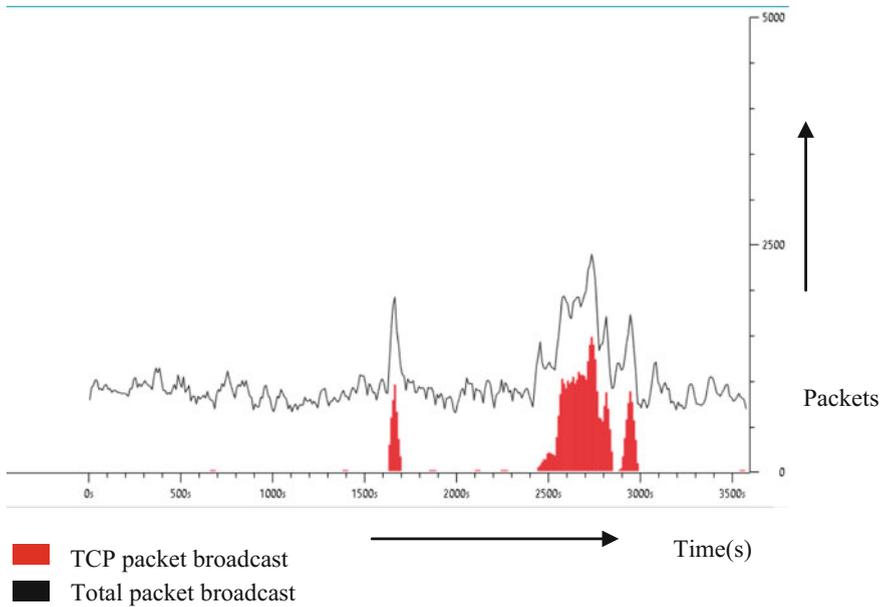


Fig. 3 TCP stream

5.2 TCP Time Sequence Graph

TCP sequence numbers over time is counted and represented by a simple graph, known as TCP Time Sequence Graph. TCP sent the bytes and these are counted as TCP sequence numbers. These counted bytes are application bytes that includes application headers sent from one side to another. The application’s behavior is identified through these graph (Fig. 4).

Now in the above figure due to fewer amounts of data, the line is diagonal with low gradient and with some interrupts in between so this indicated that this data transfer rate is low and there is some problem in the transfer. A diagonal line in time sequence graph is an indication of a good file transfer, but an interrupt diagonal line shows a problem in transfer. Fast data transfer is represented by a diagonal line with a high gradient but low rate of transfer is found by a low gradient diagonal line.

5.3 TCP Throughput Graph

The TCP throughput graph represents the TCP sequence numbers over time. The TCP sent the bytes that are actually the application’s data and counted over time. The throughput graph tell us the application throughput in bytes per second (Fig. 5).

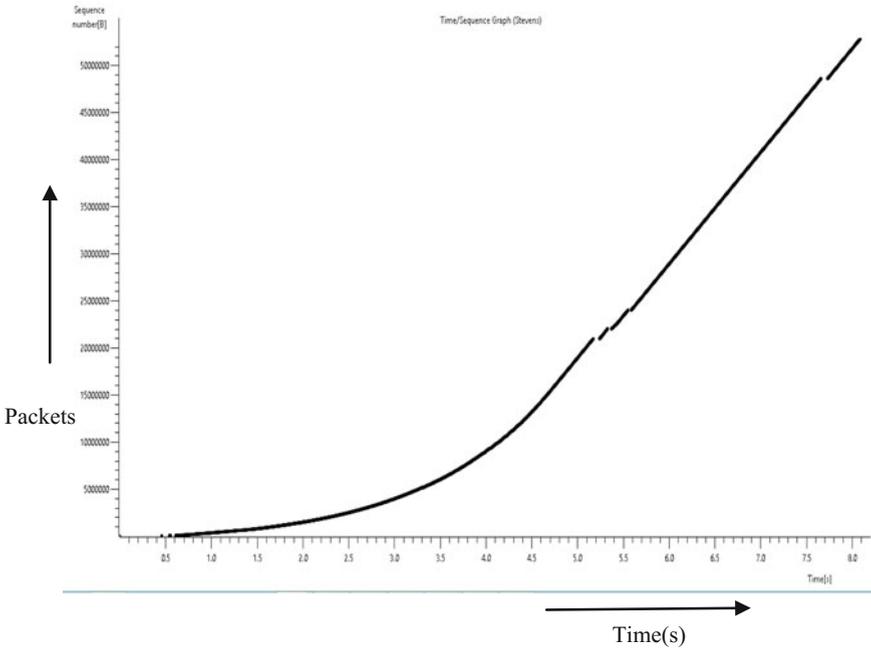


Fig. 4 TCP time sequence graph

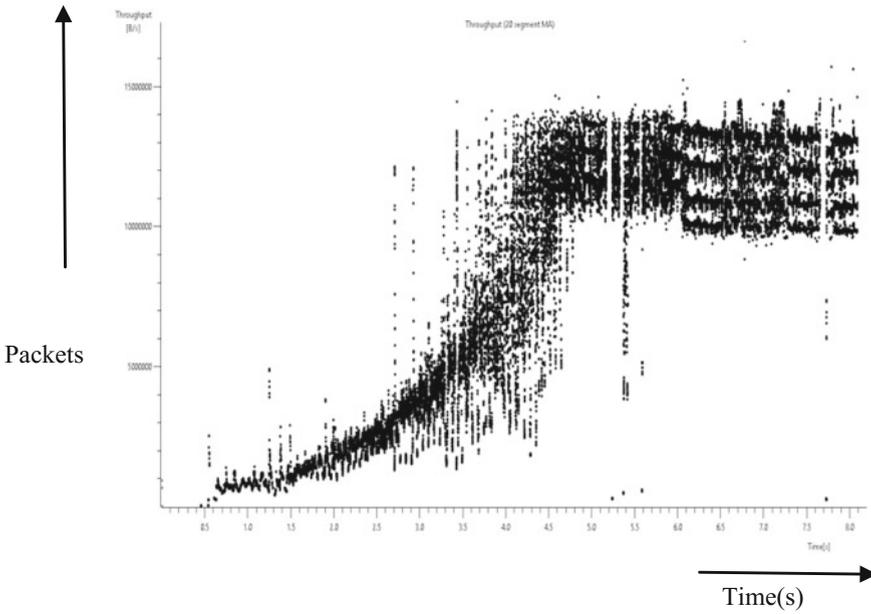


Fig. 5 TCP throughput graph

The result shows that the throughput is unstable. The throughput varies between 200,000 and 1,500,000 bytes/s. Such variation occurs due to an unstable file transfer. A straight line graph means a stable file transfer.

5.4 Roundtrip Time Graph

The TCP Round Trip Time graph represents the round trip between sequence numbers and the time they were acknowledged. The performance of the connection can be visible along with other graphs (Fig. 6).

In TCP throughput graph, it is observed that most of the sequence numbers were acknowledged in a short time till 1,500,000, after that the dots that is far and which are less shows instability in that portion which will influence the TCP performance. The result of graph summarizes as the TCP sequence number versus the time taken to acknowledge the bytes is plotted. The time indicates the time period between a sent packet and the ACK received against that packet. The instabilities in a graph do not necessarily mean a problem. This may be a representation of how the application works. An example may be, when a packet is sent it take some time to acknowledge it because either there is a problem, or may be a server is waiting for a response.

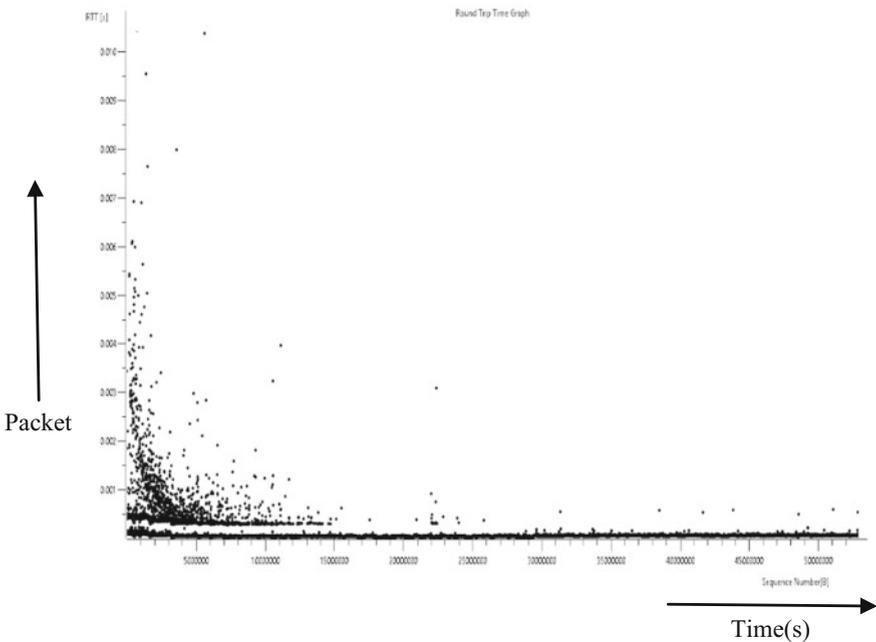


Fig. 6 Roundtrip time

6 Conclusion

In this paper, we discuss about packet analysis and network traffic monitoring over TCP protocols. We also analyzed the TCP time sequence graph, TCP Throughput graph, TCP round trip time graph. As TCP before communication always establishes connection between source and destination through three way handshake and receiver send acknowledgement to the sender after receiving of every packet, the connection is said reliable. Packet sniffer is very useful tool for network traffic analysis, network traffic monitoring, troubleshoot the network and many other purposes. Sensitive user information such as username, password can be captured through packet sniffer. Wireshark captured traffic from SMIT network. Packet sniffing is very useful tool to analyze network traffic over TCP protocol.

References

1. Kumar, M., Yadav, R.: TCP&UDP packets analysis using wireshark. *Int. J. Sci. Eng. Technol. Res.* **4**(7), 20–24 (2015)
2. Ansari, S., Rajeev, S.G. Chandrasekhar, H.S.: Packet sniffing brief introduction. *IEEE Potentials* **21**(5), 17–19 (2003)
3. Qadeer, M.A., Zahid, M., Iqbal, A., Siddiqui, M.R.: Network Traffic analysis and intrusion detection using packet sniffer. *ICCSN '10 Second International Conference*, pp. 313–317 (2010)
4. Dabir, A., Matrawy, A.: Bottleneck analysis of traffic monitoring using wireshark. *4th International Conference on Innovations in Information Technology, IEEE Innovations'07*, 18–20, pp. 158–162 (2007)
5. <http://www.sectools.org>. Accessed on 7 Aug 2016
6. https://www.wireshark.org/docs/wsug_html_chunked/ChapterIntroduction.html. Accessed on 7 Aug 2016
7. <https://www.wireshark.org/download.html>. Accessed on 7 Aug 2016

Passive Techniques of Digital Image Forgery Detection: Developments and Challenges

Santoshini Panda and Minati Mishra

Abstract Photographs and images play an important role in our lives but, in this technology era, equipped with powerful, low cost, and easy to use photo editing tools, people often forge photographs. This practice has posed a question mark on the trustworthiness of images necessitating separation of original images from the tampered lot. Because carefully edited and forged images are very hard to be distinguished from their genuine copies therefore, forgery detection and separation of the forged images from the innocent and genuine ones has become a challenging issue for image analysts. Image forgery detection procedures are generally classified into two broad categories; the active and the passive detection techniques. This paper presents a state of the art review of different passive forgery detection techniques those are proposed by different authors over time.

Keywords Copy–move forgery · Cloning · Splicing · Watermark

1 Introduction

Today, photo editing has become a common practice with the availability of free and easy to use image editing tools. Though, all edited images do not fall into forged category, some of those do. When forged images are available in huge numbers, it becomes very much important to detect and classify the genuine and the forged ones. Image forgery detection and classification can be done using active and/ or the passive detection methods. The active methods, also known as image authentication techniques, are generally based on digital signatures or watermarks. In these techniques, for verification of authenticity of an image, a watermark or a

S. Panda · M. Mishra (✉)

P. G. Department of Information & Communication Technology,
Fakir Mohan University, Balasore, Odisha, India
e-mail: minatiminu@gmail.com

S. Panda

e-mail: santoshinipanda11@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_29

digital signature is embedded into an image at the time of capture or immediately after the image is captured. This is the major drawback of the active authentication technique first, because it is not possible to embed watermarks/signatures to every captured image and second, the degradation in image quality that will be caused due to this embedding may not be acceptable by the users [1]. Passive forgery detection is an alternative to active authentication which requires no active information for the purpose of authentication. These techniques detect forgery analyzing the image statistics in the absence of any watermark/digital signature as well as in the absence of the original image for comparison.

1.1 Image Manipulation Techniques

Though all image manipulation or image editing procedures do not fall into the image forgery category, it is certain that all forged images undergo some sort of manipulation. The corrective manipulation techniques or the image enhancement techniques such as contrast and brightness adjustments, noise reduction techniques, etc., may not be included into forgery categories unless otherwise the manipulations fake some facts or make some changes to the image contents so as to convey some misleading information to the viewers [2]. According to authors of [1], image manipulation techniques can be divided into two categories: content preserving and content altering. Each of these techniques is further divided into different categories as shown in Fig. 1. Authors of [2] argue that steganography can also be classified as an image manipulation technique as that alters the image content invisibly. The subsequent sections of the paper discuss various types of image forgery and different forgery detection techniques.

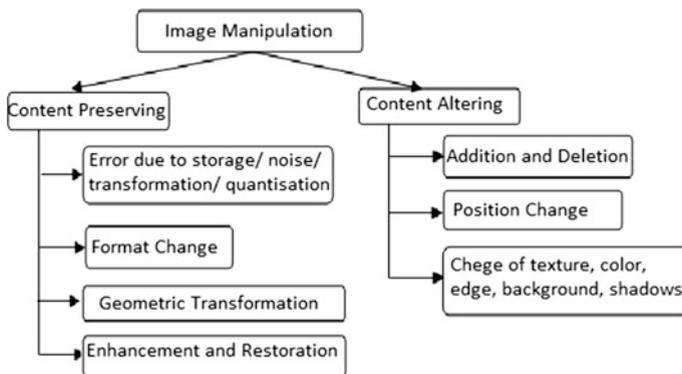


Fig. 1 Types of image manipulation

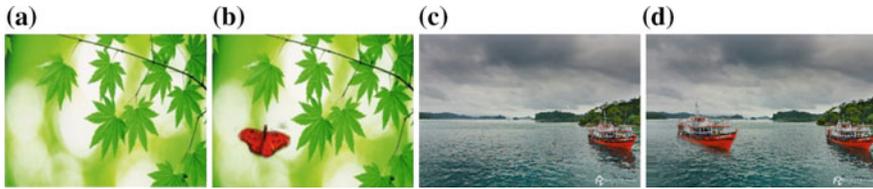


Fig. 2 a, c Original Images. b Spliced image. d Cloned Image

2 Types of Image Forgery

Image forgery is categorized into four types such as, copy–move/cloning, splicing, retouching and morphing [3]. Copy–move or cloning is a procedure where a part of an image is copy-pasted into some other area of the image itself. Whereas, splicing is another method in which parts of two or more images are stitched together to create the forged image. Retouching is achieved by adjusting color, sharpness, brightness, noise, contrast, etc. Morphing is a procedure that transfers an image to a different one through seamless transition between two images [4]. In Fig. 2 shown are four images out of which the second image is a spliced image obtained from first image and the fourth is a clone of the third.

3 Techniques of Digital Image Forgery Detection

Basically, the image forgery detection procedures fall into two classes. One is active detection and another is passive detection. Active detection requires in-built digital signature and/or watermarking whereas passive detection does not require any of these two. It works by analyzing the content and statistics of the image [5]. Image content based detection methods are further divided into the splicing detection and cloning detection methods. Cloning detection techniques are generally based on comparisons and matching. This falls into two major classes namely; the exhaustive search method and the overlapping block matching (OLBM) techniques. The exhaustive search method necessitates every possible part of an image to be compared with every other part of the image to detect a match. This gives rise to high time complexity. Just like the exhaustive search method, simple OLBM method also detects clones in a forged image, if the image is not subjected to further post-processing such as intensity variation, noise contamination, compression, etc. The time complexity of this method is less than the time complexity of the exhaustive search but, still remain as high as $O(b^2B^2)$ where, $b \times b$ is the block size and $B = (M - b + 1)(N - b + 1)$, $M \times N$ being the size of the image. In addition to the time complexity, the other dilemma with this method is choosing a right block size. If larger blocks lead to smaller detection accuracy then smaller blocks are resulted with high false positive rates. The false positives are

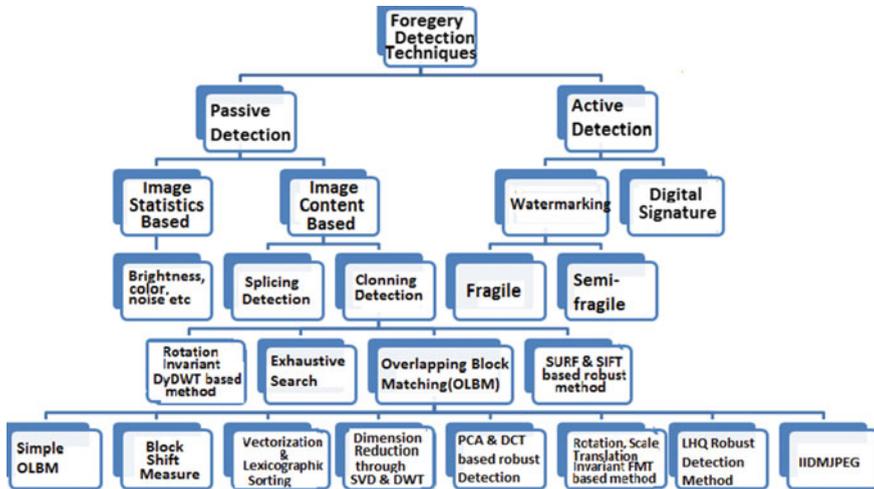


Fig. 3 Classification of forgery detection techniques

reduced by measuring the block shift and the search time can be improved using vectorization and lexicographic sorting. Further reduction in dimension and improvement in search time is achieved by using DWT and SVD methods. DCT- and PCA-based methods provides robustness against intensity changes [6]. Coming to robust detection methods, Luo et al. [7] suggested seven feature-based detection technique provides robustness against JPEG compression, noise attacks and blurring but fails to detect intensity variant clones. Whereas, the four feature based intensity invariant detection model for JPEG compressed images (IIDMJPEG) proposed by Mishra and Adhikary detects intensity variant clones in JPEG compressed images as well as provides robustness against noise attack and blurring [8]. In Fig. 3 is presented a classification structure of various digital image forgery detection methods.

3.1 Passive Detection: Non-robust Methods

The passive forgery detection techniques have evolved through several phases from the non-robust methods to highly robust techniques. The non-robust techniques fail to detect forgery when the forged image is subjected to lossy compression, blurring, rotation, scaling, noise attacks, etc. On the other hand, the robust methods succeed in detecting forgeries even if the tampered image is subjected to one or more of these above-mentioned post-processing operations. This and the following sub-section present a brief record of the evolution of the passive detection methods.

Fridrich et al. [9] give the idea about two efficient detection methods. First one is an exhaustive search method where an image and its circularly shifted version are

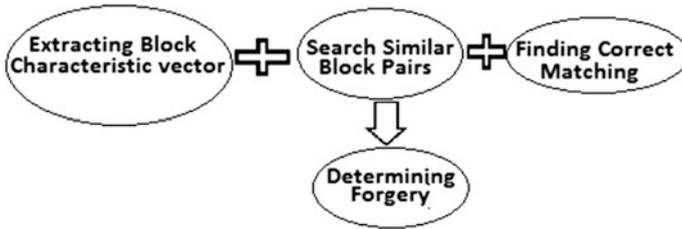


Fig. 4 Forgery detection through block matching

compared for matched parts. Though this is an effective method, the computational complexity of this method is very high which makes it impossible for practical use for large-sized images. The second method suggested by the authors is the auto-correlation method where the forgery is localized by the peaks corresponding to the cloned and the original parts in an image. The authors also suggested an overlapping block matching method that involves vectorization and lexicographic sorting. In this method, a square block of $b \times b$ pixels slides through an image one pixel at a time to give $(M - B + 1)(N - B + 1)$ blocks. These blocks are matched to locate the clones. The steps of block-based exact match are given in Fig. 4.

Murali et al. [10] in their paper suggested a DCT-based method for detecting forgery in a JPEG compressed image and a standard deviation based edge detection method. Zhouchen [11] have proposed a robust method based on double quantization (DQ) effect and have applied a trained SVM (Support Vector machines) to take tampering decision. According to the authors, DQ effect shows periodic peaks and valleys in the histogram of DCT coefficients and hence can be applied for tampering detection. Singh et al. [12] have used a DCT and SIFT based method for feature extraction and forgery detection. Zimba and Xingming [13] proposed two similar algorithms based on PCA and DWT. The method proposed by Popescu [14] divides an $N \times N$ image into $K = (N - b + 1)^2$ overlapping blocks. Each block is reshaped into b^2 long row vectors and inserted into a $K \times b^2$ feature matrix. To improve the time complexity, DWT is applied to the blocks to reduce the size of feature vector. They also proposed a method [15] based on principal component analysis and Eigen value decomposition (PCA-EVD) that reduces the dimension of the feature vectors and improves the computation time.

Mahdian and Saic [16] proposed a blur invariant method that successfully detects clones even if the cloned regions are blurred and noisy. Junhong [17] presented an LLE (Locally Linear Embedding)—a nonlinear dimension reduction technique that detects copy–move forgery as well as fused edges. Roweis et al. [18] have also used LLE for tamper localization. Kaur [19] has used Local binary pattern (LBP)—a texture descriptor for feature extraction and have used similarity criterion and Euclidean distance threshold for detection of clones.

3.2 *Passive Detection: Robust Techniques*

Jessica et al. [9] in their paper have suggested a technique robust against JPEG compression along with the techniques discussed in the above sub-section. This method is same as that of the exact match with one difference that it considers quantized DCT coefficients for comparison instead of the pixel representations for finding similarity.

Muhammad et al. in their paper [20] have proposed a dyadic Wavelet Transform (DyWT) and DWT-based robust and improved forgery detection model. Wang et al. [21] have proposed a DWT, DCT, and Eigen vector based method that is invariant to JPEG compression and additive noise. In this method first DWT and DCT are applied to the image blocks and then the resulting coefficients are multiplied to form the Eigen vectors. Block match is calculated measuring the mean and variance of distances between Eigen values of Eigen vectors for forgery localization. Zhang et al. [22] have combined phase correlation with DWT for their detection model. They have calculated spatial offset between copied and pasted region to compute the difference between the image and its shifted version. This method is claimed to reduce time complexity and is robust to jpeg compression. Li et al. [23] also have suggested a DWT- and SVD-based model where SVD (Singular Value Decomposition) have been applied to the blocks of the low frequency sub-band then the SV vectors are lexicographically sorted to locate the clones. This method is robust to JPEG compression up to quality level 70.

PCA-based robust technique proposed by Popescu and Farid [24] efficiently detects duplicated region with an computational cost in the order of $O(N\log N)$ but is sensitive to noise or lossy compression. In paper [25], Amtullah et al. have used a faster and robust to noise speeded up robust feature (SURF)—a rotation and scale invariant key-point detector and descriptor-based algorithm. To identify the duplicated regions, the authors of [26] have combined KD-tree with SURF. A K-dimensional tree or KD-tree is a binary tree with nodes as k-dimensional points and is common technique in nearest neighborhood search. A KD-tree with N-nodes needs $O(\log_2 N)$ search operations. Prerna et al. [27] also used KD-tree with SHIFT and RANSAC (Random Sample Consensus) algorithms. RANSAC algorithm has been used to find out the unreliable key points. This technique is robust to additive noise and JPEG compression. There are many more publications [28, 29] on KD-tree based methods but all are not included here. Authors of [30, 31] have detected clones based on SIFT algorithm where they have extracted the SIFT descriptors and then matched those to identify the forgery. The SIFT (scale invariant features transform) algorithm works in four steps such as, the scale space extrema (SSE) detection, the key-point localization, orientation assignment and key-point description. This method is claimed to be robust against noise attack, JPEG compression, rotation, and scaling.

Luo et al. [7] in their paper introduced a seven feature-based method that is robust against noise, blur attacks, and lossy compression. Muhammad et al. [32] have used multi-scale segmentation and denoising-based efficient technique for

clone detection. This method is claimed to be robust against noise and blurring attacks. In the paper hybrid copy–move forgery detection technique using regional similarity indices [33], the authors developed a forgery detection using local fractal dimension for image segmentation and estimating SSIM (Structural Similarity Index Measure) between each block pair in each segmented region to localize the forged regions. Bayram et al. [34] have suggested a FMT (Fourier-Mellin transform) based robust to noise, blur, rotation, scaling, and JPEG compression method. They have used counting bloom filters (CBF) instead of lexicographic sorting for computation time improvement. Solario et al. [35] have suggested a one-dimensional descriptor invariant to reflection and rotation based on log polar co-ordinates. In this method, the pixels of overlapping blocks are represented in log polar co-ordinates and summed along the angle to obtain the descriptor.

3.3 Comparison of Different Forgery Detection Methods

A comparison of various important forgery detection techniques evolved from time to time is given in the Table 1.

Table 1 Comparison of different forgery detection techniques

Method used	Paper serial	Advantages	Disadvantages
Seven feature based robust algo	[7]	Robust to blurring, noise, lossy compression	Not tested for rotation and scaling
Four feature based IIDMJPEG	[8]	Robust to blurring, noise, lossy compression. Can detect intensity variant clones	Not tested for rotation and scaling
Exhaustive Search and Autocorrelation	[9]	Detects cloned images without post-processing	Time complexity. Cannot detect if changes is subjected to post-processing operations
DCT	[10]	Detect forgery in JPEG compressed image	Fails in case of rotation, scaling
DCT and DQ	[11]	Fast, robust against JPEG compression and various forgery methods, fine grained detection	Not tested for rotation, scaling, intensity change
DCT and SIFT	[12]	Reduced time complexity, robust against rotation, scaling and noise	Not tested for rotation
DWT, PCA, EVD	[13]	Reduced feature dimension. Better accuracy	Fails in case of rotation, scaling and heavy compression

(continued)

Table 1 (continued)

Method used	Paper serial	Advantages	Disadvantages
PCA and EVD	[15]	Reduced the dimension and improved computational time	–
BLUR	[16]	Robust against noise, JPEG compression, blurring	–
LLE	[17, 18]	Detects copy-move forgery as well as fused edges	–
DWT-DCT and Eigen vector	[21]	Invariant to JPEG compression and additive noise	–
DWT, phase correlation	[22]	Reduce time complexity and robust to JPEG compression	–
DWT-SVD	[23]	Robust to JPEG compression up to QF 70, less time complexity	Not invariant to rotation and scaling
PCA	[24]	Efficient method, low false positives	Sensitive to noise and lossy compression
SURF	[25]	Invariant to rotation and scaling. Faster and robust to noise	–
KD-tree, SIFT	[27]	Robust to additive noise and JPEG compression	–
SIFT	[30, 31]	Robust against noise attack, JPEG compression, rotation and scaling	–
FMT	[34]	Robust to blurring, noise, scaling, lossy compression and transitional effects	Cannot detect forgeries which have rotation of above 10° and scaling of 10%
Log polar descriptor	[35]	Invariant to reflection and rotation	–

4 Summary and Conclusion

Even after a lot of research has been carried out during the last decades, passive forgery detection still continues to be an open research area. Because copy–move forgery or cloning is a technique commonly used by manipulators to forge digital images so, copy–move forgery detection techniques form one of the most important classes of passive detection techniques. The overlapping block matching (OLBM) method suggested in 2003 by Fridrich et al. was one of the important developments in the field of cloning detection. Many improvements have been suggested by various researchers from time to time to improve the time complexity of the algorithm as well as to make the detection algorithm robust against post-processing operations such as changes in contrast, brightness and color, noise and blurring attacks, lossy compressions, geometric transformations such as rotation, scaling and

translation to certain extends but still there exists the need to develop an efficient algorithm that will be able to detect forgeries even after multiple post-processing operations and those have been subjected to more than 10% scaling and rotated by an angle greater than 10° .

References

1. Haouzia, A., Noumeir, R.: Methods for image authentication: a survey. *Multimedia Tools Appl.* **39**(1), 1–46 (2008)
2. Mishra, M., Adhikary, M.C.: Digital image tamper detection techniques—a comprehensive study. *Int. J. Comput. Sci. Bus. Inform.* **2**(1), 1–12 (2013)
3. Kaur, H., Kaur, K.: A brief survey of different techniques for detecting copy–move forgery. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **5**(4), 875–882 (2015)
4. Shah, H., Shinde, P., Kukreja J.: Retouching detection and steganalysis. *Int. J. Eng. Innov. Res.* **2**(6), 487–490 (2013)
5. Gupta, A., Saxena, N., Vasistha, S.K.: Detecting copy-move forgery using DCT. *Int. J. Sci. Res. Publ.* **3**(5), 1–4 (2013)
6. Mishra, M., Adhikary, M.C.: Detection of clones in digital images. *Int. J. Comput. Sci. Bus. Inform.* **9**(1), 91–102 (2014)
7. Luo, W., Huang, J., Qiu, G.: Robust detection of region-duplication forgery in digital image. In: *Proceedings of 18th International Conference on Pattern Recognition IEEE*, vol. 4, pp. 746–749 (2006)
8. Mishra, M., Adhikary, M.C.: Robust detection of intensity variant clones in forged and JPEG compressed images. *ANVESA* **9**(1), 48–60 (2014)
9. Fridrich, J., Soukal, D., Lukas, J.: Detection of copy-move forgery in digital images. In: *Proceedings of Digital Forensic Research Workshop* (2003)
10. Murali, S., Chittapur, G.B., Prabhakara H.S., Anami, B.S.: Comparison and analysis of photo image forgery detection techniques. *IJCA* **2** (6), 45–56 (2012)
11. Zhouchen, L.: Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. *Pattern Recogn.* **42**(11), 2492–2501 (2009)
12. Singh, R., Oberoi, A., Goel, N.: Copy–move forgery detection on digital images. *Int. J. Comput. Appl.* **98** (9), (2014)
13. Zimba, M., Xingming, S.: DWT-PCA (EVD) based copy-move image forgery detection. *Int. J. Digital Content Technol. Its Appl.* **5**(1), 251–258 (2011)
14. Popescu, A.C., Farid, H.: Exposing digital forgeries by detecting traces of resampling. *IEEE Trans. Signal Process.* **53**(2), 758–767 (2005)
15. Zimba, M., Xingming, S.: Fast and robust image cloning detection using block characteristics of DWT coefficients. *Int. J. Digital Content Technol. Its Appl.* **5**(7), 359–367 (2011)
16. Mahdian, B., Saic, S.: Detection of copy–move forgery using a method based on blur moment invariants. *Forensic Sci. Int.* **171**(2), 180–189 (2007)
17. Junhong, Z.: Detection of copy-move forgery based on one improved LLE method. In: *Advanced Computer Control (ICACC) IEEE*, vol. 4, pp. 547–550 (2010)
18. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
19. Kaur, R.: Copy–move forgery detection utilizing local binary patterns. *Int. J. Emerg. Technol. Comput. Appl. Sci.* **7**(3), 290–294 (2013)
20. Muhammad, N., Muhammad, H., Muhammad, G., Bebis, G.: Copy–move forgery detection using dyadic wavelet transform. In: *Eighth International Conference on IEEE Computer Graphics, Imaging and Visualization (CGIV)*, pp. 103–108 (2011)

21. Wang, X., Zhang, X., Li, Z., Wang, S.: A DWT-DCT based passive forensics method for copy-move attacks. In: 2011 Third International Conference on Multimedia Information Networking and Security, IEEE, pp. 304–308 (2011)
22. Zhang, J., Feng, Z., Su, Y.: A new approach for detecting copy-move forgery in digital images. In: 11th IEEE Singapore International Conference on IEEE Communication Systems, pp. 362–366 (2008)
23. Li, G., Wu, Q., Tu, D., Sun, S.: A sorted neighbourhood approach for detecting duplicated regions in image forgeries based on DWT and SVD. In: IEEE International Conference on Multimedia and Expo, pp. 1750–1753 (2007)
24. Popescu, A., Farid, H.: Exposing digital forgeries by detecting duplicated image regions. Technical Report TR2004-515, Department of Computer Science, Dartmouth College, pp. 1–1 (2004)
25. Amtullah, S., Koul, A.: Passive image forensic method to detect copy-move forgery in digital images. *IOSR-JCE* **16**(2), 96–104 (2014)
26. Shivakumar, B.L., Baboo L.D.S.S.: Detection of region duplication forgery in digital images using SURF. *Int. J. Comput. Sci. Issues* **8**(4), (2011)
27. Prerna, C., Percy, G.J., Angaline, S., Thanga, B.I.: A key-point based copy-move forgery detection. *Int. J. Adv. Inform. Sci. Technol.* **12**(12), 175–180 (2013)
28. Sagawa, R., Masuda, T., Ikeuchi, K.: Effective nearest neighbor search for aligning and merging range images, In: Fourth International Conference on IEEE 3-D Digital Imaging and Modeling, pp. 54–61 (2003)
29. Bentley, J.L.: Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**(9), 509–517 (1975)
30. Huang, H., Guo, W., Zhang, Y.: Detection of copy-move forgery in digital images using SIFT algorithm. *Comput. Intell. Ind. Appl. Pacific-Asia Workshop IEEE* **2**, 272–276 (2008)
31. Lowe, D.G.: Distinctive image features from scale-invariant key points. *Int. J. Comput. Vision* **60**(2), 91–110 (2004)
32. Muhammad, N., Hussain, M., Muhammad, G., Bebis, G.: A non-intrusive method for copy-move forgery detection. In: International Symposium on Visual Computing. Springer, Berlin, pp. 516–525 (2011)
33. Oommen, R.S., Jayamohan, M.: A hybrid copy-move forgery detection technique using regional similarity indices. *Int. J. Comput. Sci. Inform. Technol.* **7**(4), 127–134 (2015)
34. Bayram, S., Sencar, T., Memon, N.: An efficient and robust method for detecting copy-move forgery. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1053–1056 (2009)
35. Solorio, S.B., Nandi, A.K.: Passive forensic method for detecting duplicated regions affected by reflection, rotation and scaling. In: 17th IEEE European Signal Processing Conference, pp. 824–828 (2009)

m-GeoEduNet: Mobile SDI Model for Education Information Infrastructure Network

R.K. Barik, R.K. Lenka, A.B. Samaddar, J. Pattnaik, B. Prakash and V. Agarwal

Abstract This paper represents scope for development and implementation of m-GeoEduNet prototype to share educational institute information infrastructure framework in India. It uses modern innovation on cloud computing, mobile technology, API technology, and spatial technology. The established prototype facilitates interface to collect the information related to institutes particularly IITs, NITs, and IIITs in terms of geo-spatial data. It can be attained by interacting between the users and mobile platform to achieve mobile-based Spatial Data Infrastructure (SDI) model. The collected data being documented and uploaded in Cloud database server. The present paper uses Firebase for storing of cloud database services and NOSQL database used for storing geo-spatial data. It has also been underlying third-party API, i.e., GeoFire; for efficiently handling queries on geo-spatial data.

Keywords Geo-spatial data · NOSQL · SDI and GIS

R.K. Barik (✉)

School of Computer Application, KIIT University, Bhubaneswar, India
e-mail: rabindra.mnnit@gmail.com

R.K. Lenka · B. Prakash · V. Agarwal

Department of Computer Science & Engineering, IIIT Bhubaneswar, Bhubaneswar, India
e-mail: rakeshkumar@iiit-bh.ac.in

B. Prakash

e-mail: b112012@iiit-bh.ac.in

V. Agarwal

e-mail: b112057@iiit-bh.ac.in

A.B. Samaddar

NIT Sikkim, Ravangla, India
e-mail: absamaddar@yahoo.com

J. Pattnaik

CSM Technologies Pvt Ltd, Bhubaneswar, India
e-mail: jashikapattnaik24@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_30

291

1 Introduction

The overall growth of any country technically and economically permits a fast development in the sector of Education. Education plays an important role in creating quality human assets to serve the nation. In the other side, this sector is also confronting huge competition in terms of globalization and technology employed. Hence there is requirement of innovative initiative to meet the upcoming challenges. Harmonized efforts should be taken to extract quality information regarding educational stuffs in precise way. It can be achieved with the incorporation of latest technologies such as cloud computing, mobile technology, spatial technologies, and API technology. To analyze geographical information for planning related activities, spatial technology is used to determine proper dissemination and deployment of resources. Spatial Technology is assemblage of tools to store and retrieve spatial data and showcase those data from the practical world to fulfill many purposes. GIS, remote sensing and GPS with mobile cloud are spatial technologies used to generate Spatial Data Infrastructure (SDI) model holds potential to remove bottlenecks that impede the productivity in Education sector. Further, easy perceiving spatial information is required for planners and decision-makers for the implementation of Right To Education (RTE) Act [1]. Major challenges upon the Education system in India comprise infrastructure, quality, trained manpower, access to education, remoteness, high dropout rate, and constrained resource distribution. The rate of participation by the population is very low in our country especially in Higher Education. As per Research Findings, the cultural, social, economic, and geographical barriers resist people to go for higher learning. Therefore, it is essential to facilitate quality education as well as to aware about information on educational infrastructure to aspiring candidates. This ultimately augments professional opportunities for students and promotes actions supporting the increase in student enrollments in institutes [2].

In India, the higher education sector observed a marvelous surge in institutional capacity post-independence. It ranks third in terms of enrollment of students after the United States of America and China worldwide. The information regarding all institutes is not obtainable in a single platform causing hitches for people. It would be really wise if users can acquire information about institute like its location, contact details, ongoing events, and vacancy for various degrees and post in a single platform. The availability of such information in web is not well systematized which misinforms most of the people [3]. With the integration of mobile cloud technology with GIS, access to spatial data can be made obtainable anywhere and anytime [4, 5]. It can facilitate data on real time which dynamically used to display spatial data via maps on mobile platform. This application allows users to generate collaborative queries, examine spatial information, modify map data and shows the overall result operations in fresh way represented on the map. So, need arise to launch well-organized mobile SDI model that act as interface for access and exchange of information proficiently from anywhere [6–8].

Keeping this in mind, the present work prototype m-GeoEduNet, a mobile-based SDI model has been developed and implemented for sharing of educational institute

information infrastructure framework in India. It offers a platform for assembling geo-spatial data assimilated from the interaction between users and mobile application. Geo-spatial data collected and mapped to documented data and uploaded to Cloud database server. The present research uses Firebase for storage of cloud database services and NOSQL database; used for storing spatial data and the underlying third-party API, i.e., GeoFire; has been explicitly designed for competently handling queries on geo-spatial data. Furthermore, the brief review of related works giving detail description for educational and other information infrastructures to derive SDI model is described in Sect. 2. Section 3 describes on the proposed mobile-based SDI model. Section 4 describes about the prototype development. Various experimental results with set of analysis are discussed in Sect. 6. The concluding remarks and way forward of the research is described in Sect. 7.

2 Related Works

Initially, SDI model proposed and developed for educational information infrastructure in Service-Oriented Architecture (SOA) environment. Further, several works has been done for better visualization of technical institute of India by open-source software (OSS) tools. The developed prototypes only operated through thin client architecture, i.e., only through web browsers [1, 9–12]. In the era of mobile cloud computing with spatial technology, several architectures has been proposed for visualization of spatial data in wide range of application areas with OSS resources [13, 14]. Several designing architectures implemented for mobile-based SDI model. Figure 1 has shown the general architecture to manage reports of road pavement damage data via mobile and desktop. System architecture displays that the Android-based APP developed for the data acquisition and better visualization of spatial and nonspatial data [15, 16].

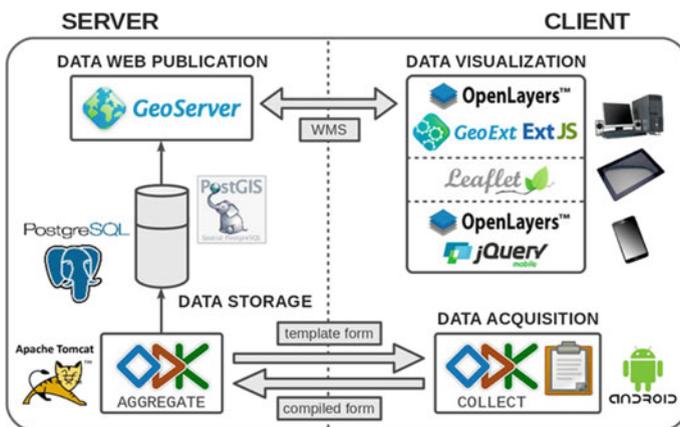


Fig. 1 General system architecture for mobile and other devices [17]

It has been perceived that mobile computing with cloud devices have great potential for geo-spatial data visualization and collection in diverse field. In forest inventory management, mobile mapping with high-resolution remote sensing images, grid maps, and vector maps used for data processing and surveying, those are essential tools for the forestry science workers [18]. Air pollution mobile app has also developed with the help of OSS for air quality management, greenhouse gases, and emission of air pollutants by Protection Agencies such as California Environmental Protection Agency and United States Environmental Protection Agency [19]. Mobile-based SDI has been one of the important requirements for geo-scientist for accurate raw data collection [20]. EpiCollect and Creek Watch are two examples of mobile apps and are popular for the field of environmental monitoring and watershed management [21–23]. It is benefitted for the general people in health sector. m-health has been developed and implemented as a mobile app for public health information collection [24]. With the help of earth observation data, an Android-based mobile app also developed for the accurate dispersal of rainfall estimations [25]. From these several reviewing of the research works, it has been summarized that spatial data infrastructure model in mobile platform is need of the hour for educational data infrastructure hub.

3 Objective of the Present Work

The major aim of the current assignment is development of the prototype m-GeoEduNet, a SDI model in mobile platform for sharing the educational institute data network in India. The objective of the developed model is to provide a platform for gathering geo-spatial data acquired from interaction between users and mobile interface. The prototype generates valid information where authenticated user submits information and is openly provided to other users. Geo-spatial data can be collected and mapped to the documented data and uploaded to the Cloud database server. The present prototype uses Firebase for storing of cloud database services and also NOSQL database model; has been used for storing geo-spatial data. It has been also using third-party API, i.e., GeoFire; designed for efficiently handling spatial queries. Thus the next section describes the detail system architecture of m-GeoEduNet.

4 The System Architecture of m-GeoEduNet

The proposed system architecture of m-GeoEduNet consists of user layer, data security layer, and data layer as shown in Fig. 2. The user layer consists of android-based mobile software application where users connect to GIS database service to query information on location of various institutions, which further mapped to documented information of the respective institution. The security layer consists of security rules for reading from and writing to the database. When client

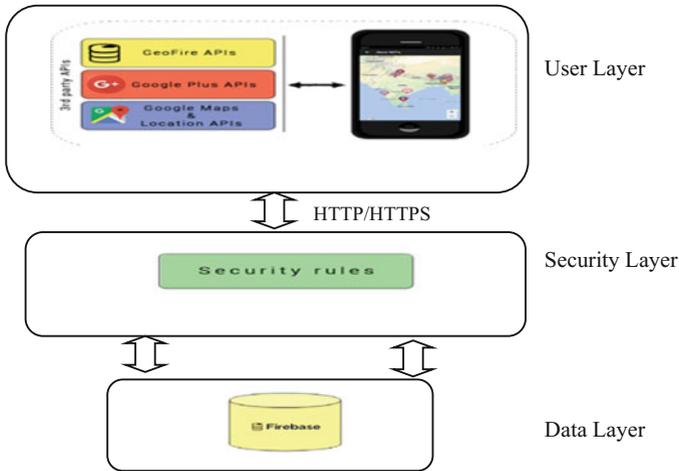


Fig. 2 System architecture of m-GeoEduNet

send a request to the database service such as GeoFire query to get locations of all institutes established between 1990 and 2000, request will be parsed by the GeoFire server and required geo-spatial data retrieved using HTTP connection in JSON (Java Script Object Notation) format and then a new layer containing the location markers will be added to the map using Google Maps APIs. The data layer consists of geo-spatial data that are stored in a NOSQL database. In the proposed SDI model, Firebase NOSQL database is used to store and manage geo-spatial data in JSON format. The Firebase database uses JSON as its data model, and stores the data in key value pair repository [26–28].

5 Prototype Developments for m-GeoEduNet

For the prototype development of m-GeoEduNet, the primary emphasis is on the real-world method to determine and spread the thought of using the model in academic sector. The established model has to deliver a proficient means of allocation of geo-spatial and nonspatial data in mobile platform. The prototype is based on Object-Oriented Software Engineering (OOSE) proposed by Jacobson’s method to conglomerate the time critical nature and strong user focus [29]. Figure 3 represents the fully win-win procedure model for the development of m-GeoEduNet.

The model incorporates modules namely Registration, Educational Institutions Information Mapping and Data Filtering Services in Module I, II, and III respectively. The registration process for authenticated user is mentioned in Module I. Post process, the user can access model in complete functionalities. The comprehensive designing and showing of Education Data Infrastructure Network by means of numerous dynamics allied with accessibility in the Nation mentioned in

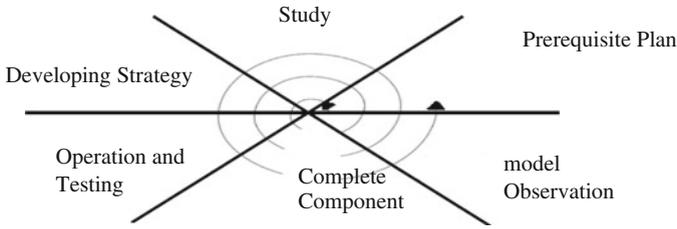


Fig. 3 Win-win spiral model for m-GeoEduNet

Module II. The present studies have been taken information from IITs, NITs, and IIITs. The data filtering process filters the retrieved dataset based on various parameters like established year, state in which the institute is located in Module III. The prerequisite phase to design the app intends to stipulate the model performance by user perception which has been shown through state diagram in Fig. 4. Apprehending proficient app consumers according to the know-how of the situation facilitates system development.

The developed prototype, m-GeoEduNet is very advantageous for people of any age group and locality to gain detailed information about the universities and their quality of education. Apart from the general user, m-GeoEduNet used by planners and builders to analyze the spatial dissemination of various institutes in study area and chose the best-suitable location in the future to establish new institutes. Although, there are many open source libraries and GIS suites available, m-GeoEduNet proposes the use of third-party location API, i.e., GeoFire which maintains spatial-data ineffectual manner by mapping data with geographic location using unique key. This makes the queries return result faster and offers functionality to retrieve data within a certain radius of a particular location. GeoFire maintains

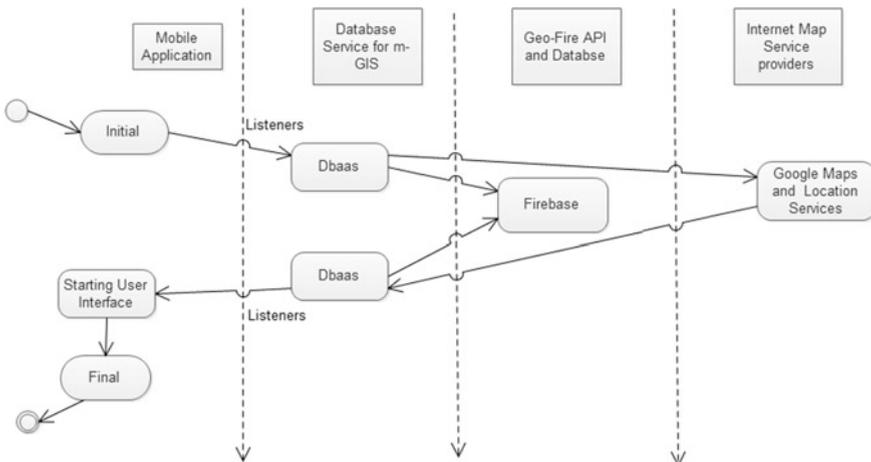


Fig. 4 State diagram for m-GeoEduNet

data on NOSQL-based database called Firebase which offers better performance over traditional RDBMS for basic GIS operations.

m-GeoEduNet perform complex geo-queries, NOSQL seems to be a better fit for m-GeoEduNet than the traditional RDBMS to achieve gain in performance. NOSQL has benefits like horizontal scalability, dynamic schemas over RDBMS which fits into the requirements of m-GeoEduNet. So, the application of the proposed SDI model is implemented through Android mobile application: m-GeoEduNet which includes JAVA and XML for frontend development, Java Script Object Notation (JSON) for formatting the data and transfer the collected spatial information to and from the Firebase and third party APIs such as Google Maps API, Google Location API, GeoFire API.

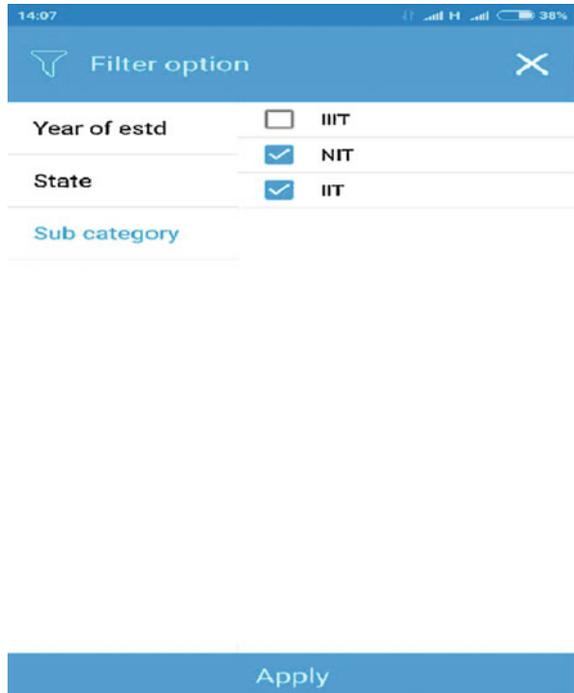
6 Experimental Results

After designing phase, the m-GeoEduNet coded through Android Studio IDE. The various screenshots have been illustrated as follows. Figures 5 and 6 illustrate the view of all old NITs in India and the filtering page for various kind of filtration option according to IIT, NITs and IIITs.

Fig. 5 Old NITs in India



Fig. 6 Filter page for m-GeoEduNet



7 Concluding Remarks

In this paper, we have developed and implemented mobile SDI model to map spatial data with documented data to provide an expedient platform for decision-makers to decide potential locations for building new institutions. It built on a reliable platform for users to view information related to various institutions and notify them about ongoing events faster. Also, we have analyzed the constraint of distributed information obtainable via Internet and aimed to solve few of them. To achieve the data availability faster, we used real-time database and for providing quality information, we fetched data from reliable sources. It is being used for spatial data mapping and analysis based on GIS to improve the dissemination of educational services. The m-GeoEduNet utilizes the power of GPS to create SDI model that maps the documented data with spatial data. This mobile SDI model provides an expedient platform for user to access data about various institutes. Rather than using a disseminated platform, such as the list of increasing websites, m-GeoEduNet aims to provide single integrated platform to access information more easily as well as more efficiently.

References

1. Barik, R.K., Samaddar, A.B.: Service oriented architecture based SDI model for education sector in India. In: International Conference on Frontiers of Intelligent Computing Theory and Applications, pp. 555–562 (2014)
2. Lagrab, W., AKNIN, N.: Analysis of educational services distribution-based geographic information system (GIS). *Int. J. Sci. Technol. Res.* **4**(3), (2015)
3. Internet 1: <http://www.worldbank.org/>. India country summary of higher education. World Bank. Accessed on 23 June 2016
4. Barik, R.K., Dubey, H., Samaddar, A.B., Gupta, R.D., Ray, P.K.: FogGIS: Fog Computing for geospatial big data analytics. *IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering* 613–618 (2016)
5. Smith, J., Mackaness, W., Kealy, A., Williamson, I.: Spatial data infrastructure requirements for mobile location based journey planning. *Trans. GIS* **8**(1), 23–22 (2004)
6. Li, M., Wang, H.: Design for mobile GIS based on embedded database. In: Conference Anthology, IEEE, pp. 1–3 (2013)
7. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless Commun. Mobile Comput.* **13**(18), 1587–1611 (2013)
8. Delipetrev, B., Jonoski, A., Solomatine, D.P.: Development of a web application for water resources based on open source software. *Comput. Geosci.* **62**, 35–42 (2014)
9. Samaddar, S.G., Barik, R.K.: A mobile framework for geographical indication web services. In: Third International Conference on Computational Intelligence and Information Technology, 2013. CIIT 2013. pp. 420–426 (2013)
10. Sakhare, P., Mascarnes, S., Chaudhari, A.: Development of WebGIS framework for Indian technical institutes using open source GIS tools. In: International Conference on Computer, Communication and Control (IC4), pp. 1–4 (2015)
11. Internet 2: <http://schoolgis.nic.in/index.html>. Accessed on 1 Aug 2016
12. Okan, E.R.A.Y.: Application of geographic information system (GIS) in education. *J. Tech. Sci. Technol.* **1**(2), 53–58 (2012)
13. de Abreu Freire C.E., Painho, M.: Development of a mobile mapping solution for spatial data collection using open-source technologies. *Procedia Technol.* **16**, 481–490 (2014)
14. Athanasios, N., Karagiannis, F., Palaiologou, P., Vasilakos, C., Kalabokidis, K.: AEGIS App: wildfire information management for windows phone devices. *Procedia Comput. Sci.* **56**, 544–549 (2015)
15. Quwaider, M., Jararweh, Y., Al-Alyyoub, M., Duwairi, R.: Experimental framework for mobile cloud computing system. *Procedia Comput. Sci.* **52**, 1147–1152 (2015)
16. Brovelli, M.A., Minghini, M., Zamboni, G.: Public participation in GIS via mobile applications. *ISPRS J. Photogrammetry Remote Sens.* **114**, 306–315 (2016)
17. Brovelli, M.A., Minghini, M., Zamboni, G.: Public participation GIS: A FOSS architecture enabling field-data collection. *Int. J. Digital Earth* **8**(5), 345–363 (2014)
18. Dong, C., Liu, F., Wang, H., Chen, F.: Application research of mobile GIS in forestry informatization. In: 5th International Conference on Computer Science and Education (ICCSE), pp. 351–355 (2010)
19. Gkatzoflias, D., Mellios, G., Samaras, Z.: Development of a web GIS application for emissions inventory spatial allocation based on open source software tools. *Comput. Geosci.* **52**, 21–33 (2013)
20. Chen, H., Xiao, K.: The design and implementation of the geological data acquisition system based on mobile GIS. In: 19th International Conference on Geoinformatics, pp. 1–6 (2011)
21. Aanensen, D.M., Huntley, D.M., Feil, E.J., Spratt, B.G.: EpiCollect: linking smartphones to web applications for epidemiology, ecology and community data collection. *PLoS ONE* **4**(9), e6968 (2009)

22. Tang, Z., Liu, T.: Evaluating Internet-based public participation GIS (PPGIS) and volunteered geographic information (VGI) in environmental planning and management. *J. Environ. Planning Manage.* **59**(6), 1073–1090 (2016)
23. IBM, Creek Watch. <http://creekwatch.researchlabs.ibm.com/>
24. Boonchieng, E., Boonchieng, W., Senaratana, W., Singkaew, J.: Development of mHealth for public health information collection, with GIS, using private cloud: A case study of Saraphi district, Chiang Mai, Thailand. In: 2014 International Computer Science and Engineering Conference (ICSEC), pp. 350–353 (2014)
25. Mantas, V.M., Liu, Z., Pereira, A.J.S.C.: A web service and android application for the distribution of rainfall estimates and Earth observation data. *Comput. Geosci.* **77**, 66–76 (2015)
26. Internet 2: FIREBASE. <https://www.firebase.com/docs>. Access 23 June 2016
27. Internet 3: Google developers. <https://developers.google.com>. Access on 23 June 2016
28. Internet 4: Android developers. <https://developer.android.com/index.html>. Access on 23 June 2016
29. Pressman, R.S.: *Software engineering: a practitioner's approach*. Palgrave Macmillan, USA, 2005

Investigation of Structural and Electronic Environments of Nitrogen-Doped CVD-Grown DLC Films

Ranjan Kr. Ghadai, Soham Das, Subhash Chandra Mondal
and Bibhu Prasad Swain

Abstract Nitrogen-doped diamond-like carbon (N-DLC) films were synthesized by thermal chemical vapor deposition technique by varying the nitrogen flow rate at a constant C_2H_2 flow. The influence of nitrogen incorporation on the microstructure and structural electronic properties of N-DLC films were investigated by Raman spectroscopy, scanning electron microscopy (SEM), X-ray diffraction (XRD), and X-ray photoemission spectroscopy (XPS). SEM images reveal the DLC nanoparticles formation and carbon nanowires formations for 70 and 130 sccm nitrogen flow rate respectively. The Raman spectra of N-DLC investigated D and G peaks at 1300 and 1600 cm^{-1} respectively. X-ray diffraction of N-DLC indicated the crystallite sizes 1.74 and 8.78 nm at different peak position. Due to nitrogen incorporation in DLC thin films the core orbital spectra of C(1s) shifted from 284 to 286 eV for nitrogen flow increased from 70 to 130 sccm. In the present paper, an attempt has been made to synthesize nitrogen incorporated DLC films using thermal chemical vapor deposition technique to study their structural and microstructural properties.

Keywords DLC thin films • SEM • XRD • Raman and XPS

R.Kr.Ghadai (✉) · S. Das
Department of Mechanical Engineering, Sikkim Manipal Institute of Technology,
Majitar, Rangpo 737136, India
e-mail: ranjan.g@smit.smu.edu.in

S. Das
e-mail: sohamengg10@gmail.com

S.C. Mondal
Department of Mechanical Engineering, IEST, Shibpur, Howrah, India
e-mail: scmondall@gmail.com

B.P. Swain
Centre for Materials Science and Nanotechnology,
Sikkim Manipal Institute of Technology, Majitar, Rangpo 737136, India
e-mail: bibhu.s@smit.smu.edu.in

1 Introduction

Diamond-Like Carbon (DLC) films have attracted a great deal of research attention due to their excellent mechanical, biological, optical, and tribological properties [1]. Although DLC films have strong hardness, low fractional coefficient, high optical transmittance superior, and very good anti-scratch property, they also possess good electrical insulation properties because of the coexistence of trigonal (sp^2) and tetrahedral (sp^3) bonded carbon matrices [2]. Nowadays, DLC film is considered as an attractive carbon electrode material showing similar potential windows with respect to the aqueous solvent stability and a good electrochemical reactivity [3]. DLC films present a good thin film example whose properties strongly depend upon the nature of coating and the technique used for the film deposition [4]. Different techniques have evolved for the deposition of diamond-like carbon thin films and the majority of the techniques used so far involved is chemical vapor deposition (CVD) [5]. Metal-containing diamond-like carbon (Me-DLC) films have attracted a great deal of research attention owing to their excellent mechanical, tribological, and biological properties [1].

The nitrogen flow rate has significant effect on the compositions and structures and hence on the hardness and elastic modulus of the films, and increasing nitrogen flow rate decreases drastically the hardness and elastic modulus of the films [6]. Atomic force microscopy (AFM) results of nitrogen-doped DLC film revealed that the surface morphology of the films became smooth due to doping nitrogen and FTIR results showed that amounts of C=N and C=N bonds increased gradually with increasing nitrogen partial pressure [7]. It has been found that the fraction of aromatic ring structures in the film increases with the increase of nitrogen (N_2) content and nitrogen doping is a feasible approach to improve the thermal stability of films [8].

2 Experimental Details

Figure 1 shows the image of thermal CVD deposition. Nitrogen DLC thin film were deposited on silicon substrates by thermal chemical vapor deposition technique. The substrates were first cleaned in 2% HF solution for 2 min and then it is ultrasonically cleaned in deionised water for 10 min. After the cleaning process the substrates were loaded into the vacuum chamber along with the carbon powder. During deposition diamond-like carbon thin film has been made by varying nitrogen flow rate and keeping acetylene (C_2H_2) flow rate constant. Four samples have been made for nitrogen flow rates of 70, 90, 110, and 130 sccm. The deposition temperature and deposition time were 1200° and 2 h respectively. The rate of increasing the temperature was $3^\circ C/min$ and the rate of decreasing temperature was $5^\circ C/min$.



Fig. 1 Image of thermal CVD set up

The influence of nitrogen incorporation on the microstructure and electrochemical properties of N-DLC films were investigated by Raman spectroscopy scanning electron microscopy, X-ray diffraction, and X-ray photoemission spectroscopy. Film characterization in terms of microstructure, composition, and chemical state of components was carried out by scanning electron microscopy (SEM) and X-ray photoelectron spectroscopy (XPS).

3 Results and Discussions

Figure 2 shows the SEM images of nitrogen-doped DLC films prepared at different nitrogen flow rates while C_2H_2 flow remained constant. The SEM images show different nanostructure of DLC thin films. At the low nitrogen doped ($N_2 = 70$ sccm) the nanoparticle were formed. The particle size of $N_2 = 70$ sccm flow rate indicates in the DLC thin films 50–500 nm. Due to low N_2 flowrate the particle are separated to each other and not agglomerate. However, at the higher N_2 doping the nanoparticles are agglomerated and formed as nanowires and agglomerated with each other. The diameters of the DLC nanowires are 500 nm with uniform distribution. It is also observed that the length of the nanowires varied from 2 to 5 μm . Cui et al. investigated corrosion and tribocorrosion performance of multilayer diamond-like carbon film in NaCl solution and find out similar microstructure in the single-layer DLC films [8]. Han et al. described the Influence of Nitrogen Flow Rate on the Microstructure of DLC film and they observed a smooth and dense surface due to the flow of nitrogen [9].

Figure 3 shows the XRD patterns of nitrogen DLC film prepared at different nitrogen flow rates which demonstrates the formation of nitrogen crystalline phase

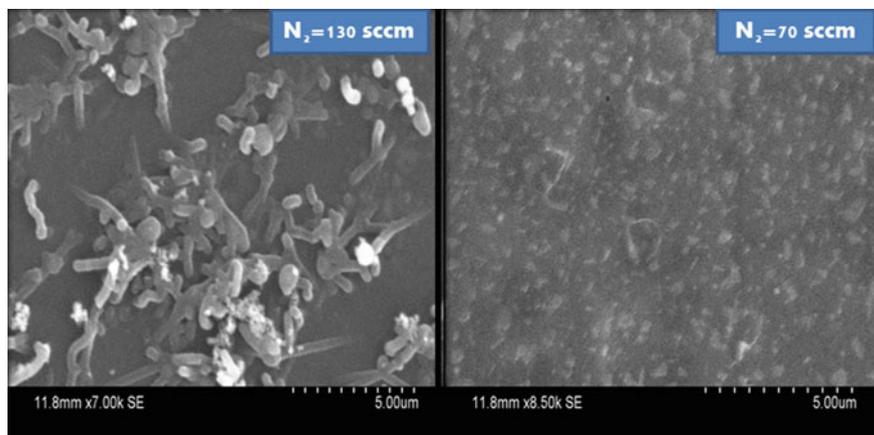
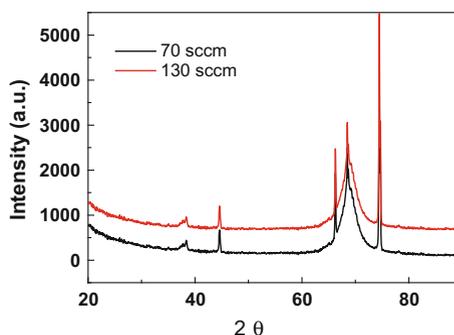


Fig. 2 SEM images of nitrogen DLC films prepared at the nitrogen flow rates of 70 and 130 sccm

Fig. 3 XRD spectra with N₂ flow rate 70 and 130 sccm



apart from the DLC base. Here the graph has plotted for two different nitrogen flow rates 70 and 130 sccm. For nitrogen DLC film two different peaks are identified at 68° and 76°. It is also showing that the intensity is higher for higher nitrogen flow rates. By using Scherrer equation the grain size can be calculated at different peaks. The peaks are coming at two different values of Bragg angles. For Bragg angle 38° and 34° the grain size are 8.78 and 1.74 nm. Han et al. described the influence of nitrogen on XRD peak and observed the peak at 37.6° [9]. They indexed 37.6° as (111) crystal plane of face-centered cubic (fcc) of DLC film.

Figure 4 shows the visible Raman spectra of the nitrogen-doped DLC films deposited at different nitrogen flow rates. For D-band the peak is centered at 1325 cm⁻¹ and for G-band the peak is centered at 1602 cm⁻¹. From the graph it is also showing that the intensity of peak is more for higher nitrogen flow rate. Ray et al. studied the nitrogen-doped diamond-like carbon (DLC) thin films and found the D peak and G peak at 1350 and 1550 cm⁻¹ respectively [6]. Zhou et al. analyzed the Raman spectroscopy of nitrogen-doped DLC film prepared by PECVD

Fig. 4 Raman spectroscopy for different N_2 flow rate 70 and 130 sccm

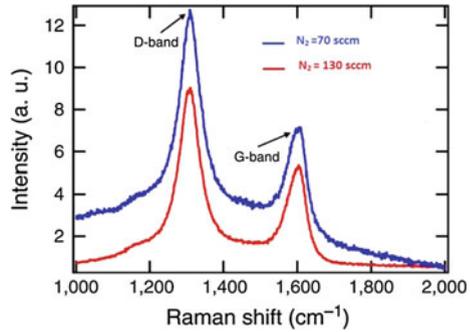
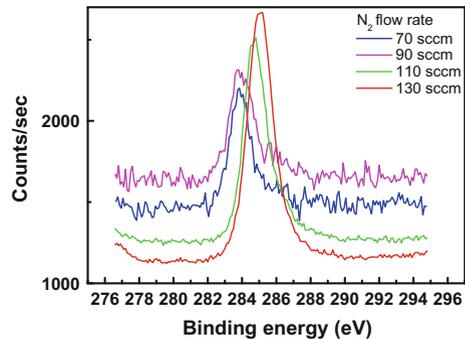


Fig. 5 XPS of nitrogen DLC films prepared at the nitrogen flow rates of 70, 90, 110 and 130 sccm



technique. They observed two obvious peaks, the first-order peak of carbon in the range of $800\text{--}1800\text{ cm}^{-1}$ centered at about 1500 cm^{-1} and the second-order peak of centered at about 960 cm^{-1} [10].

Figure 5 shows a typical C(1s) core orbital spectrum of nitrogen-doped DLC films at different nitrogen flow rate, deposited by thermal chemical vapor deposition technique. The peaks are at 284, 284.5, 285, and 286 eV for the nitrogen flow rate of 70, 90, 110, and 130 sccm respectively. The C(1s) peaks shifts its position 2 eV and the peak intensity increased with N_2 flow rate. The peak at a binding energy of 284 and 284.5 eV may be due to sp^2 hybridized carbon atoms, while the peak at 285 and 286 eV are due to the sp^3 hybridized carbon atoms. Ray et al. analyzed the N-doping DLC films and found the peak at 398 eV [6]. Jian-rong et al. deposited the nitrogen-doped DLC film and observed the XPS peak at 286 eV [11]. Swain et al. investigated C–N binding energy at 284.7 eV for C(1s) for a-CN:H films [7].

4 Conclusion

N-doped DLC films were successfully deposited by a thermal CVD by using C_2H_2 and N_2 gases. The morphology of DLC films prone to grown as nanoparticles and nanowires nanostructure. The estimated grain size of nanoparticles and nanowires

are 8.78 and 1.74 nm for 70 and 130 sccm N₂ flow rate DLC thin films respectively. In D-band and G-band, the peaks are centered at 1325 and 1602 cm⁻¹ which indicated graphitic carbon mixed with DLC films [12]. The N incorporation in the DLC indicated increasing of binding energy from 284 to 286 eV.

References

1. Bewilogua, K., Hofmann, D.: History of diamond like carbon films-from first experiment to worldwide application. *Surf. Coat. Technol.* **242**, 214–225 (2014)
2. Alfred, Grill: Diamond-like carbon: state of the art. *Diam. Relat. Mater.* **8**, 428–434 (1999)
3. Yoo, K.S., Miller, B., Kalish, R., Shi, X.: Electrodes of nitrogen-incorporated tetra-hedral amorphous carbon a novel thin-film electro catalytic material with diamond-like stability. *Electrochem. Solid-State Lett.* **2**, 233–235 (1999)
4. Robertson, J.: Diamond-like amorphous carbon. *Mater. Sci. Eng.* **R37**, 129–281 (2002)
5. Nakazawa, H., Mikami, T., Enta, Y., Suemitsu, M., Mashita, M.: Structure, chemical bonding and these thermal stabilities of diamond-like carbon (DLC) films by RF magnetron sputtering. *Japan. J. Appl. Phys.* **42**, 676 (2003)
6. Han, Z., Li, H., Lin, G., Dong, C.: Influence of nitrogen flow rate on the microstructure and properties of N and Me (Me=Cr, Zr) co-doped diamond-like carbon films. *J. Mater. Sci. Technol.* **26**(11), 967–972 (2010)
7. Xiao, J.R., Li, X.H.: WANG, Z.X.: Effects of nitrogen content on structure and electrical properties of nitrogen-doped fluorinated diamond-like carbon films. *Trans. Nonferrous Met. Soc. China* **19**, 1551–1555 (2009)
8. Liu, X.F., Xin, Z.H.O.U.: GAO, J.D.: Influence of nitrogen doping on thermal stability of fluorinated amorphous carbon thin films. *Trans. Nonferrous Met. Soc. China* **16**(1), 54–58 (2006)
9. Cui, Mingjun, Pu, J., Liang, J., Wang, L., Zhang, G., Xue, Q.: Corrosion and tribocorrosion performance of multilayer diamond-like carbon film in NaCl solution. *RSC Adv.* **5**, 104829–104840 (2015)
10. Ray, S.C., Pong, W.F., Papakonstantinou, P.: Iron, nitrogen and silicon doped diamond like carbon (DLC) thin films: a comparative study. *Thin Solid Films* **610**, 42–47 (2016)
11. Zhou, Kai, Ke, P., Li, X., Zou, Y., Wang, A.: Microstructure and electrochemical properties of nitrogen-doped DLC films deposited by PECVD technique. *Appl. Surf. Sci.* **329**, 281–286 (2015)
12. Swain, Bibhu P., Swain, Bhabani S., Hwang, Nong M.: Effect of H₂ dilution on a-CN: H films deposited by hot-wire chemical vapour deposition. *Appl. Surf. Sci.* **255**, 9264–9267 (2009)

A Novel Region Growing Based Method to Remove Pectoral Muscle from MLO Mammogram Images

Manasi Hazarika and Lipi B. Mahanta

Abstract Digital Mammography is the most efficient screening technique for early detection of breast abnormalities. Automated computer aided methods have been very effective in identifying subtle signs of breast cancer like microcalcification and masses. However such methods for detection of masses are highly affected by the presence of pectoral muscle in mediolateral oblique view. So it is highly recommended to remove pectoral muscle as a preprocessing step. In this paper, a novel method for pectoral muscle removal from Mediolateral Oblique mammogram images is presented. The method has three main phases. In the first phase a triangular region is defined over the mammogram that separates the pectoral muscle from the rest of the tissue. In the second phase, a local region growing method is applied within the triangular area defined to suppress the pectoral muscle. In the last phase, the pectoral edge is refined using gradient information of the edge. Results and Conclusion: The system is tested over 150 images taken from Mini-MIAS dataset. Hand-drawn segmentation masks are used to compare the segmentation accuracy for the proposed method.

Keywords Mammogram · Pectoral muscle · Region growing · Mediolateral oblique (MLO) view · Segmentation

1 Introduction

Cancer is one of the most life-threatening diseases since many decades. Breast cancer among women has emerged as very common type of cancers across the entire world. Unlike most of the cancers, breast cancer is completely curable if

M. Hazarika (✉)

Department of Computer Science, Gauhati University, Guwahati, Assam, India
e-mail: manasi.hazarika@gmail.com

L.B. Mahanta

Center for Computational and Numerical Studies, Institute of Advanced Studies in Science and Technology, Guwahati, Assam, India
e-mail: lbmahanta@iasst.gov.in

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_32

307

detected in the initial stages. Mammography is found to be one of the most efficient radiographic screening techniques for early diagnosis of breast cancer. But many a times it is difficult for a radiologist to detect the subtle signs of malignancy because of low contrast mammograms leading to unnecessary biopsy.

A good computer-aided diagnosis system can help in identification of early symptoms of breast cancer. In developed countries screening mammography has become a routine checkup activity amongst women. In such a scenario, a CAD system will be really helpful to the radiologists to inspect a large number of mammograms. However the performance of any CAD system is highly affected by the quality of the images. Mammogram images often come with low contrast due to the limitations of X-Ray hardware systems making it difficult to interpret the characteristics of mammograms. So enhancing the quality of images is required for most of the images before analyzing further. The presence of pectoral muscle in Mediolateral Oblique (MLO) affects in the segmentation process of mass detection algorithms. Intensity levels of both mass and pectoral muscle are very much similar. They possess higher intensity values than the rest of the image. So it is highly recommended to remove the pectoral muscles before applying any segmentation algorithm to get more accurate results.

In the literature many methods are available for the removal of pectoral muscle. A two phase based method to remove the pectoral muscle was presented in [1]. In this method, an adaptive histogram based method was used to isolate the breast area from the background in phase I and in phase II they applied a selective region growing algorithm to suppress the pectoral muscle. A pixel value mapping based approach was presented in [2]. The boundary was sharpened using morphological opening and closing operations. In [3] a segmentation method using Markov Random Field and Bayesian Segmentation was presented. The entire breast tissue including the pectoral muscle was clustered based on the similarity of intensity level. They used an eight-connected neighborhood to make the clusters. But the results of this method were not found satisfactory.

The work presented in [4] estimated the area where the pectoral muscle was based on some mathematical modeling. An iterative thresholding method was applied in that area to segment out the pectoral muscle. A gradient magnitude ridge traversal algorithm was proposed in [5] to segment the pectoral muscle region. The traversal algorithm was applied at small scale and then using voting scheme the resultant multiple edges were solved. A mammogram segmentation method based on breast density and detection of asymmetry was presented in [6]. The method they used for pectoral muscle segmentation was based on the technique proposed in [4]. They have improvised the method by extending the bottom end of the pectoral margin by drawing a straight line, parallel to the straight line which best fits the iteratively refined estimate instead of the line estimated initially as in [4]. Pectoral muscle segmentation method using histogram thresholding, edge detection, contour growing, and polynomial fitting was presented in [7].

A model-based segmentation method for removal of pectoral muscle is presented in [8]. In this method the image is divided into quarters and then an iterative thresholding based approach is applied only within the quarter where the

presence of the pectoral muscle is estimated. Hough transformation and different fitting methods were used to obtain the binary mask of the image. In [9] a seeded region growing method is proposed for pectoral muscle segmentation. Prior to segmentation, the images were preprocessed by removing noise, artifacts, and labels. Median filtering was used for removal of noise, artifacts and thresholding and morphological operations were used for removal of labels. The technique proposed in [10] was based on global thresholding approach for pectoral muscle removal. First the images were normalized to improve the contrast and then a DCT-based filter was applied for noise removal. A histogram based eight-connected neighborhood component labeling method was proposed in [11] to extract breast region and pectoral muscle removal. Gray-level normalization and median filtering were used to improve the image quality prior to the extraction method.

The paper is organized into five sections. In Sect. 2 a brief discussion on region growing technique is presented. In Sect. 3 a segmentation method for removal of pectoral muscle from MLO mammogram images using a region growing technique is proposed. Experimental results along with the description of the dataset used are presented in Sect. 4. Finally the paper is concluded in Sect. 5.

2 Region Growing

Region growing is one of the state of the art technique used for image segmentation. It is based on the principle of homogeneity that is at least one characteristic remains consistent for each and every pixel within a region. The method starts with a seed point and propagates through the connected points iteratively according to specific homogeneity criteria. This method is widely used in detection of mammographic masses.

Appropriate seed selection for region growing can help in acquiring high-level knowledge of image components. This property makes SRG very suitable for content-based image database applications [12].

A seeded region growing algorithm that takes multiple initial seed point was proposed in [13]. In this method, the SRG was controlled by the initial seeds chosen rather than tuning homogeneity parameter as in conventional region growing. The results of most of the region growing methods are highly sensitive to initial seed selection. A concept of symmetric region growing which is insensitive to initial seed selection was proposed in [14].

3 Proposed Method

The proposed method has mainly three phases viz. Phase I: defining a triangle over the mammogram image to isolate pectoral muscle, Phase II: seeded region growing and Phase III: refinement of the resultant image. Figure 1 depicts different stages of the proposed methodology. A mammogram is preprocessed before applying the

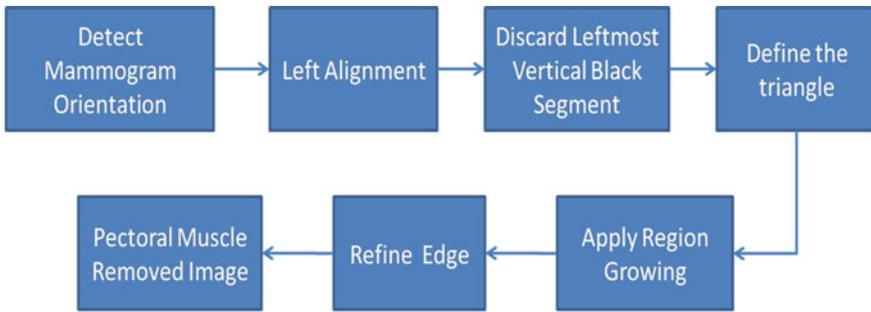


Fig. 1 Block diagram for pectoral muscle removal

steps of the proposed pectoral muscle removal method. The mammograms are either left aligned or right aligned. In the preprocessing step, the right mammograms are left aligned and the vertical black bar appearing in the leftmost part of the mammograms (in case of both left and right mammograms) are cropped out from the images. In phase I, a triangle is defined over the mammogram image that covers the pectoral muscle isolating the main breast region. In the second phase a seeded region growing algorithm is applied to suppress the pectoral muscle. This algorithm is applied only within the triangle defined in the first step. Last step is to refine the pectoral muscle suppressed image.

3.1 *Define Triangle to Isolate Pectoral Muscle from Breast Tissue*

In MLO images, the pectoral muscle appears either on left-top corner or right-top corner depending on the orientation of the mammogram. They have higher intensity values as compared to the main breast tissue. A method to remove pectoral muscle by using seeded region growing technique was proposed in [15]. In this technique, a triangular area was defined over the mammogram that covers the pectoral muscle and the seeded region growing was applied within that area only. But while defining the triangular area they did not considered the fact that for some images the pectoral muscle area may be spread out up to the lower boundary of the images.

Considering this fact we have redefined the triangular area in this work in such a way that it can cover the pectoral muscle in all cases. The triangle is defined as shown in Fig. 2. In the left aligned mammograms, the pectoral muscle appears on the left-top corner. The mammogram is cropped by removing the vertical black bar appearing on the leftmost side of the mammogram. The left boundary line of the cropped image gives two end points (A and B) of the triangle. The third end point, C is placed at a position on the top boundary of the image which separates the pectoral muscle from the breast tissue. The point C is chosen automatically based

Fig. 2 Triangle ABC that isolate the pectoral muscle from rest of the tissue



on the change in intensity differences. The point C is found by comparing the difference of intensity between neighboring pixels, as the pectoral muscle has higher intensity than the rest of the tissue. In this method instead of calculating first-order or second-order derivatives, we calculate the intensity difference between every i th and $(i + 5)$ th. It is observed that the intensity difference is much higher at the transition place. Experimentally we have found 20 as the optimum threshold.

A diagonal is drawn connecting the left-top corner and the right-bottom corner of the image intersecting the line BC at point D. It is observed that the area under the triangle ACD has intensity value much higher than the remaining parts of the triangular area ABC. Since the pectoral muscle possesses higher intensity value than the breast tissue, we consider the average pixel intensity of the region ACD as a parameter to set the selection criteria of pixels while performing region growing.

To draw the straight lines other than AB, we have used the Digital Differential Analyzer (DDA) line drawing algorithm. The algorithm works as given below. The equation of a straight line with slope m and y intercept at b can be expressed as given in Eq. (1).

$$y = mx + b \tag{1}$$

For a line with $|m| < 1$, k th value of x and y are calculated as given in Eqs. (2) and (3).

$$x_k = x_{k-1} + 1 \quad (2)$$

$$y_k = y_{k-1} + m \quad (3)$$

For a line with $|m| > 1$, k th value of x and y are calculated as given in Eqs. (4) and (5).

$$x_k = x_{k-1} + \left\lfloor \frac{1}{|m|} \right\rfloor \quad (4)$$

$$y_k = y_{k-1} \pm 1 \quad (5)$$

3.2 Removal of Pectoral Muscle

To eliminate the pectoral muscle we have applied a region growing method within the triangle ABC. To minimize the space complexity a non-recursive region growing method is applied. As we can see in Fig. 2 the pectoral muscle is approaching from line AB and stopping at line BC. So we have divided the regions within ABC logically into many subregions. Each horizontal line within ABC is considered as one subregion. The region growing starts from the leftmost point of a subregion and stops until the criteria specified in Eq. (6) is satisfied or reaches the end point of the subregion.

$$I_{(x,y)} > I_{avg} - (I_{lmax} - I_{avg}), \quad (6)$$

where $I_{(x,y)}$ is pixel intensity at position (x,y) in the image I , I_{avg} is the average intensity of the region ACD and I_{lmax} is the local maxima of the subregion.

3.3 Refinement of the Resultant Image

The method discussed in the previous section misclassified some of the image pixels during region growing. So we have devised a mechanism for further refinement of the resultant area. It is observed that the inner boundary of the pectoral muscle decreases smoothly from the topmost horizontal line towards the lower lines. By following the gradient of the inner margin of pectoral muscle, we can easily check if there is any ragged line and the same can be corrected accordingly.

4 Experimental Results

The proposed method to eliminate pectoral muscle was tested on the images collected from Mini-MIAS database [16]. This dataset consists of a total of 322 MLO mammogram images. The images have a spatial resolution of 200 μm per pixel and comprises of 1024×1024 pixels with 8 bits each. This database is publicly available to the researcher.

The proposed pectoral muscle suppression method is applied to 150 images of Mini-MIAS database. Segmentation accuracy for the said method is obtained by comparing resultant images with the respective hand-drawn segmentation masks. The segmentation accuracy is categorized into acceptable, partially acceptable, and unacceptable. The images for which the results obtained are almost close with the hand-drawn segmented images and are considered as acceptable. For some images, the region growing method misclassifies some of the image pixels, resulting in slight over segmentation or under segmentation; but does not affect the visual inspection of the abnormalities present in the image and hence may be used for further analysis. This set of results is considered as partially acceptable. The unacceptable results are the ones where there is complete mismatch in the result obtained from the proposed method and the hand-drawn segmented images. This category of images cannot be used for further analysis. We have achieved an 86.67% acceptable result and 5.33% partially acceptable result. This result is presented in Table 1.

The set of mammograms we have chosen for our experiment includes fatty, fatty glandular, and dense glandular tissues. The effect of the proposed method to fatty glandular mammogram mdb046 can be seen in Fig. 3a–c. Figure 4a–c shows the result obtained after suppressing the pectoral muscle from the fatty mammogram mdb028. Figure 5a, b shows the effect on dense glandular mammogram mdb058. The method shows a considerably better result over some of the technique discussed in this paper. In Table 2 a comparative analysis of different pectoral muscle segmentation techniques along with the proposed technique is presented.

Table 1 Categorical division of automatically segmented pectoral muscles from mini-MIAS images according to the segmentation acceptability

Category	No. of images	Percentage
Acceptable	130	86.67
Partially acceptable	8	5.33
Unacceptable	12	8

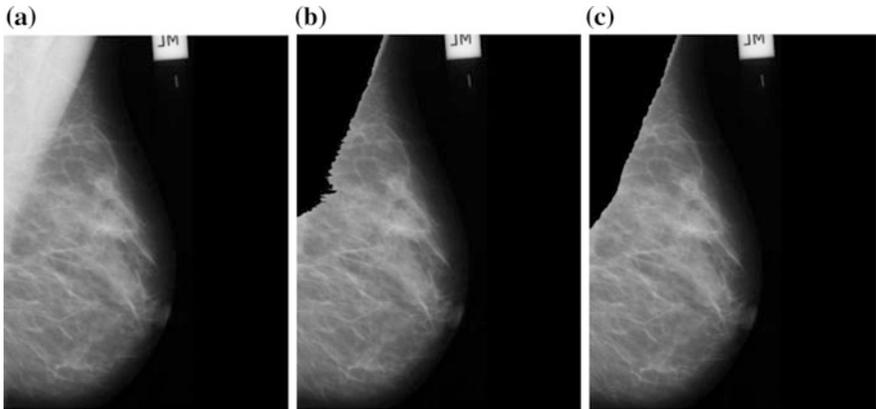


Fig. 3 a Original image mdb046, b mdb046 after suppressing pectoral muscle, c mdb046 after refinement

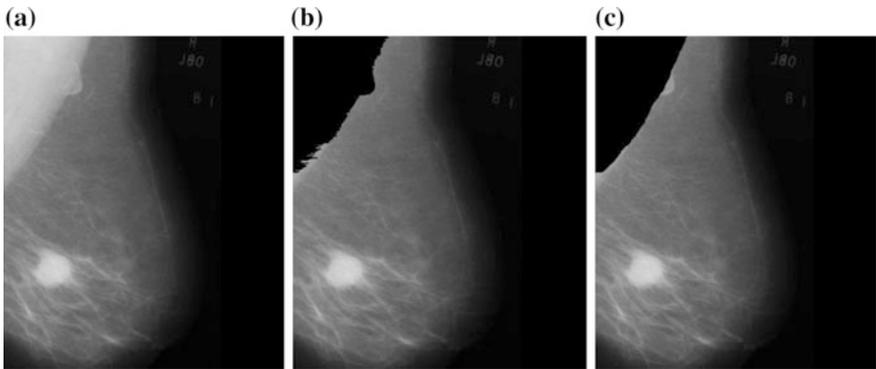


Fig. 4 a Original image mdb028, b mdb028 after suppressing pectoral muscle, c mdb028 after refinement

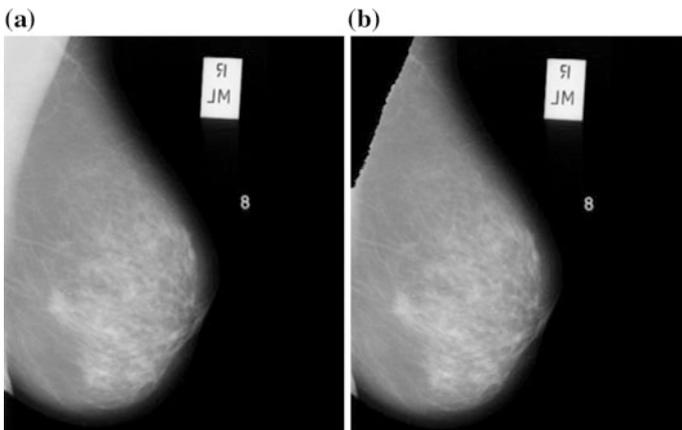


Fig. 5 a Original image mdb058, b mdb058 after suppressing pectoral muscle

Table 2 Comparative analysis of different segmentation approaches for pectoral muscle removal

Methods proposed by	Techniques used for segmentation	Number of images used	Segmentation accuracy (%)
[1]	Histogram based thresholding	322	86
[4]	Straight line estimation	322	83.9
[8]	Thresholding	60	81.7
[9]	Wavelet		85
[10]	Global thresholding, canny edge detection method and connected component labeling	161	90.06
[11]	Histogram based 8-connected neighborhood	322	89.5
Proposed	Region growing	150	92 (including partially acceptable result)

5 Conclusion

A seeded region growing method is presented to remove pectoral muscle from MLO mammogram images. The region growing is restricted within a well-defined triangular area which divides the image into two areas, one with pectoral muscle and one without pectoral muscle. Thus it helps in restricting region growing within the area that contains the pectoral muscle only. This triangular area defined over an image covers the pectoral muscle for almost all types of images unless there is absence of pectoral muscle. The method shows considerably good result for pectoral muscle removal.

References

1. David, R., Arnau, O., Joan, M., Marta, P., Joan, E.: Breast segmentation with pectoral muscle suppression on digital mammograms. *Lect. Notes Comput. Sci.* **3523**, 471–478 (2005)
2. Saltanat, N., Hossain, M.A., Alam, M.S.: An efficient pixel value based mapping scheme to delineate pectoral muscle from mammograms. In: *IEEE Fifth International Conference on Bio-Inspired Computing, Theories and Applications (BIC-TA)*, pp. 1510–1517 (2010)
3. Adel, M., Rasigni, M., Bourennane, S., Juhan, V.: Statistical segmentation of regions of interest on a mammographic image. *EURASIP J. Adv. Signal Process.* **2**, 1–8 (2007)
4. Kwok, S.M., Chandrasekhar, R., Attikiouzel, Y., Rickard, M.T.: Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. *IEEE Trans. Med. Imag.* **23**(9), 1129–1140 (2004)
5. Aylward, S.R., Hemminger, B.M., Pisano, E.D.: Mixture modeling for digital mammogram display and analysis. *Digit. Mammography, Comput. Imag. Vision* **13**(1), 305–312 (1998)
6. Tzikopoulos, S.D., Mavroforakis, M.E., Georgiou, H.V., Dimitropoulos, N., Theodoridis, S.: A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. *Comput. Methods Programs Biomed.* **102**(1), 47–63 (2011)

7. Zhili, C., Reyer, Z.: Segmentation of the breast region with pectoral muscle removal in mammograms. *Med. Image Understanding Anal.*, (MIUA). pp. 71–76 (2010)
8. Weidong, X., Shunren, X.: A model based algorithm to segment the pectoral muscle in mammograms. In: *International Conference on Neural Networks and Signal Processing*, IEEE, pp. 1163–1169 (2003)
9. Mustra, M., Bozek, J., Grgic, M.: Breast border extraction and pectoral muscle detection using wavelet decomposition. In: *EUROCON 2009, IEEE*, pp. 1426–1433 (2009)
10. Sreedevi, S., Elizabeth, S.: A novel approach for removal of pectoral muscles in digital mammogram. In: *International Conference on Information and Communication Technologies*, *Procedia Computer Science*, vol. 46, pp. 1724–1731 (2015)
11. Subash, C.B.R., Thangavel, K., Arul P.D.D.: Automatic mammogram image breast region extraction and removal of pectoral muscle. *Int. J. Sci. Eng. Res.* 229–35 (2013)
12. Fan, J., Elmagarmid, A.K.: An automatic algorithm for semantic object generation and temporal tracking. *Signal Process, Image Comm.* **17**, 145–164 (2002)
13. Adams R. and Bischof L.: Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 641–647 (1994)
14. Wan, S., Higgins, W.: Symmetric region growing. *IEEE Trans. Image Process.* **12**, 1007–1015 (2003)
15. Maitra I.K., Nag S., Bandyopadhyay S.K.: Technique for preprocessing of digital mammogram. *Comp. methods program med.* **107**, 175–188 (2011)
16. Suckling, J., Parker, J., Dance, D.R., Astley, S., Hutt, I., Boggis, C.R.M., Ricketts, I., Stamatakis, E., Cernaez, N., Kok, S.L., Taylor, P., Betal, D., Savage, J.: The mammographic image analysis society digital mammogram database. In: *Proceedings of the 2nd International Workshop on Digital Mammography*, pp. 375–378. Elsevier Science (1994)

Microstrip Patch Antenna: A Review

Arun Kumar Singh, Rabindranath Bera and Bansibadan Maji

Abstract Technology is getting advanced day by day. Antennas used in earlier days were generally larger in size. Now days in most of the applications, larger antennas are replaced by patch antennas. There are different shapes of patch antennas like triangular, circular, rectangular, fractal, etc. Each has its own characteristics, limitations, advantages, and disadvantages. Apart from this, the feed technique also has a great impact on performance of the antenna. In this paper a review is done on the basis of feed techniques, shape and the dielectrics used during the design process of the antenna.

Keywords Microstrip patch antenna (MPA) · Fractal antenna · Coaxial feed
Microstrip feed · Return loss

1 Introduction

In recent years, communication technology has reached to a new standard especially wireless communication. The size of the system has reduced drastically. Digital Signal Processing in the baseband of system has become stronger. In the RF domain the antenna size also has reduced and thus the system has become more portable. Instead of using larger antenna, Microstrip Patch Antenna (MPA) is now the first choice of any system designer. This kind of antenna is easy to fabricate

A.K. Singh (✉) · R. Bera

Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India
e-mail: arunsingh.smit@gmail.com

R. Bera

e-mail: rbera50@gmail.com

B. Maji

Department of Electronics and Communication Engineering, National Institute of Technology Durgapur, Durgapur, West Bengal, India
e-mail: bmajiecnit@yahoo.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_33

with involvement of very less capital. The MPA geometry can change as triangular, rectangular, circular etc. [2]. Out of these shapes rectangular patch antenna is considered as favorites because of ease of fabrication, analysis, and low cross polarization [3]. Initially the MPA was designed to operate in a single band. Later there was growing demand of multiple band antennas, i.e., antenna which can operate for several bands which leads the device to become capable to receive multiple services offered by different wireless technology networks like Bluetooth (2.4 GHz), WiMAX (3.5 GHz), WLAN (2.4 and 5.7 GHz).

The above-said application requires wider bandwidth, high speed data transmission. For best performance, high radiation efficiency, small size, and low return loss impedance matching are the pre-requisites for the designing of MPA.

A proper selection of patch parameter and position of feed will meet the desired system performance like resonating frequency, gain, and directivity [4]. The resonating frequency changes due to inductive loading provided by the feed position on the patch. The dielectrics used in the antenna also play a very important role in evaluating the performance of the antenna (Fig. 1).

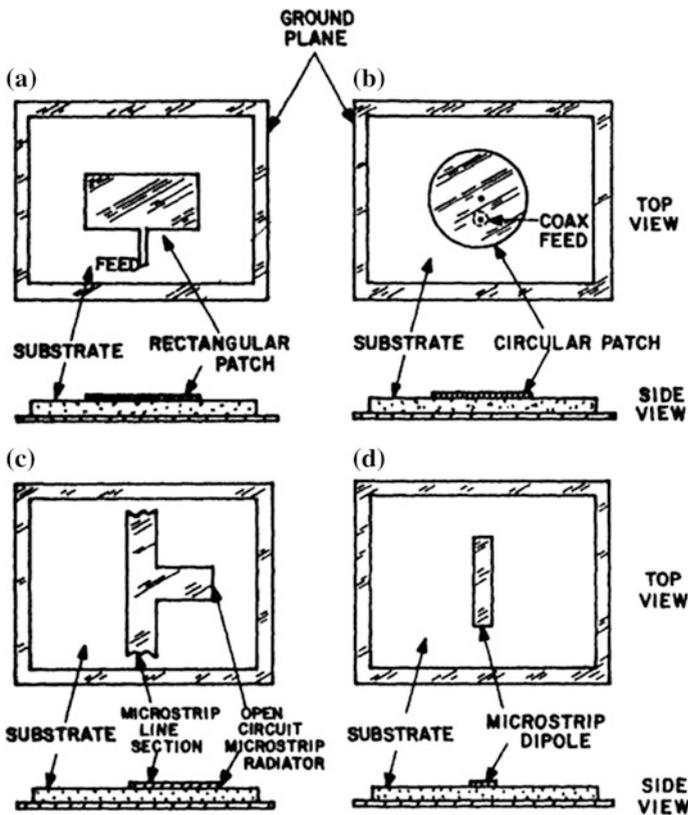


Fig. 1 a Rectangular MPA, b circular MPA, c open-circuit microstrip radiator, d microstrip dipole antenna [1]

2 Design Review of Microstrip Patch Antenna

In paper [1], MPA of leaf structures are discussed. These types of antennas have advantages like wider bandwidth, high radiation performance, small size, and multiple band operating frequency (Fig. 2).

These antennas were designed to operate in X Band (8–12 GHz) and Ku Band (12–18 GHz). Artificial neural network (ANN) is implemented in this paper for best optimization. The input parameters high, low resonance frequencies and their return losses and the output parameters were the patch dimension (D, R1, R2, W2). To optimize the design, backpropagation algorithm is used as an ANN tool (Figs. 3 and 4).

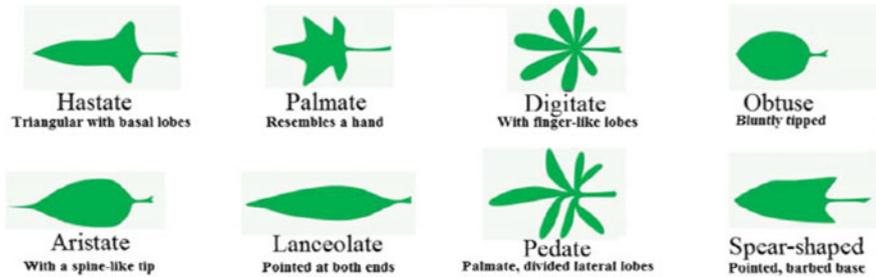
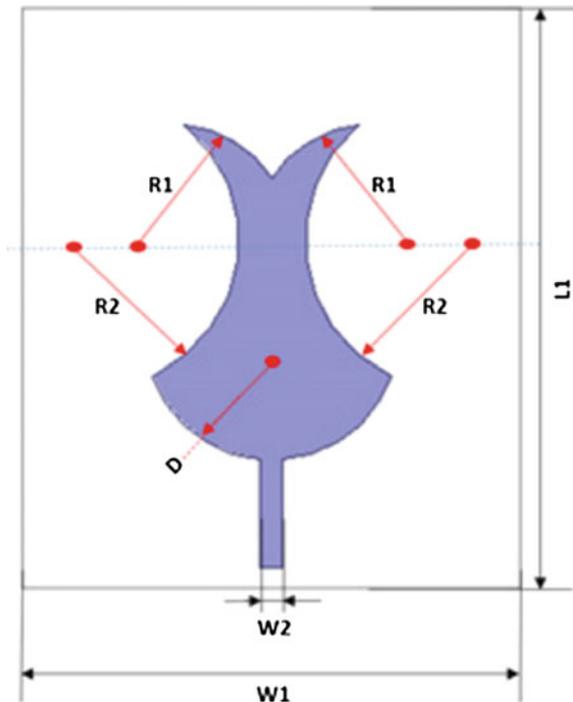


Fig. 2 Leaf-shaped MPA [1]

Fig. 3 Simulated design of tulip leaf shaped MPA [1]



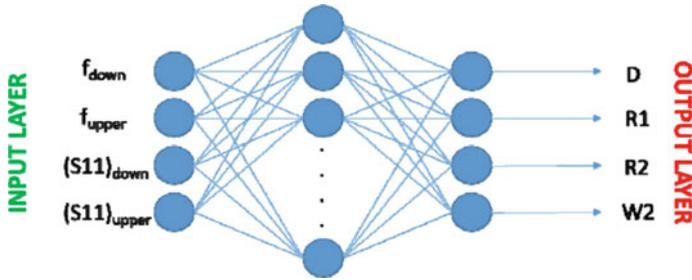


Fig. 4 ANN model [1]

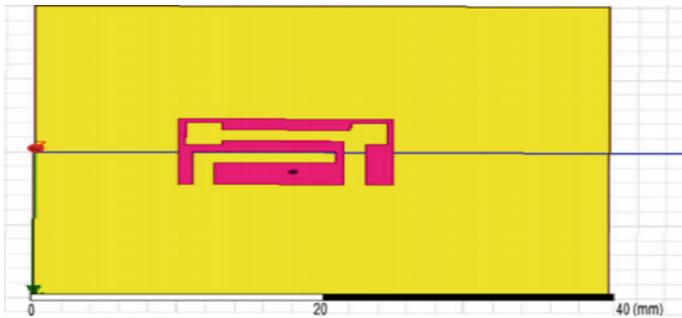


Fig. 5 Top view of triple frequency slot antenna [2]

However the designed discussed is only for single antenna. For better performance results the design can be extended towards array form.

In paper [2], slotted MPA operating in dual band is discussed. The operating frequencies are in ISM band (2.4 GHz), and C band (4.85 and 6.1 GHz) (Fig. 5).

In this paper, different slots are created so as to increase the bandwidth and to reduce the patch size. The feed type used is coaxial feed. The benefit of coaxial feed is the return loss is minimum with higher impedance matching. If gain of the antenna is to be increased the only way is to modify the design towards array and with coaxial feed is slightly difficult.

In paper [3], a triband microstrip patch antenna is being discussed. This type of antenna is used for mobile communication as well as for WLAN, WiMAX applications. Optimization tool viz. Genetic algorithm is used to optimize the antenna for better performance.

The number of bands in a particular patch depends upon the length and width of the slots. In (Fig. 6) design for the specified slot is shown and there are three bands created as shown in (Fig. 7).

However as discussed earlier for optimization, genetic algorithm tool is used. The flowchart for the search is as follows (Fig. 8).

The design shown in this paper is also for a single patch antenna. However the same can be extended to array patch antenna for better gain results.

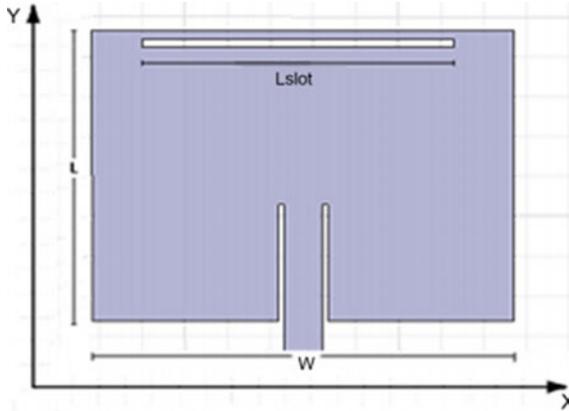


Fig. 6 Microstrip patch slotted antenna [3]

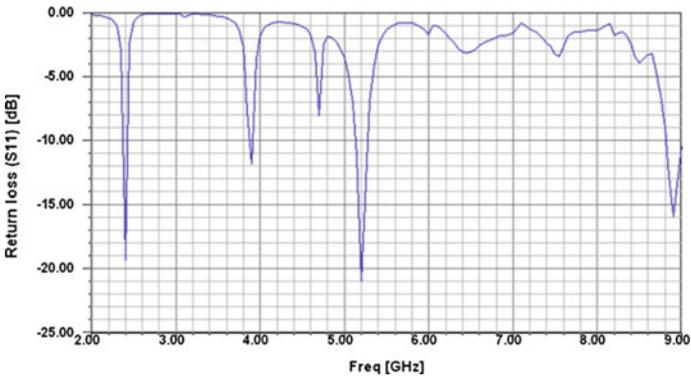


Fig. 7 Return loss at three bands

In paper [4], a unique design of duo triangle shape patch antenna for lower band application is shown. Both the triangles are derived from a single patch. The proposed geometry is shown in (Fig. 9).

Using this structure the size is reduced, gain, and bandwidth is increased. The antenna efficiency is tremendously increased to 90%.

In paper [5] a slotted antenna of swastika shape is discussed. The reason behind this geometry is increase in bandwidth, dual band nature, and overall size reduction. The geometry is shown in (Fig. 10).

Where $Sub_w = 21\text{ mm}$, $Sub_L = 21\text{ mm}$, $Sub_h = 12.5\text{ mm}$, $l = 3\text{ mm}$.

The proposed antenna provides a better performance with high bandwidth, low return loss, and good gain. The efficiency of the antenna in terms multiband operation and cost makes it superior over the conventional design.

Fig. 8 Genetic algorithm embedded with HFSS framework

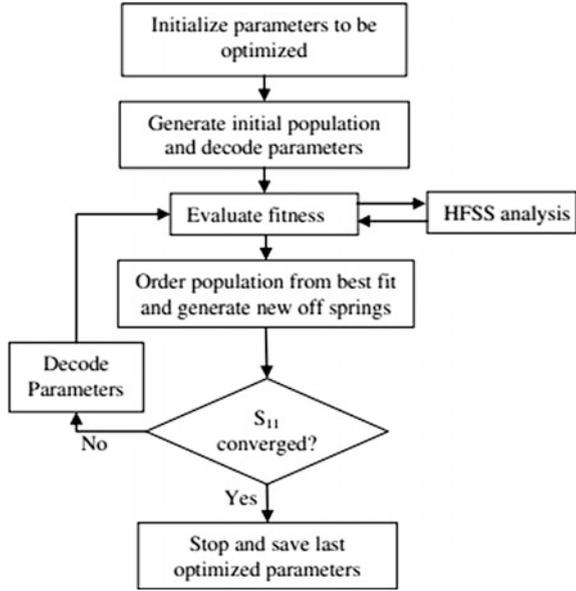
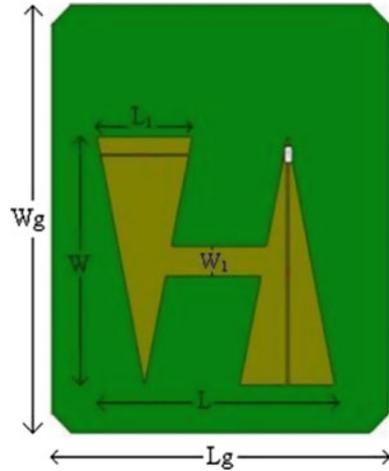


Fig. 9 Duo triangle-shaped patch



In paper [6], patch antenna in biomedical field is discussed. It shows a wearable patch antenna. Devices such as pacemakers, drug delivery system, heart rate monitors which are implanted within the body needs to communicate with external world via base station devices. This kind of antenna will help in those cases where a continuous monitoring is required even when the patient is not within the hospital.

For this ISM bands are used to provide high bandwidth communication link. The same design can be extended towards array design for enhance gain.

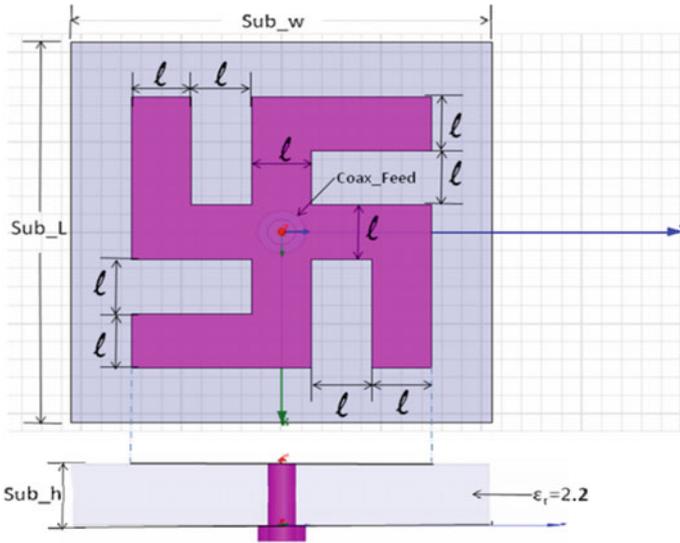
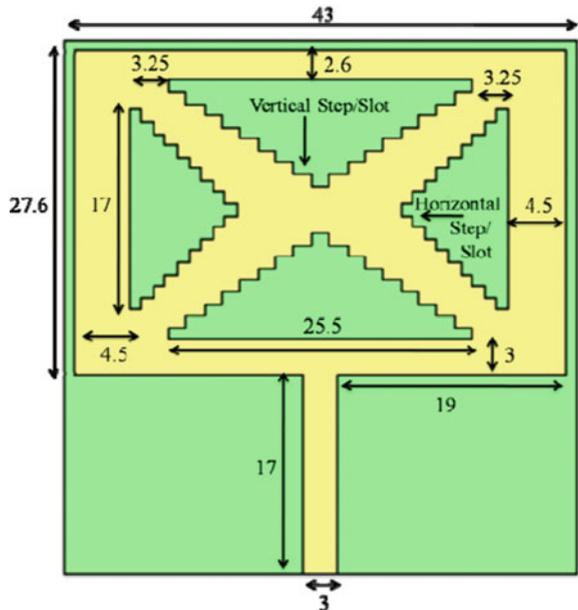


Fig. 10 Swastik-Shaped patch geometry

Fig. 11 Chebyshev patch antenna



In paper [7], multiband patch antenna is designed with the help of Chebyshev on each side. This features helps the antenna to have multiband with wide band nature. The geometry is shown in (Fig. 11).

It was found that the designed antenna increases the bandwidth and the feed line movement along the Chebyshev is used for controlling the frequency shifts. This kind of antenna found its application majorly in the field of GSM, WiFi, and WiMAX.

3 Conclusion

In this paper review on design of patch antenna is done. The parameters of review were size of the patch, return loss (S_{11}) parameter, gain, bandwidth and efficiency. Different geometry of antenna has different impact on its performance. However in none of the paper array design was discussed. The performance in terms of gain will further increase if array design is used.

References

1. Ozkaya, U., Seyfi, L.: Dimension optimization of microstrip patch antenna in X/Ku band via artificial neural network. In: World Conference on Technology, Innovation and Entrepreneurship, pp. 2520–2526 (2015)
2. Bhavanam, S.N., Kalyan, R.S.: Design of a novel coaxial feed triple frequency patch antenna with slots and shorting pin. In: International Conference on Computational Modeling and Security (CMS 2016), pp. 345–351 (2016)
3. Sami, G., Mohanna, M., Rabeh, M.L.: Tri-band microstrip antenna design for wireless communication applications. In: National Research Institute of Astronomy and Geophysics, NRIAG J. Astron. Geophys. 39–44 (2013)
4. Srivastava, S., Singh, V.K., Singh, A.K., Ali, Z.: Duo triangle shaped microstrip patch antenna analysis for WiMAX lower band Application. In: International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA), pp. 554–563 (2013)
5. Patel, S., Kosta, Y., Soni, H., Patel, S.: Wideband miniaturized patch antenna design and comparative analysis. AIM 2011, CCIS 147, pp. 111–116 (2011)
6. Rucker, D.G., Khaleel, H.R., Raheem, S.S., Al-Rizzo, H.M.: Microstrip antenna arrays for implantable and wearable wireless applications. In: MobiHealth 2010, LNICST 55, Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 135–143 (2011)
7. Abbas, S.M., Saleem, I., Parveen, A., Zahra, H., Khan, S.A.: Chebyshev multiband patch antenna. In: IMTIC 2012, CCIS 281, pp. 10–19 (2012)

A Comparison Study of Face, Gait and Speech Features for Age Estimation

Prachi Punyani, Rashmi Gupta and Ashwani Kumar

Abstract With the growing importance of age estimation in the recent years, Researchers have been trying to use different human body biometrics to estimate the age of a person. Face, gait and speech are the three main biometric traits which have been reported to investigate the human age successfully. Each feature has specific characteristics which employ the prediction of age. Like Wrinkles, skin and shape of the face; speed, head to body ratio and height of the gait; and pitch and heaviness of the speech define the baselines for the age estimation. We have compared these three features and evaluated their performance. Conventional techniques have been used from the literature and experimental results are compared in terms of MAE and accuracy. Face is found to have most detailed features to predict the age and hence minimum mean absolute error of 5.36. It is followed by gait and then speech which are found to have mean absolute error of 6.57 and 6.62 respectively.

Keywords Human age estimation • Biometrics • Face • Gait • Speech Comparison

1 Introduction

Soft Biometrics [1] has emerged as a new topic in the field of biometrics which includes the estimation of demographic characteristics from the biometric organs. Soft biometrics is defined as the estimation of age, gender, ethnicity, eye-, hair-,

P. Punyani (✉) · A. Kumar (✉)
Indra Gandhi Delhi Technical University for Women, Kashmiri Gate, Delhi, India
e-mail: prachipunyani90@gmail.com

A. Kumar
e-mail: drashwnikumar@gmail.com

R. Gupta (✉)
Ambedkar Institute of Advanced Communications and Research,
Shastri Park, Delhi, India
e-mail: rashmig71@yahoo.com

skin colour and other biological details from body parts like face, gait, fingerprints, palm prints. It may also be estimated from acoustic waves like speech patterns.

Evolution of this field comes from the growing online market. Both customers and sellers demand the need of a personalized shopping experience where product choices are shown to the customers according to their age group interests and gender category. Moreover, we can use the technology of age prediction at entrance doors in order to prevent the below 18 children to watch adult movies or from buying alcohol bottles. Age and gender estimation is an alternative way of performing security identification and verification tasks when just biometric recognition does not solve the purpose.

Till date, work has been done to estimate the age and categorize the gender based on face, gait and speech only. In this paper, we compare these three techniques of age estimation and report the results. Widely used state of the art techniques are used for the implementation of each technique and comparisons are made in terms of mean absolute error (MAE).

Face has various features which modify with age. Like size of face is different for a child, young and adult. Similarly, a child face does not have wrinkles, while an adult face is full of wrinkles. Age estimation models based on these features helps in a better way of human-computer interaction for online marketing, visual surveillance and criminal investigation. There are certain situations when the person is a distance from the camera and face image could not be taken clearly. Gait-based age estimation plays the role in such situations. Herein, we exploit the fact that walking speed and even height is different for different age groups. Research has also been conducted on speech sounds of different age groups to estimate the age. Acoustic waves produced by males and females have different pitch. Females create high pitched acoustic waves and males produce low pitched acoustic waves. Moreover, children have high pitched voice and adults have heaviness in their voice. Hence, speech-based age estimation is also reported.

We compare these three techniques and analyze their effectiveness for the purpose. Face-based age estimation is found to have the best Recognition accuracy for finding age. Gait and speech also have comparable accuracy. Different techniques are used depending on the purpose and the conditions given. Figure 1 shows the basic block diagram of age estimation process.

Remainder of the paper explains the literature work in Sect. 2 and methodologies used in Sect. 3. Experimental results, conclusion and references are described in Sects. 4, 5 and 6 respectively.

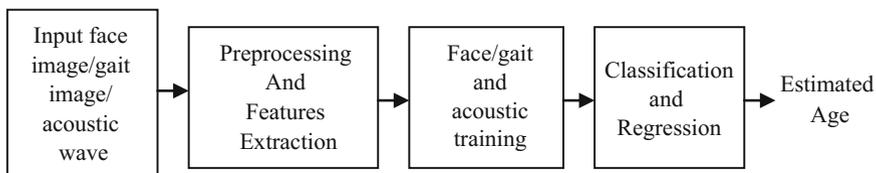


Fig. 1 Basic block diagram of age estimation model

2 Literature Review

Although Age estimation is a very latest research topic, many scientists have successfully researched on the ways of estimating age using three main biometric traits like face, gait and speech. Biological changes with age in Fingerprints, palm-print and iris could also be utilized to judge the human age, but research is yet to be done in this shell.

With the growing age, our face changes in terms of shape, skin and wrinkles. These changes have defined various models defined in literature to estimate the features using facial characteristics of a person. These models are named as Anthropometric model [2], aging manifold analysis [3, 4], Aging Pattern Subspace [5], Active appearance model [6, 7] and Local Binary Patterns [8]. From a feature representation, final age estimation is described by classification [7], regression [6–9] or Hierarchical method [10]. Each model is based on different details acquired from a human face which may be shape, texture or body part landmark distances.

Gait based age estimation is appropriate when target person and camera are at a distance. Resolution of facial age estimation decreases as the distance of the person from the camera increases, and hence gait based age estimation becomes relevant. A few works have been done by scientists to estimate the age using gait features as listed in [11, 12]. Gait based age estimation depends on the differences in stride frequency, head to body ratio, head size, height and walking speed of different age groups.

Pitch of speech sounds is different for both males and females. Moreover, speech rate, pitch, zero crossing, features of short time energy are different for different age groups. Researchers have done a few works to estimate the age of a person using speech models in [13, 14]. These works mainly include age group classification rather than the estimation of exact age. Some of the acoustic feature models are proposed in [15, 16] and classification algorithms are detailed in [17, 18].

3 Methodologies Used

We have used state of the art methodologies estimate the age from the database. First step involves the extraction of features, which is followed by age estimation and classification process to indentify or verify the age. We have paragraphed below the methodologies used for face, gait and speech each for the purpose of comparison.

3.1 Face

To classify the age using face images, we have used the approach utilized by Elhocine Boutellaa, Abdenour Hadid, Messaoud Bengharabi and Samy Ait-Aoudia

[19]. Work is tested on FG-NET face database [20] which is widely used for experiments that need age information. Firstly, face images are preprocessed by properly orienting and cropping them for proper analysis. Next step is to extract the facial features. Binarized statistical image features (BSIF) are considered for encoding texture and shape features of the face. Finally after calculating the age, classification is performed using SVM classifier having an RBF kernel.

3.2 *Gait*

Gait images are investigated using USF gait database [21] available publically. Knowledge given by Jiwen Lu and Yap-Peng Tan [12] is implemented to get the results. The sequence of steps followed includes features extraction using gait energy image (GEI), label encoding approach to encode the decimal image into a binary value, dimensionality reduction using Multi-label guided subspace learning model which helps in correlating age and gender information for better age estimation. Final step explains the recognition process using ML-KNN and label decoding to get the final identified age.

3.3 *Speech*

Age estimation using speech sounds is reported on NIST SRE 2010 telephony test set [23, 24] by Yorktown heights using senone posterior based I-vectors [22]. Acoustic features are extracted using I-vector representation. Linear dimensionality reduction is performed to reduce the dimensions and select only the most relevant feature subset for the task of age estimation. Classification is done using Support vector regressors (SVR) model. Interested readers can go through the detailed procedure given in [22].

4 Experimental Results

Face-based, gait-based and speech-based age estimation is performed and results are reported in terms of mean absolute error (MAE). MAE is defined by following formula:

$$\frac{1}{N_t} \sum_{i=1}^{N_t} |A - \tilde{A}| \quad (1)$$

Table 1 Comparison of face, gait and speech for age estimation in terms of MAE

Method	MAE
Face	5.36
Gait	6.57
Speech	6.62

Here A and \tilde{A} are the true and estimated ages respectively, N_t is the total number of test images. Hence it is defined as the average of all the differences between the true and the estimated ages.

We have used FG-NET face database [20], USF gait database [21] and NIST SRE 2010 speech database [23, 24] for the studies which are publically available online and include age information. Table 1 provides the comparison results for age estimation based on face, gait and speech in terms of MAE. Results validate that face-based age estimation shows minimum MAE of 5.36 and hence the best way to describe the age of a human. However, Gait-based age estimation is minutely better than speech-based age estimation, but the difference is very less. The kind of technique used for the estimating age hence depends more on the input situations and the purpose of the task.

Figure 2 presents a bar graph for recognition accuracy of each technique. We have tested the results for three main age groups individually for better validation of the results. These three age groups under test are child (0–18 years), young (>18–50 years) and adults (>50–80 years). In each of these age groups, face-based age estimation highlights better accuracy than gait-based and speech-based methods.

Lastly, Basis for the technique used for age estimation task is not only defined by MAE but also depending on the input situation and the application for which it is used. Hence, we enlist the applications of each technique in Table 2 of the paper.

Fig. 2 Recognition accuracy of face, gait and speech for child, young and adult (in percentage)

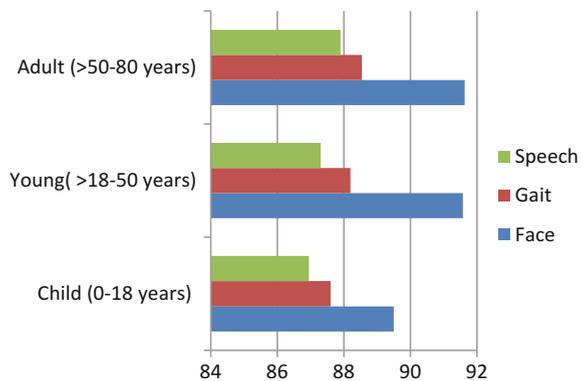


Table 2 Comparison of applications of face, gait and speech for in terms of age estimation

Method	Applications
Face	Criminal investigation, visual surveillance, personalized e-commerce experience based on age group, access control, biometric identification and verification
Gait	Long distance visual surveillance, airport, railway and building entries, automatic vending machines of cigarette and alcohols, online marketing
Speech	Natural interaction with the dialogue systems, target advertising based on age, forensic studies, user characterization and pairing of caller and agent in call centers

5 Conclusion and Discussion

The paper compares the three main researched ways for age estimation named as gait-based, face-based and speech-based age estimation. Conventional techniques are used for experimentation. Comparisons are reported in terms of mean absolute error (MAE) and Recognition accuracy percentage. Face-based age estimation is seen to have best performance with minimum MAE of 5.36. Gait-based and speech based methods are found to have MAEs of 6.57 and 6.62 respectively which are though less than that of face-based age estimation but comparative to each other.

We also analyze that the type of method used depends not only on the performance metrics but also the application or the purpose for which it is used. Face-based methods are generally applied when face images are close to the camera and gait is used when camera is quit away from the face image and face could not be seen with high resolution. When only speech sounds are available, then speech-based methods come into role.

In this paper, a novel algorithm has been used to compare face, gait and speech based features. We can extend this work to a different level and can think of optimizing the performance of age estimation systems further by using fusion of any two or all three ways of age estimation. Fusion can be performed at features level, score level or decision level. The decision of the level used for fusion will depend on the algorithm used and its compatibility at various fusion levels.

References

1. Dantcheva, A., Elia, P., Ross, A.: What else does your biometric data reveal? A survey on soft biometrics. *IEEE Trans. Inf. Forensics Secur.* **11**(3), 441–467 (2015)
2. Kwon, Y.H., Loba, N.V.: Age classification from facial images. *Comput. Vis. Image Underst.* **74**(1), 1–21 (1999)
3. Guo, G., Fu, Y., Dyer, C., Huang, T.S.: Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Process.* **17**(7), 1178–1188 (2008)
4. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimedia* **10**(4), 578–584 (2008)
5. Geng, X., Zhou, Z., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2234–2240 (2007)

6. Lanitis, A., Taylor, C.J., Cootes, T.F.: Towards automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 442–455 (2002)
7. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. *IEEE Trans. Syst. Man Cybern. Part B* **34**(1), 621–628 (2004)
8. Bereta, M., Karczmarek, P., Pedrycz, W., Reformat, M.: Local descriptors in application to the aging problem in face recognition. *Pattern Recogn.* **46**, 2634–2646 (2013)
9. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimedia* **10**(4), 578–584 (2008)
10. Han, H., Otto, C., Liu, X., Jain, A.K.: Demographic estimation from face images: human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(6), 1148–1161 (2015)
11. Lu, J., Tan, Y.P.: Gait-based human age estimation. *IEEE Trans. Inf. Forensics Secur.* **5**(4), 761–770 (2010)
12. Chuen, B.K.Y., Connie, T., Song O.T., Goh, M.: A preliminary study of gait-based age estimation techniques. In: *Proceedings of APSIPA Annual Summit and Conference 2015*, pp. 800–806 (2015)
13. Schuller, B., Steidl, S., Batliner, A., Burkhardt, F., Devillers, L., Muller, C., Narayanan, S.: Paralinguistics in speech and language—state-of-the-art and the challenge. *Comput. Speech Lang.* **27**(1), 4–39 (2013)
14. Dobry, G., Hecht, R., Avigal, M., Zigel, Y.: Supervector dimension reduction for efficient speaker age estimation based on the acoustic speech signal. *IEEE Trans. Audio Speech Lang. Process.* **19**(7), 1975–1985 (2011)
15. Spiegl, W., Stemmer, G., Lasarczyk, E., Kolhatkar, V., Cassidy, A., Potard, B., Shum, S., Song, Y.C., Xu, P., Beyerlein, P., Harnsberger, J.D.: Analyzing features for automatic age estimation on cross-sectional data. In: *Proceeding of INTERSPEECH*, Brighton, UK, pp. 2923–2926 (2009)
16. Schlitz, S., Muller, C.: A study of acoustic correlates of speaker age. In: Muller C. (ed.) *Speaker Classification II*, ser. *Lecture Notes in Computer Science*, vol. 4441, pp. 1–9. Springer, Berlin, Heidelberg (2007)
17. Safavi, S., Russell, M., Jančovič, P.: Identification of age-group from children speech by computers and humans. In: *Proceeding of INTERSPEECH*, Singapore, pp. 243–247 (2014)
18. Kockmann, M., Burget, L., Cernocký, J.: Brno university of technology system for Interspeech 2010 paralinguistic challenge. In: *Proceeding of INTERSPEECH*, Makuhari, Japan, pp. 2822–2825 (2010)
19. Boutella, E., Hadid, A., Bengherabi, M., Ait-Aoudia, S.: On the use of Kinect depth data for identity, gender and ethnicity classification from facial images. *Pattern Recogn. Lett.* **68**, 1–8 (2015)
20. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 442–455 (2002)
21. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The human ID gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(2), 162–177 (2005)
22. Sadjadi, S.O., Ganapathy, S., Pelecanos, J.W.: Speaker age estimation on conversational telephone speech using senone posterior based I-Vectors. *ICASSP*, pp. 5040–5044 (2016)
23. Brandschain, L., Graff, D., Cieri, C., Walker, K., Caruso, C., Neely, A.: Mixer 6. In: *Proceeding LREC*, Valletta, Malta (2010)
24. Cieri, C., Corson, L., Graff, D., Walker, K.: Resources for new research directions in speaker recognition: The Mixer 3, 4 and 5 corpora. In: *Proceeding of INTERSPEECH*, pp. 950–953, Antwerp, Belgium (2007)

A Review on MIMO Based V2V Communication

Samarendra Nath Sur, Rabindranath Bera and Bansibadan Maji

Abstract This paper provides a review on the various channels for vehicular application and utilization of multiple input and multiple output (MIMO) to improve the reliability of communication system for the application in intelligent transportation system (ITS). We also present some key research areas related to MIMO in vehicular ad hoc networks (VANETs) and possible signal processing tools to encounter the severe dynamic vehicular channel.

Keywords MIMO · ITS · VANET · V2V · V2I

1 Introduction

From the point of view of road safety, driver protection, on the way communication etc., vehicle-to-vehicle (V2V) communication researches have gain significant importance in research world. And also today's advancement in communication technologies are moving towards the exchange of high data rate multimedia information.

Vehicular ad hoc networks (VANETs) is basically designed to exchange the dedicated information between the vehicles. VANET can be categorized into two forms: vehicle to vehicle V2V and vehicle to infrastructure V2I [1]. The main motive of the use of VANETs technology is to the development of robust and

S.N. Sur (✉) · R. Bera

Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar 737136, Sikkim, India
e-mail: samar.sur@gmail.com

R. Bera

e-mail: rbera50@gmail.com

B. Maji

Department of Electronics and Communication Engineering, National Institute of Technology Durgapur, Durgapur, West Bengal, India
e-mail: bmajiecnit@yahoo.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_35

333

secure communication system [2, 3] like multi input multi output (MIMO) system to provide satisfy, less road traffic congestion, accident warning, road exploration, etc. [4] (Fig. 1).

In contrast to the cellular channel, V2V channels are highly dynamic and severe due to high mobility of the transmitter (Tx)/receiver (Rx) system [5, 6]. And this dynamic behavior of the channel puts lots of challenges towards the development efficient channel model and reliable physical communication system model. MIMO systems have also attracted considerable attention as an efficient solution to meet the increasing demands of high data rate, spectral efficiency and reliability in wireless communications. Therefore, there are enough opportunity to utilize the advantages of MIMO in highly dynamic vehicular channel condition.

This paper deals with the basic of vehicular channel modelling, understanding the dynamics nature of V2V scenarios. As further advancement, MIMO channel modelling for vehicular application and utility of MIMO signal processing for in V2V scenarios.

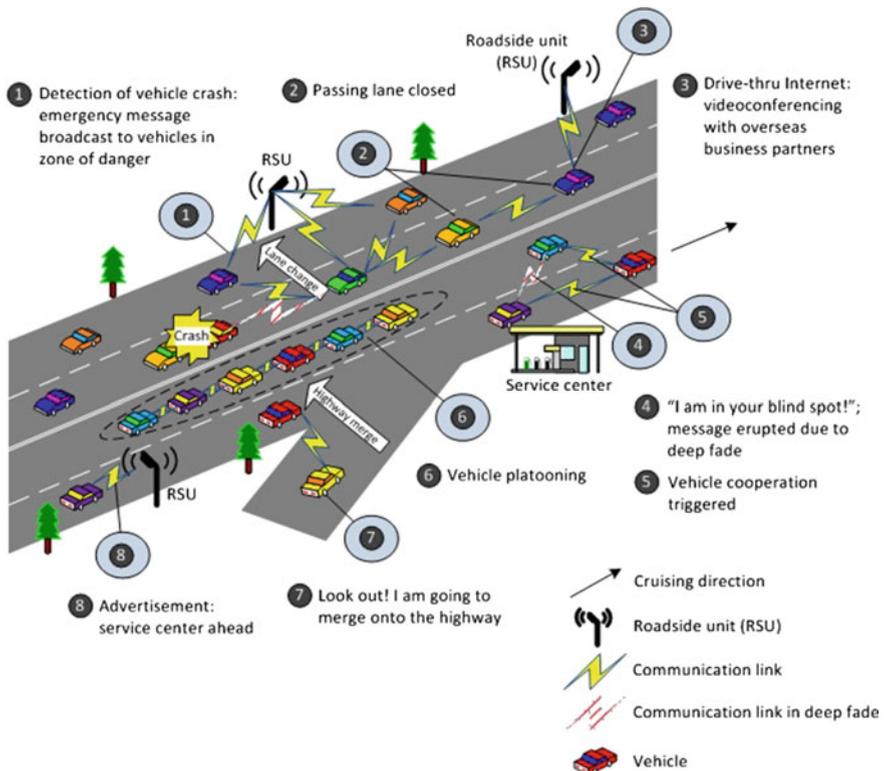


Fig. 1 Typical road situation [5]

2 VANETs Standard

VANETs are created for exchanging data wirelessly in case of inter-vehicle communication (IVC). The country responsible for the standardization of VANETs are USA, Australia, and Japan etc. [7]. In 1992 [6, 8], Dedicated Short Range Communication (DSRC) technology was recognized by FCC for VANET. DSRC is having licensed band of 5.9 GHz to support ITS. The DSRC spectrum is consist of seven 10 MHz channels. Out of which one channel for controlling signal and six for service related. Further, in 2004, IEEE 802.11p [9] was proposed for possible application in ITS. The purpose of the IEEE 802.11p standard is to enhance the system to communicate in rapidly changing environment [9, 10]. It uses 75 MHz of the licensed bandwidth, from 5.85 to 5.925 GHz, as part of ITS for DSRC in the USA [11].

3 Vehicular Channel

The modeling of wireless channel needs proper understanding of the statistical behavior of the channel [12, 13]. It can be classified into two categories: deterministic and stochastic. In terms of computational complexity, statistical modeling is much simpler than its counterpart [4].

A channel can better be represented keeping multipath propagation in consideration. A channel can be represented by channel impulse response (CIR) and can be characterized by channel coherence bandwidth (B_c), coherence time (T_c), delay spread, Doppler spread, path loss characteristic etc. [14, 15]. Similarly, in case of a stochastic wireless channel modeling, it can be characterized by autocorrelation and power spectral density (PSD) functions. Practically used model in cellular system is wide sense stationary (WSS) but its assumption does not hold for vehicular channel [4, 15, 16] (Fig. 2).

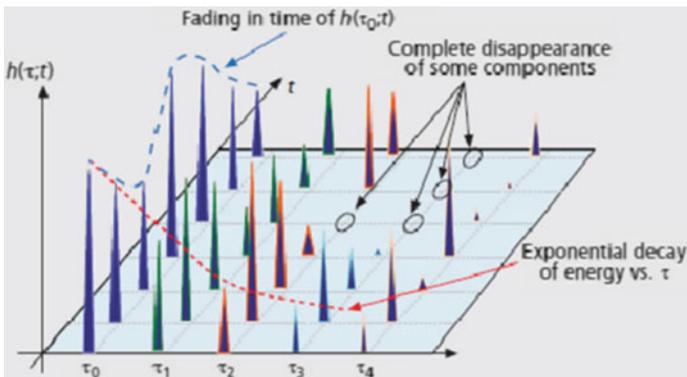


Fig. 2 Time variation of the CIR [15]

There is a tremendous growth in the field of V2V and V2I research [4, 17]. Lots of research works are going on for the designing of efficient channel model. But vehicular channel model is different from the typical wireless cellular channel model and that makes it more interesting. As V2V channel is dynamic in nature, there may exist line of sight (LOS) between Tx and Rx or may not. And also channel variation rates, can be larger than cellular system, which makes the channel severely frequency selective channel. Along with the multiple scattering, the V2V channel behavior becomes much more severe in comparison to normal Rayleigh fading model. In case of a V2V channel, along with other parameters, Angles of Arrival (AoA) and Angles of Departures (AoD) of the waves, which implicates scattering function. The statistical modeling for vehicular channel depends on the scenario such as highway, rural streets, suburban streets, urban streets etc. [1, 18–23].

4 Geometry-Based MIMO Channel

As it is already discussed in previous section, the common ways to develop a vehicular channel model are statistical and deterministic channel model. Over past few years lots of advanced channel model for V2V communication have been proposed and studied in details. Such kinds of advanced models are categorized as geometry-based deterministic models (GBDMs) (ray-tracing models) [24] and stochastic models, which are further categorized as non-geometrical stochastic models (NGSMs) [25] and geometry-based stochastic models (GBSMs) [6, 26, 27]. GBDM is basically 3D ray-tracing approach for propagation model. On a simple word it basically simulates the possible paths between the transmitter and receiver based on certain geometric rules. Whereas in case of NGSM, it does not depend on any geometric consideration. A GBSM is modeled based on the predefined stochastic distribution of scatterers by applying the fundamental laws of wave propagation. From the point of view of implementation of MIMO in V2V communication, Molisch and other co-authors have proposed a channel model, namely Geometry-based Stochastic MIMO Channel Model (GSCM) [4, 27–29]. Beauty of this model is that, here, scatterers are placed randomly according to some statistical distributions and which makes the model more realistic [4]. Boban et al. proposed Geometry-based efficient propagation Model for V2V communication (GEMV²) [30, 31]. Characteristic of this model is that it uses different types of path loss models for LOS and non-LOS conditions [31, 32].

5 MIMO Technologies for VANET

As MIMO system is emerging as a powerful tool to provide more appropriate solution for future generation communication system. MIMO systems provide higher capacities and reliable communication. It provides several advantages including

improved coverage area, higher throughputs, and security [33]. But VANET system required high data rate at high quality of service (QoS) [3, 33] and that is the most important challenge for the researchers to build suitable MIMO signal processing algorithms. Challenges that need to be addressed are channel modeling, space time signal processing, efficient channel coding and low complex receiver design etc.

The MIMO diversity techniques introduced with space-time block coding (STBC) and cyclic shift diversity (CSD) which extend the range and improve the signal quality. On the other hand, receiver diversity techniques like maximum ratio combining (MRC) have also been implemented to improve the system performance [34].

Another category of MIMO is the spatial multiplexing. For high data rate communication, spectral efficiency is also an important performance evaluation parameter for MIMO. In this regards, vertical Bell Laboratories Layered Space-Time (V-BLAST) is a spatial multiplexing MIMO systems to achieve much higher spectral efficiencies than SISO systems, in rich scattering environments [35–38].

In dispersive channel, optimum MIMO detectors namely ML [37] receiver is designed for combating both ISI and ICI MIMO system. But it is practically impossible to implement in hardware. Simple linear detection scheme like zero-forcing (ZF) or minimum mean square error (MMSE) algorithms are result in performance degradation [39–42]. In context to dispersive channel, Decision feedback equalization (DFE) [43–47] has the potential to achieve the information-theoretic performance bounds for both fading and static MIMO channels. Therefore an adaptive or iterative nonlinear MIMO-DFE based on the least squares algorithm has been studied to suppress the ISI and ICI.

There are some other variant of MIMO detectors with design flexibility, including the successive interference cancellation (SIC) detector, the parallel interference cancellation (PIC) detector, the multistage interference cancellation (MIC) detector have also been investigated [37, 48–50]. And also blind equalization is also there. All these receives comes at a cost of receiver complexity and these indeed increase the power consumption of the system. Hence low complexity and high-performance suboptimum MIMO detection algorithms are needed for practical MIMO applications.

Development of low complex and scalable receivers for MIMO systems have gained lots of attention. Different types of advanced receiver systems have been proposed, such as, tree-search based MIMO detection, the probabilistic data association (PDA) based MIMO detection, sphere decoder (SD), lattice-reduction (LR) based MIMO detectors [51–56] and the semi-definite programming relaxation (SDPR) based MIMO detection. This kind of low complex receivers make the massive MIMO or Large scale MIMO to a realizable and practically implementable solution for future generation communication system.

6 Conclusion

This article has provided a brief survey of V2V channel model and utilization of MIMO for the betterment of communication in highly dynamic channel condition. Here we have discussed some important parameters for modeling a vehicular channel and also discussed some advanced channel model like GBDMs, NGSMs, GBSM, GSCM and GEMV². Finally, we have discussed some advanced MIMO signal processing techniques to encounter severe dynamic channel related problem.

References

1. Boban, M., Barros, J., Ozan, K.T.: Geometry-based vehicle-to-vehicle channel modeling for large-scale simulation. *IEEE Trans. Veh. Technol.* **63**(9), 4146–4164 (2014)
2. Wang, S.Y.: The effects of wireless transmission range on path lifetime in vehicle-formed mobile ad hoc networks on highways. In: *Proceedings of the IEEE International Conference on Communications*, vol. 5, pp. 3177–3181. Seoul, Korea (2005)
3. Jonathan, L., Boeglen, H., Anne-Marie, P., Benoit, H., Rodolphe, V.: A semi-deterministic channel model for VANETs simulations. *Int. J. Veh. Technol.* **2012**, 1–8 (2012)
4. Boeglen, H., Benoît, H., Pascal, L., Jonathan, L., Anne-Marie, P., Rodolphe, V.: A survey of V2V channel modeling for VANET simulations. In: *Eighth International Conference on Wireless on-Demand Network Systems and Services (WONS)*, pp.117–123 (2011)
5. Source: http://bbcr.uwaterloo.ca/~htcheng/Current_Projects.html
6. Wang, C.X., Cheng, X., David, I.L.: Vehicle-to-vehicle channel modeling and measurements: recent advances and future challenges. *IEEE Commun. Mag.* **47**(11), pp. 96–103 (2009)
7. IEEE P802.11p/D2.01.: Standard for Wireless Local Area Networks Providing Wireless Communications while in Vehicular Environment. (2007)
8. Al-Khalil, A., Al-Sherbaz, A., Turner, S.: Enhancing the physical layer in V2V communication using OFDM-MIMO Techniques. In: *PGNet, Liverpool* (2013)
9. Miao, L., Djouani, K., Wyk, B., Hamam, Y.: Evaluation and enhancement of IEEE 802.11p standard: a survey. *mobile computing*, vol. 1, no. 1 (2012)
10. Han, C., Dianati, M., Tafazolli, R., Kernchen, R.: Throughput analysis of the IEEE 802.11p enhanced distributed channel access function in vehicular environment. In: *Vehicular Technology Conference Fall (VTC 2010-Fall)*, pp 1–5 (2010)
11. IEEE: IEEE Draft P802.11-REVmb™/D12. Institute of Electrical and Electronics Engineers, New York (2011)
12. Paetzold, M.: *Mobile Fading Channels*. Wiley (2002)
13. Molisch, A.: *Wireless Communications*. Wiley-IEEE Press (2005)
14. Parsons, J.D.: *The Mobile Radio Propagation Channel*. 2nd ed., Wiley, West Sussex (2000)
15. David, W.M.: Channel Modeling for Vehicle-To-Vehicle Communications. *IEEE Commun. Mag.* **46**(5), 76–83 (2008)
16. Bello, P.: Characterization of randomly time-variant linear channels. *IEEE Trans. Commun. Syst.* **11**(4), 360–393 (1963)
17. Acosta-Marum, G., Ingram, M.A.: Six time- and frequency selective empirical channel models for vehicular wireless LANs. In: *IEEE 66th Vehicular Technology Conference, 2007, Fall, 2007*, pp. 2134–2138 (2007)
18. Meireles, R., Boban, M., Steenkiste, P., Tonguz, O.K., Barros, J.: Experimental study on the impact of vehicular obstructions in VANETs. In: *Proceeding of IEEE VNC*, pp. 338–345. Jersey City, NJ, USA (2010)

19. Karedal, J., et al.: Radio channel measurements at street intersections for vehicle-to-vehicle safety applications. In: Proceeding of IEEE VTC-Spring, pp. 1–5 (2010)
20. Tan, I., Tang, W., Laberteaux, K., Bahai, A.: Measurement and analysis of wireless channel impairments in DSRC vehicular communications. In: Proceeding of IEEE ICC, pp. 4882–4888 (2008)
21. Matz, G.: On non-WSSUS wireless fading channels. *IEEE Trans. Wirel. Commun.* **4**, 2465–2478 (2005)
22. Sen, I., Matolak, D.: Vehicle channel models for the 5-GHz band. *IEEE Trans. Intell. Transp. Syst.* **9**, 235–245 (2008)
23. Matolak, D.W.: Channel modeling for vehicle-to-vehicle communications. *IEEE Commun. Mag.* **46**, 76–83 (2008)
24. Maurer, J., et al.: A ray-optical channel model for mobile to mobile communications. In: 4th MCM COST 2100, Wroclaw, Poland, 6–8 Feb. (2008)
25. Sur, S.N., Bera, S., Bera, R., Maji, B.: MIMO channel capacity in non-uniform phase distributed Nakagami channel with ZF receiver. In: 3rd International Conference on Foundations and Frontiers in Computer, Communication and Electrical Engineering (C2E2–2016), pp 289–293 (2016)
26. Cheng, X., Wang, C.-X., Laurenson, D.I.: A Geometry-based stochastic model for wideband MIMO mobile-to-mobile channels. In: Proceeding of IEEE GLOBECOM '09, pp. 1–6. Hawaii, USA (2009)
27. Karedal, J., Tufvesson, F., Czink, N., Paier, A., Dumard, C., Zemen, T., Mecklenbrauker, C. F., Molisch, A.F.: A geometry based stochastic MIMO model for vehicle-to-vehicle communications. *IEEE Trans. Wirel. Commun.* **8**(7), 3646–3657 (2009)
28. Molisch, A.F., Kuchar, A., Laurila, J., Hugel, K., Schmalenberger, R.: Geometry-based directional model for mobile radio channels—principles and implementation. *European Trans. Telecommun.* **14**(4), 351–359 (2003)
29. Molisch, A.F.: A generic model for MIMO wireless propagation channels in macro- and microcells. *IEEE Trans. Signal Process.* **52**(1), 61–71 (2004)
30. Viriyasitavat, W., Boban, M., Hsin-Mu, T., Athanasios, V.V.: Vehicular communications: survey and challenges of channel and propagation Models. *IEEE Veh. Technol. Mag.* **10**(2), 55–66 (2015)
31. Boban, M., Barros, J., Tonguz, O.: Geometry-based vehicle-to-vehicle channel modeling for large-scale simulation. *IEEE Trans. Veh. Technol.* **63**(9), 4146–4164 (2014)
32. Al-Khalil, A.B., Turner, S., Al-Sherbaz, A.: Utilising SCM–MIMO channel model based on V-BLAST channel coding in V2V communication. In: Lecture Notes in Computer Science, vol. 9066, pp 3–11. Springer (2015)
33. Ahmed, A., Ahmad, A.E., El-Keyi, A., ElBatt, T., Bai, F., Saraydar, C.: MIMO vehicular networks: research challenges and opportunities. *J Commun.* **7**(7), 500–513 (2012)
34. Foschini, G.J.: Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.* **1**(2), 41–59 (1996)
35. Wolniansky, P.W., Foschini, G.J., Golden, G.D., Valenzuela, R.A.: V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. In: Proceedings of URSI International Symposium on Signals, Systems, and Electronics (ISSSE'98), pp. 295–300. Pisa, Italy (1998)
36. Yang, S., Hanzo, L.: Fifty years of MIMO detection: the road to large-scale MIMOs. *IEEE Commun. Surv. Tutorials* **17**(4), 1941–1988 (2015)
37. Sur, S.N., Bera, S., Singh A.K., Bera, R., Maji, B.: Performance analysis of VBLAST Based MIMO OFDM system in vehicular channel. *Int. J. Intell. Syst. Appl (IJISA)*, **6**(12), 62–68, (2014). ISSN: 2074-904X (Print), ISSN: 2074-9058 (Online)
38. Honig, M.L.: *Advances in Multiuser Detection*. Wiley, Inc., Hoboken (2009)
39. Jiang, Y., Varanasi, M.K., Li, J.: Performance analysis of ZF and MMSE equalizers for MIMO systems: an in-depth study of the high SNR regime. *IEEE Trans. Inf. Theor.* **57**(4), 2008–2026 (2011)

40. Sur, S.N., Ghosh, D., Bhaskar, D. Bera, R.: contemporary MMSE and ZF receiver for V BLAST MIMO system in Nakagami-m fading channel. In: IEEE INDICON 2011, pp. 16–18. Hyderabad, India (2011)
41. Sur, S.N., Bera, S., Bera, R. Maji, B.: Capacity analysis of ZF receiver in correlated MIMO Nakagami-m channel. In: Intelligent Computing and Applications, Advances in Intelligent Systems and Computing, vol. 343, pp 435–443 (2015)
42. Zhu, X., Murch, R.D.: MIMO-DFE based BLAST over frequency selective channels. In: IEEE GLOBECOM 2001, pp. 499–503 (2001)
43. Lee, J., Toumpakaris, D., Jang, E.W., Lou, H.: DFE-based receiver implementation for MIMO systems employing hybrid ARQ. In: IEEE GLOBECOM 2008, pp. 1–5 (2008)
44. Sur, S.N., Bera, R., Maji, B.: Decision feedback equalization for large scale MIMO System. In: International Conference On Foundations And Frontiers In Computer, Communication And Electrical Engineering (C2E2–2016), pp 295–298 (2016)
45. Sur, S.N., Bera, R., Maji, B.: Decision feedback equalization for MIMO system. In: Intelligent Computing and Applications, Advances in Intelligent Systems and Computing, vol. 343, pp. 205–212 (2015)
46. Kofidis, E., Athanasios, A.R.: Adaptive BLAST decision-feedback equalizer for MIMO-FBMC/OQAM systems. In: 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 841–846 (2010)
47. Golden, G.D., Foschini, G.J., Valenzuela, R.A., Wolniansky, P.W.: Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture. *Electron. Lett.* **35**(1), 14–16 (1999)
48. Kohno, R. Hatori, M.: Cancellation techniques of co-channel interference in asynchronous spread spectrum multiple access systems. *Electron. Commun. Jpn. (Part I: Communications)*, **66**(5), 20–29 (1983)
49. Wang, Z.: A hybrid MIMO detector combining SIC and PIC based on wideband MIMO channels. In: IET International Conference on Wireless, Mobile and Multimedia Networks, pp. 1–4 (2006)
50. Duel-Hallen, A., Holtzman, J.M., Zvonar, Z.: Multiuser detection for CDMA systems. *IEEE Pers. Commun. Mag.* **2**(2), 46–58 (1995)
51. Bai, L., Choi, J.: Low complexity MIMO detection. Springer (2012)
52. Rahman, M., Rohani, E., Jingwei, X., Choi, G.S.: An improved soft decision based MIMO detection using lattice reduction. *Int. J. Comput. Commun. Eng.* **3**(4), 264–268 (2014)
53. Ahn, J.K., Yu, S.J., Lee, E.Y. Song, H.K.: An improved lattice reduction aided detection scheme for MIMO-OFDM System. In: World Academy of Science, Engineering and Technology, vol. 5, no. 11, pp. 856–859 (2011)
54. Gan, Y.H. Mow, H.W.: Complex lattice reduction algorithm for low complexity MIMO detection. In: Proceeding of IEEE PIMRC, vol. 2, pp. 1516–1521 (2004)
55. Ubben, D.W., Seethaler, D., Jalden, J., Matz, G.: Lattice reduction. *IEEE Signal Process. Mag.* **28**(3), 70–91 (2011)
56. Yang, S., Lv, T., Hanzo, L.: Semidefinite programming relaxation based virtually antipodal detection for MIMO systems using gray-coded high-order QAM. *IEEE Trans. Veh. Technol.* **62**(4), 1667–1677 (2013)

Energy-Efficient Packet Routing Model for Wireless Sensor Network

Madhu Patil and Chirag Sharma

Abstract Wireless sensor network (WSN) in recent time has seen enormous growth. The sensor has been used in various industries and organization for various application services such as disaster management, surveillance, monitoring of patients, and so on. The WSN are tiny and battery-powered and is deployed in non-rechargeable location and preserving the battery is a most crucial factor in such network. Many technique have been proposed in recent times to enhance the lifespan of sensor network among them cluster is the most sort-out technique. Many clustering have been proposed in recent times to improve the energy efficiency of sensor network. The drawback of existing technique is that they did not consider the packet failure likelihood among inter and intra; as a result there exist scheduling bottleneck and degrades the energy of sensor devices. To overcome, this work present a packet failure estimation model by adopting a cross-layer design in order to improve the scheduling. Experiments are conducted for lifetime efficiency for varied sensor devices for proposed and existing LEACH. The result shows that the proposed model performs better than existing LEACH in term of network lifetime and energy efficiency.

Keywords WSN · Clustering · Cross layer · MAC

1 Introduction

WSN is a network composed of small and tiny sensor devices which are powered by batteries which are self-organized, deployed remotely and form a communication network. The sensor devices are getting inexpensive every passing day which has resulted in increased uses of sensor device by various industries and organization such for various application purposes such as in catastrophic managements, military strategic operation, and tracking of animal and human movement in forest

M. Patil (✉) · C. Sharma
Nitte Meenakshi Institute of Technology, Bangalore, India
e-mail: p25.madhu@gmail.com

for protection of wildlife and so on. The work of sensor devices in all these application is to collect the sensory information from surrounding and transmit this information to the nearby sink or the base stations. The issues of data aggregation and routing in sensor network are been widely studied and analyzed by various researchers [1–7].

The sensor device is battery-powered and the deployed nature of these devices poses a difficulty in recharging since these devices are deployed in an unattended environment. Therefore preserving the batter/energy is a key factor in prolonging the lifetime of sensor networks. There has been various model has been presented [7–9] to improve and prolong the lifetime of sensor network. But still the sensor network poses the following issues and challenges such as the data rate for communication, storage capacity, computation complexity, and range of communication [10–12]. The transmission capacity of sensor devices is limited due to the communication range of these devices. Due to limited range of communication the data is transmitted multiple times through intermediate/hop nodes using certain routing path selected which results in flooding of packets which can cause collision of packets, data redundancy, and energy dissipation. To overcome these short coming, clustering technique is being adopted where the WSN is partitioned into several clusters and each portioned cluster contains a cluster head and set of member devices associated to one cluster head at a time. The cluster head collects data from its member and transmit it to base station/sink or nearby cluster head. This reduces data redundancy and improves the energy efficiency of sensor network.

Various research works are going on in the field of WSN for enhancing the network lifetime in which clustering approach gain popularity. Low Energy Adaptive Clustering Hierarchy (LEACH) is the base for all clustering approach and it is the revolutionary approach [13, 14]. In this approach sensor network divided into number of cluster and each cluster has its own cluster head (CH), member of cluster communicate with CH. This approach is a remark for energy saving in the WSN. Further a centralized approach of LEACH is presented, in this approach central node, i.e., base station take all the decision regarding cluster formation, CH selection based on network and sensor nodes details. This approach is beneficial for the bandwidth utilization and energy preservation [13]. In [15] some other approach is presented for enhanced version of LEACH-C, in this approach a vice cluster head is selected by the CH to save their energy. In LEACH-CE [16] residual energy is considered for all the candidate cluster head. In [17] K-mean clustering approach is considered for WSN.

The problem with K-mean clustering is to find initial centroid vector which resulting in inappropriate size of clusters division resulting in loss of connectivity.

However, the benefit of these clustering protocols induces an overhead for formation of cluster and selection of cluster head. These issues can be overcome by utilizing a proper medium access control (MAC) design which can improve the reliability sleep scheduling of sensor devices. The MAC can be broadly classified into two types contention based and contention free based protocol. The MAC layer that adopts contention based protocol such as CSMA/CA is not efficient due to long

contention of nodes for channel access [18]. The CSMA/CA protocol induces high collision due to concurrent channel access and the packet drop increase when the node density and network is increased as a result the contention based protocol is not suitable for large networks. To overcome this contention free based protocol such as TDMA is been adopted by various clustering approach [19, 20], where each cluster member is assigned a time slot for transmission and based on this time slot scheduling is done. The adoption of TDMA improves the sleep time of sensor nodes thus the reliability and energy efficiency of network is improved.

The existing clustering methodology [19, 20] that adopted TDMA did not consider likelihood of packet loss in network channel as a result there induces a scheduling overhead among intra- and inter-cluster communication which degrades the energy consumption and reliability of network. This problem can be overcome by developing a packet failure estimation bounding model. To overcome issues presented in this paper, here our approach is to design an energy efficient network. For this a cluster head selection technique is present which is based on connectivity of nodes and packet failure using TDMA channel. This paper is organized as follows: Sect. 2 contains the proposed approach. In Sect. 3, we presented the simulation study of our work and overall work is concluded in Sect. 4.

2 Proposed Packet Optimization Based Hop Selection Model

An energy-efficient WSN formation approach is presented here based on packet failure and network connectivity. Our propose cluster formation approach is similar to the LEACH approach [13, 14]. Hence first we discuss about the LEACH clustering approach. A cluster consists of several devices. In each round a new cluster head is elected based on some threshold value $H(d)$, in this approach let the random device be called as d , obtain a uniform value between 1 and 0. This obtained value is compared with $H(d)$. If device obtained value is less than $H(d)$ then it elected itself as cluster head for that particular round. Threshold value is updated after each round.

$$H(d) = \begin{cases} \frac{r}{1-r \times \lfloor \varphi \bmod (1/r) \rfloor} & \text{if } d \in \bar{S}; \\ 0, & \text{Otherwise.} \end{cases} \quad (1)$$

In the above Eq. 1, r represents the cluster head ratio with respect to the total number of devices in each round. The Current round number is represented by φ , $0 \leq \varphi < \infty$ and collection of devices represented by \bar{S} which is not elected as cluster head of period $1/r$ rounds. Based on above Eq. 2 every sensor devices behave as CH for a particular period in a round. Already elected cluster head sensor node removed from the next round of candidate selection procedure.

In the proposed model, the nodes are randomly deployed over the network with density δ and each nodes has same radius S which has the overlapping section and

based on above Eq. 1, here authors design their own cluster formation approach for a particular device d in a round. Value of threshold and parameter r changed per normalized value of overlapped region. Hence for a device d we get the following values:

$$r(d) = \alpha \times \omega(d), \quad (2)$$

α represents the average amount of CH. Based on this threshold value $H(d)$ represented for sensor device d as

$$H(d) = \begin{cases} \frac{r}{1-r(d) \times [\phi \bmod (1/r(d))]} & \text{if } d \in \bar{S}; \\ 0, & \text{Otherwise.} \end{cases} \quad (3)$$

Number of sensor device which is not elected a cluster head is represented as \bar{S} for that particular round. In that particular round $1/r(d)$, d act as a CH. Based on this approach there is a chance for electing each device a cluster head in the network.

The likelihood of cluster head neighbor can be obtained by following equation:

$$\mathcal{F}_d(d) = (\delta \pi S^2)^d e^{-\gamma \pi S^2} / \delta!, \quad \delta = 0, 1, 2, 3, \dots, \infty. \quad (4)$$

The transmission in cluster network is classified into following as inter-clustering and intra-clustering transmission.

To evaluate the likelihood of packet failure in intra-cluster communication, let us consider that all the member devices of a cluster will not decode the data correctly. We consider the Rayleigh fading channel based on which the received Signal-to-Noise Ratio (SNR) γ_1 , which is at distance s from cluster head is obtained as follows

$$\gamma_1 = \mathcal{G} D_1 |c|^2 / s^2 \mathcal{N}_0, \quad (5)$$

where $\mathcal{G} = \frac{(\mathcal{G}_{\text{trns}} \mathcal{G}_{\text{rcvr}} \lambda^2)}{(L_m F_n (4\pi)^2)}$, where λ represent the carrier wavelength, F_n represent noise param of receiver, $\mathcal{G}_{\text{trns}}$ and $\mathcal{G}_{\text{rcvr}}$ represent the antenna gain of transmitter and receiver respectively and L_m is the additive background noise of hardware devices, \mathcal{N}_0 is the receiver end Gaussian noise power, $|c|$ represent Rayleigh distribution, D_1 is the intra-cluster communication energy dissipation per bit and c is the wireless channel gain among sensor node of a cluster and its cluster head. Therefore for a particular D_1 and s , the mean bit error rate (BER) is obtained as follows:

$$L_1^b = D(\mathbb{Q}(2\gamma_1)) \cong s^2 \mathcal{N}_0 / 4\mathcal{G}D_1, \quad (6)$$

where $D(\cdot)$ average param related to distribution of $|c|$, $\mathbb{Q}(y)$ is Q-function and is represented as $\mathbb{Q}(y) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_y^\infty e^{-x^2/2} dx$. we consider the packet length to be B bit based on first order energy model. Therefore the likelihood of packet failure L_1^p is defined as follows:

$$L_1^p = 1 - (1 - L_1^b)^B. \quad (7)$$

The distance s among cluster head and its adjacent neighbor within each cluster, the likelihood density function is obtained as follows:

$$\mathcal{F}_s(s) = \frac{2s}{S_1^2}, S_1 \geq s > 0 \quad (8)$$

The L_1^b is trivial, therefore we obtain $1 - (1 - L_1^b)^B \cong BL_1^b$. Therefore using Eq. (7) and (8) we obtain the mean likelihood of packet failure in a cluster is as follows:

$$\bar{L}_1^p \cong \int_0^{S_1} B(s^2 \mathcal{N}_0 / 4\mathcal{G}D_1) 2s / S_1^2 ds = B\mathcal{N}_0 S_1^2 / 8\mathcal{G}D_1. \quad (9)$$

The hop nodes channel is consider being independent. Therefore likelihood of a devices that successfully decode packet is as follows:

$$L_a(a) = \sum_{d=a}^{\infty} \binom{d}{a} (\delta\pi S_1^2)^d e^{-\delta\pi S_1^2} / d! (\bar{L}_1^p)^{d-a} (1 - \bar{L}_1^p)^a, \quad (10)$$

where $a = 0, 1, 2, 3, \dots$. Therefore the mean number of devices in the cluster that decode packet correctly is given as follows:

$$\bar{a} = \sum_{a=0}^{\infty} a * L_a(a) = \delta\pi S_1^2 (1 - \bar{L}_1^p). \quad (11)$$

Similarly to obtain the likelihood of packet failure in inter clustering communication is obtained based on Eq. (7). Which is follows:

$$L_2^p = 1 - (1 - L_2^b)^B \cong BL_2^b. \quad (12)$$

In inter-cluster communications, cluster head transmit the message to its members with mean energy depletion D_1 /bit, based on Eq. (4) number of member device $\delta\pi S^2$ with in a cluster that is trying to obtain packet is obtained. Therefore

the mean energy consumption of intra cluster transmission I'_D per packet is formulated as follows:

$$I'_D = B((1 + \beta)D_1 + D_{\text{trns}} + \delta\pi S^2 D_{\text{rcvr}}), \quad (13)$$

where B represents the packet length bit size, D_{trns} represent the energy consumption per bit of transmitter, D_{rcvr} represent the energy consumption per bit of receiver circuit and β represent the power amplifier transmission efficiency. For simplicity, here we consider that all nodes in the network has same D_{trns} and D_{rcvr} . To select the hop devices in inter-cluster communication the mean of $\bar{a} + 1$ devices including cluster head energy dissipation I''_D per packet is formulated as follows:

$$I''_D = B((1 + \beta)(\bar{a} + 1)D_2 + (\bar{a} + 1)D_{\text{trns}} + D_{\text{rcvr}}) \quad (14)$$

The overall energy consumption of proposed cluster head/hop selection model is obtained as follows:

$$I_D = I'_D + I''_D \quad (15)$$

$$I_D = D((1 + \beta)(D_1 + \delta\pi S_1^2(1 - \bar{L}_1^p)D_2)) + B((\delta\pi S_1^2(1 - \bar{L}_1^p) + 2)D_{\text{trns}} + (\delta\pi S_1^2 + 1)D_{\text{rcvr}}). \quad (16)$$

The Eq. (16) shows the total energy dissipation of a wireless sensor network. Therefore the total energy dissipation depend on following parameters B , δ , S_1 , D_1 , D_2 and \bar{L}_1^p depends on D_1 . The simulation and experimental study of proposed packet failure optimization based energy model is evaluated and compared with existing methodology in next section below.

3 Simulation Result and Analysis

Window 10 enterprise edition with 64-bit Quad core processor, 2 GB NVIDIA CUDA Dedicated Graphic card, with 16 GB of RAM is used for simulation study. For sensor network deployment we used sensoria simulator. It used C# programming language. We considered various network parameters for simulation study which are represented in the Table 1. A large network size is considered for our system analysis and deployed 1000 and 2000 sensor node. We analyzed our system performance and compared our network lifetime and energy efficiency with LEACH-based protocol.

The network lifetime is obtained for 40% of sensor node death. In Figs. 1 and 2, we can see that the proposed approach network lifetime efficiency is better than existing LEACH approach. The proposed approach enhanced the lifetime of sensor

Table 1 Network parameter considered for simulation

Network parameter	Value
Network size	40 m * 40 m
No. of sensor nodes	1000, 2000
No. of base station	1
Sensor device initial energy	0.2 J
Radio energy dissipation	50 nj/bit
Data packets length	2000 bits
Transmission speed	100 bit/s
Bandwidth	5000 bit/s
Idle energy consumption (Eelec)	50 nj/bit
Delay in data processing	0.1 ms
Amplification energy (Emp)	100 pJ/bit/m ²

Fig. 1 Network lifetime analysis for 1000 nodes

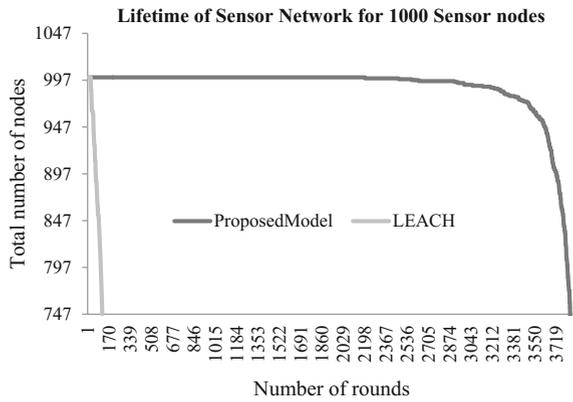
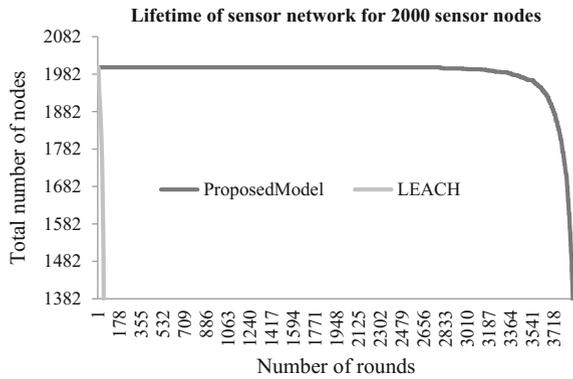


Fig. 2 Network lifetime analysis for 2000 nodes



network by over 96.44 and 98.64% when size of network is considered as 1000 and 2000 nodes respectively over LEACH.

As from above Fig. 3, we observed that as number of sensor device increased performance of the proposed model get better while LEACH performance degraded when number of nodes increased. The result shows that LEACH protocol is not suitable for large network and the proposed model is adaptive in nature with increase in node density.

The energy remaining is obtained for 40% of sensor node death. In Figs. 4 and 5, we can see that the proposed model performs better than the existing LEACH algorithm in term of energy remaining efficiency. The experimental result shows that the energy efficiency of the proposed model over the existing LEACH.

When we considered, the sensor node equal to 1000 improved energy percentage over the LEACH is 76.16% and when size of network is 2000 nodes improvement percentage is 79.12%. With the increased number of sensor node performance of

Fig. 3 Network lifetime analysis

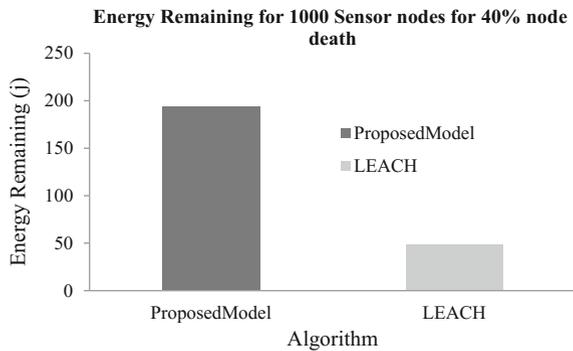


Fig. 4 Energy remaining for 1000 sensor nodes for 40% sensor node death

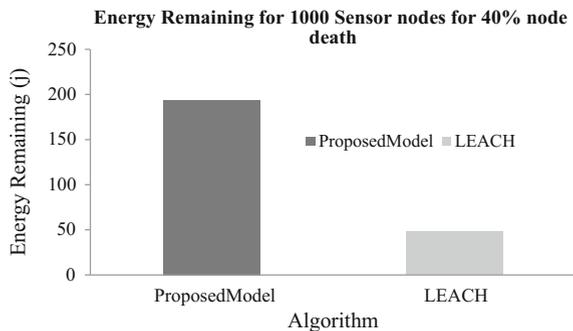
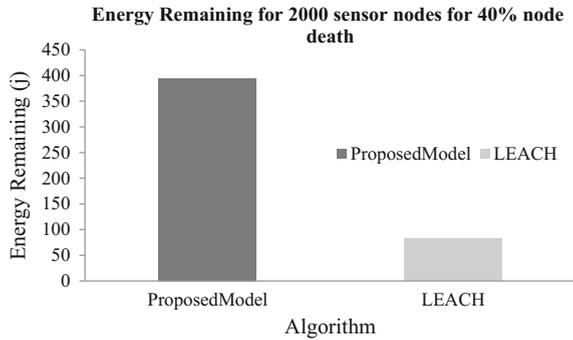


Fig. 5 Energy remaining for 2000 sensor nodes for 40% sensor node death



the proposed model get improved but LEACH model performance decreases with the increase in the network size.

4 Conclusion

Clustering technique play a major role in improving network lifetime efficiency of sensor network. In recent time, when size of network is large existing LEACH-based clustering approach is not sufficient for network lifetime since these protocols do not consider packet failure likelihood among inter- and intra-cluster transmission. To address the energy-efficiency issue of existing approach, this work proposes packet failure likelihood estimation model by adopting cross-layer design for hop selection. Our proposed model improves the lifetime of network by over 96.44 and 98.64% over LEACH when sensor node is equal to 1000 and 2000 respectively; and the energy efficiency is improved by 76.16 and 79.12% over LEACH when sensor node equal to 1000 and 2000 respectively. Experimental results show that the proposed model better than LEACH in term lifetime efficiency and energy efficiency. The outcome of proposed model achieved show that it achieves significance performance when compared to E^2E^2 (energy efficiency and reliable routing) proposed by H.K. Deva Sarma et al. Further in the future, we will extend our system and conduct the simulation performance for other network parameter such node decay and active rate. We will also develop a sleep/active scheduling for inter- and intra-cluster communication.

References

1. Dow, C.R., Lin, P.J., Chen, S.C., Lin, J.H., Hwang, S.F.: A study of recent research trends and experimental guidelines in mobile ad-hoc network in IEEE AINA, vol. 1 (2005)
2. Fang, X., Gao, H., Li, J., Li, Y.: Approximate multiple count in wireless sensor networks. In: IEEE INFOCOM (2014)

3. Guo, S., Wang, C., Yang, Y.: Mobile data gathering with wireless energy replenishment in rechargeable sensor networks. In: IEEE INFOCOM (2013)
4. Liao, Q., Shi, L., He, Y., Li, R., Su, Z., Striegel, A., Liu, Y...: Visualizing anomalies in sensor networks. In: ACM SIGCOMM, vol. 41(4) (2011)
5. Wang, J., Cao, Z., Mao, X., Liu, Y.: Sleep in the dins: insomnia therapy for duty-cycled sensor networks. In: IEEE INFOCOM (2014)
6. Akyildiz, I.F., Su, W., Sankara subramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* **40**(8), 102–114 (2002)
7. Schurgers, C., Srivastava, M.B.: Energy efficient routing in wireless sensor networks. In: IEEE MILCOM, vol. 1 (2001)
8. Ma, J., Lou, W., Wu, Y., Li, M., Chen, G.: Energy efficient TDMA sleep scheduling in wireless sensor networks. In: IEEE INFOCOM (2009)
9. Bandyopadhyay, S., Coyle, E.J.: An energy efficient hierarchical clustering algorithm for wireless sensor networks. In: IEEE INFOCOM, vol. 3 (2003)
10. Cheng, J., Ye, Q., Jiang, H., Wang, D., Wang, C.: STCDG: an efficient data gathering algorithm based on matrix completion for wireless sensor networks. *IEEE Trans. Wirel. Commun.* **12**(2), 850–861 (2013)
11. Li, Y., Sheng, M., Wang, C., Wang, X., Shi, Y., Li, J.: Throughput-delay tradeoff in interference-free wireless networks with guaranteed energy efficiency. *IEEE Trans. Wirel. Commun.* **14**(3), 1608–1621 (2015)
12. Lee, H.J., Lim, J.T.: Cross-layer congestion control for power efficiency over wireless multihop networks. *IEEE Trans. Veh. Technol.* **58**(9), 5274–5278 (2009)
13. Heinzelman, W.B., et al.: An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **1**(4), 660–670 (2002)
14. Liu, X.: A survey on clustering routing protocols in wireless sensor networks. *Sensors* **12**(12), 11113–11153 (2012)
15. Zhao, F., Xu, Y., Li, R.: Improved leach routing communication protocol for a wireless sensor network. *Int. J. Distr. Sensor Netw.* **2012** (649609), 6 (2012)
16. Tripathi, M., et al.: Energy efficient clustered routing for wireless sensor network. In *Proceeding of IEEE MSN'13*, pp. 330–335, 11–13 (2013)
17. Mehta, D., Harous, S., Alem, I., Khebbab, D.: LEACH-CKM: Low energy adaptive clustering hierarchy protocol with K-means and MTE. In *IIT'14, Al Ain*, pp. 99–103 (2014)
18. Ye, M., Li, C., Chen, G., Wu, L.: EECS: An energy efficient clustering scheme in wireless sensor networks. In *Proceedings of the Performance, Computing, and Communications Conference, Phoenix, AZ, USA*, pp. 535–540 (2005)
19. Mammu, A.S.K., Hernandez-Jayo, U., Sainz, N., de la Iglesi, I.: Cross-layer cluster-based energy-efficient protocol for wireless sensor networks. *Sensors* **15**, 8314–8336 (2015)
20. Deva Sarma, H.K., Mall, R., Kar, A.: E2R2: Energy-efficient and reliable routing for mobile wireless sensor networks. In *IEEE Syst. J.* **10**(2), 604–616 (2016)

An ECC-Based Algorithm to Handle Secure Communication Between Heterogeneous IoT Devices

S. Sasirekha, S. Swamynathan and S. Suganya

Abstract Internet of Things (IoT) is an interconnection of tiny real-world objects using sensor, actuators, and software's designed for exchanging data and controlling them. It invades business landscape on massive scale integrating heterogeneous devices. Some of the application includes media, environmental monitoring, infrastructure and energy management, medical, healthcare system, and transportation. Earlier, in these applications, the primary issue is handling efficient communication among the interconnected devices to have an extended network lifetime. Nowadays, as data grows rapidly in the IoT, security and privacy is also a major issue to be concentrated in these applications to reduce vulnerabilities. Recent studies show that the existing IP security protocols and algorithm have technical limitation in applying it to the context Internet of Things. In particular, when a new device enters to an existing secured network, the network should ensure that the new device is authenticated and authorized properly with respect to the properties of existing network. Otherwise, similar to other network, IoT network also will be subjected to various attacks such as physical attacks, service disruption attacks, data attacks, and denial-of-service attacks. Hence, in this work, an algorithm to authenticate and authorize when a new device gets added to the existing network is proposed. The algorithm is tested for various attacks and the results are proven to be efficient and secure.

Keywords Internet of Things (IoT) · PAN network · IoT security
ECC algorithm

S. Sasirekha (✉) · S. Suganya
SSN College of Engineering, Chennai 603310, India
e-mail: sasirekhas@ssn.edu.in

S. Suganya
e-mail: suganyasrithar@gmail.com

S. Swamynathan
Anna University, Chennai 600025, India
e-mail: swamyns@annauniv.edu

1 Introduction

The Internet of Things (IoT) is a contemporary domain which has the ability to connect all real-world object/things around us and assign a unique ID and get connected to Internet. This ID helps the user to fetch the data from the objects easily without involving human-to-thing interaction similar to a web data access. The term IoT implicitly states that the standard Internet protocol is used communicating between things. The concept may also be entrusted as the Internet of Everything. The popularity of this domain is because this IoT system fits in easily in the Internet infrastructure. The main goal of this domain is to empower things to connect things around us eventually irrespective of the time, location, and person using any network and service [1]. The applications of IoT are essentially endless. It consists of environmental monitoring, framework management, energy conservation, home and office automation, transportation, medical and healthcare systems, and a particularly aggressive application of smart cities. Smart cities would enclosure all the devices within the city into a single network with global connectivity. This innovative concept can now be applied in automating everyday tasks in the office, at homes, and practically anywhere to obtain better results with minimal labor. However, IoT is advantageous. In recent days, it has few research challenges that need to be addressed. Some of them include complexity, compatibility, and security/privacy. Initially, in IoT developing stages, security was not a major concern, but with the rapid growth of data security has become an issue of high concern. IoT security as an evolving paradigm which needs much concentration on security, as it handles enormous data eventually [2]. It is a major issue, because, as one security incident may harm the good infinite edges of IoT. The option that differentiates IoT security problems [3] from the standard ones is the heterogeneous and huge scale objects and networks. The huge challenge for security is device authentication [4], mainly deals with automatic systems. The gateways that bind IoT devices to company and manufacturer networks need to be secured as well as the devices themselves [5]. IoT data is being deposited in vast repositories, which can become fetching targets for corporate hackers and industrial spies who depend on big data to make profits.

IoT devices have been bringing forward the IoT security to promote the stack of issues that require to be addressed quickly. IoT devices are always attached and always on. In this proposed work, when a new device adds or joins to an existing PAN network, then it authenticates and authorizes the new device using the proposed algorithm based on Elliptic Curve Cryptography. Considering the resource constraint nodes and sensor devices, REST with JSON data format is adopted to handle the communication between them.

The remaining of the work is organized as follows: In Sect. 2 some of the related works are discussed. In Sect. 3 the system model for the proposed system is introduced. Section 4 shows about the authentication algorithm for the proposed network. Section 5 confers the implementation work carried out for the experimental setup. In Sect. 6, analysis of the proposed work is evaluated. In Sect. 7, conclusion of the proposed work is examined.

2 Related Work

There are many research works carried out related to IoT security. Some of them are discussed as follows:

Dong Hee Kim addresses IoT as primary-based product and service area which is widely utilized in numerous business space and our daily lives. There are several burdens regarding the security issues in using IoT-based application; however, the questions of safety are not verified [6]. R. Roman refers that the safety design ought to be applicable to the lifecycle of an issue and its capabilities. It implies that a security space is essential for a trusted-outsider amid this strategy, or when a kind of conventions actualized. Security conventions should manufacture lightweight security components that are unit gettable to be kept running on little things. Furthermore to permit end-to-end security and area particular convention variations, conventions should get to be to bolster interpretations done by portals [7]. The bunch security ought to be considered and IoT brings correspondence designs that range unit remarkable in old systems, and are not adequately bolstered by end-to-end web security conventions. Protocol style ought to any can take the difficulty of packet fragmentation on security, with explicit target doable DoS attacks.

Pawlowski defines regarding the authentication protocol employed in Wireless Local Area Networks (WLANs) is EAP (Extensible Authentication Protocol). This authentication mechanism dwells into three services (Authentication server, critic, and supplicant), and two protocols that are unit carrying EAP frames (RADIUS and EAPOL). EAPOL may be a mechanism that encloses EAP messages and relocates them between supplicant and also the critic over the link layer [8]. By abuse EAPOL and RADIUS, authentication system comprises three stages: authentication start, authentication message trade, and authentication call. Leandro Marin gives PC code improvement for secure correspondence in heterogeneous Internet of Things systems. The ask signature subject has been reused as partner praiseworthy answer using our streamlined PC code calculations [9]. Key transaction convention is planned and it is a bolstered huge topic that exhibits the ease of use of the offered PC code enhancements. Trust Extension Protocol for Authentication generally amplified article-related sensors over the manufacturer (TEPANOM) is an enlargement of this work

N. Thangarasu offers PC code on the grounds that the most elevated quality per-key-piece of any radiant open key arrangement of beginning era procedures like RSA, Diffie-Hellman [10]. PC code offers indistinguishable level of security and microcircuit house is limited for system location misuse bunch, keen card, remote gadget, and cell phones. The proceeding with advancement of gages is to a great degree important position for the work of a cryptosystem.

Jive Park uses a unnatural device which will use DTLS communication with none of the ASCII text files for a DTLS handshaking. The system may have endless alternative of cipher suite. It uses the DTLS protocol into a handshaking section associated with a secret writing section, end-to-end security is contracted as a result

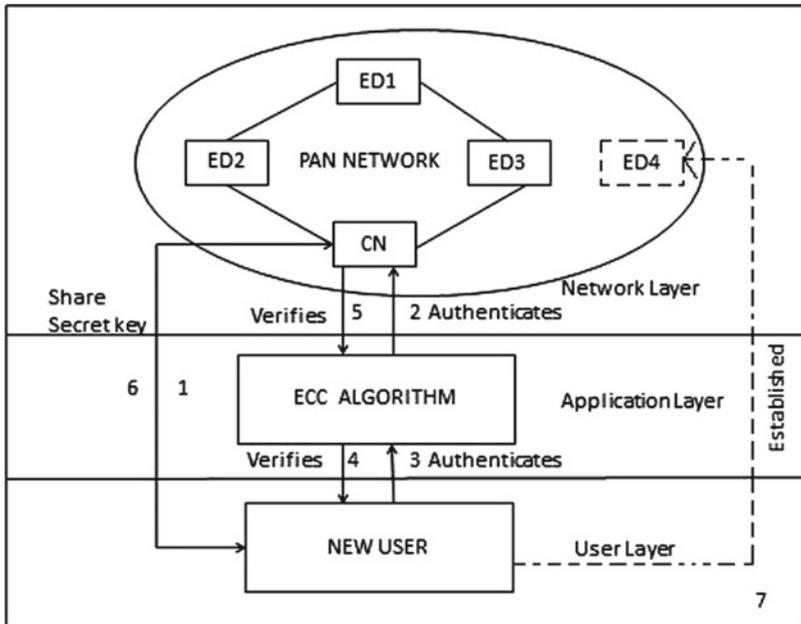
of encryption, and secret writing area unit finished within the end node [11]. Our framework will encourage sending unnatural gadgets in an extremely secure way in constrained situations. Usman Sarwar talks about viewing 6LoWPAN somebody as a versatile element device which may perform bi-directional correspondence with the gadget hubs and skilled of performing expressions various sorts of correspondence with the 6LoWPAN gadget organize [12]. Huansheng Ning arranged partner-accumulated verification based on generally evaluated confirmation subject for the U2IoT plan. Inside the APHA, two sub-conventions of zone unit intended to deliver base-up security insurance [13]. It furthermore gives learning privacy and information honesty by the coordinated way descriptor and closeness-based generally tumultuous maps, starts trust connections by means of the lightweight systems, and applies progressively hashed qualities to accomplish session freshness. Shahid Raza portrays that the DTLS might be packed misuse 6LoWPAN institutionalized instruments [14]. DTLS may be compressed and enlarged to device networks. Liu Yongliang proposes a replacement protocol referred to as glorious key security [15]. Manoj Prabhakar introduces Mobile Authentication Model (MAM) which is intended to get passcode for authentication method. As a general request, such assurance should not yield confirmation power or acquaint potential security helplessness with the hidden validation topic [16].

In like manner some ID-based Authentication (IBA) depends on ID-based marks with REST URI. IBA depends on IBE with REST URI. It utilizes the ID-based cryptography and REST unmistakable URI for validation. In IBA, customers and servers get their own keys from PKG. A purchaser needs to capture the URI of the asset server (RS) to perform confirmation. The shopper will create the RS's open key by consolidating RS' URI with the ace open key. Similarly, the RS will get the customer's open key by consolidating customer's URI with the ace open key. Every gathering will encipher a message beneath the recipient's open key and sign underneath its very own key. This straightforwardness incorporates a pleasant preferred standpoint for IoT. IoT gadgets might be honest to goodness by IBA while not keeping up a session or customer standing.

Based on the above discussion, the proposed system is designed to facilitate an option to deploy unnatural devices in a very secure manner in forced environments using ECC key exchange algorithm.

3 System Model for Authenticating Devices in IoT Network

The working model for the proposed authentication and authorization algorithm based on ECC is represented in three layers as shown in Fig. 1. The three layers are, namely,



ED-End Device
CN-Coordinator Node
PAN-Personal Area network

Fig. 1 System model for authenticating devices in IoT network

- Layer 1: Network layer
- Layer 2: Application layer
- Layer 3: User layer

3.1 Network Layer

In this layer a Personal Area Network (PAN) is established among the nodes. The nodes are interconnected among them using a mesh topology and coordinated by a centralized/gateway node. The nodes in the network are named as end node (ED) and the centralized node is named as the coordinator node (CN). The nodes are connected within a range of 10 m. In this layer the sensors and a communication

interface ZigBee are used for communication purpose in the network. A sensor is a device that identifies and responds to some type of input from the physical environment. ZigBee [17] is a technical standard for communication protocols for establishing a PAN network using low-range digital radios. The ZigBee plays three different roles such as coordinator, router, and end device. The major benefits are low power transmission, transmission with specified delay, and large network capacity. ZigBee network is more flexible so that any node can transfer data to coordinator node which is connected with web server.

3.2 Application Layer

The application layer provides the end user with a variety of shared network services such as establishing the flow of data exchange between the nodes and also providing a provision to enlarge the network based on the application requirements as shown in Fig. 2.

When one or more new devices enter into the same network, the network range expands. This concept is known as enlargement of networks. When more nodes enter into the same network, there arises a security issue which is a major one in application layer. This is because, when there is a fast arrival of new nodes, the coordinator node takes some time to identify whether the node is authorized or

Fig. 2 Flow chart of communication in application and network layers

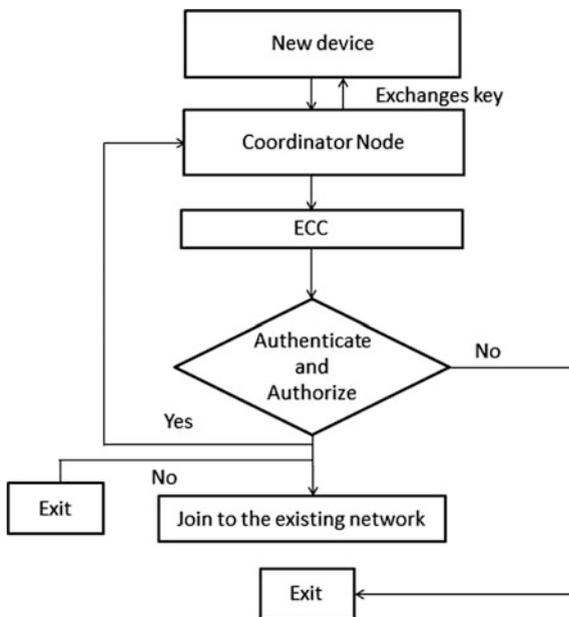


Table 1 Comparison between authentication algorithms with various parameters

Parameters →	Algorithm type	Key/value	Key length	Computation
RSA	Asymmetric type	3072 bits	Increased key size	Factorization
ECC	Asymmetric type	256 bits	Small key size	Discrete logarithm
MD5	Hashing type	128-bit hash value	Hash values used	Hashing method
SHA	Hashing type	160-bit hash value	Hash values used	Hashing method

unauthorized. Therefore, to design a security algorithm a study on exiting IP security algorithm is done to adapt those for resource constraint devices.

Table 1 shows about a small analysis of different algorithms, which is to build our considering resource constraint sensor devices. Compared to other algorithms, the key size of ECC is small, even if the message size is too big. Key size is one of the major advantages in ECC, so that it avoids the problems between coordinator node and new node in the network in draining its energy fast. ECC algorithm is briefly elaborated in Sect. 4.

3.3 User Layer

The novice device requests the coordinator node to join in that particular network. The coordinator authenticates and authorizes the new user by verifying the pass-code; if it matches then corresponding user can join in the network.

4 Authentication and Authorization Algorithm for IoT Networks

The proposed algorithm to authenticate and authorize new device is elaborated as follows:

- Step 1. Initial set up → PAN network is created at device level with heterogeneous devices (i.e., Arduino and Raspberry pi) and the communication between these heterogeneous devices have been taken place.
- Step 2. In PAN network, data transmission occurs between the devices which are manually incorporated by the anonymous user.
- Step 3. By modifying the PAN network with those heterogeneous devices, it detects the unauthorized user. For this, algorithm based on ECC is used in the application layer which communicates to other arduino devices in the user layer.

Step 4. The authorized and unauthorized user should be authenticated by the coordinator node using the following key generation method.

Step 5. **The ECC key generation steps are illustrated as follows:**

1. Assume a and b are the equation coefficients of the elliptic curve.
2. Choose a random integer, primary key k , where $k \in [1..n - 1]$.
3. Calculate the public key H in this point, $H = kG$ (where G is the base of the subgroup).
4. To establish the secure connection between new node N and the coordinator node C the public keys H_N and H_C are inter-exchanged.
5. Using the exchanged public key, the secret key S is determined as $S_C = H_N * k_C$ and $S_N = H_C * k_N$, where k_N and k_C are the private keys of new node and coordinator node, respectively.
6. Check whether the values computed by both the parties are same. If matches it is authenticated.

Step 6. REST [18, 19] with JSON data format is adopted for web services used in algorithm based on ECC.

For exchanging a passcode many data formats are used in various applications like JSON, XML, etc. In this work, JSON format is adopted for exchanging authentication and authorization messages. The code snippet for key generation in JSON format is given below:

```
var data = [{
  "Key": "AQAB", Email: "test@test.com"}];
function getByKeyValue(key) {
  var detect = null;
  for (var j = 0; j < value.length; j++) {
    var node = value[j];
    if (node.Key == key) {
      detect = element;
    }
  }
  return detect;
}
```

5 Implementation

The implementation is carried out in two stages such as creation of PAN network and authenticating and authorizing a new node joining the network.

5.1 Creation of PAN Network

The implementation is done for a simple environmental monitoring scenario which involves a temperature, humidity, and a moisture sensor. For this purpose, sensor boards, sensors, and suitable communication techniques are used. A PAN network is established using heterogeneous IoT devices which are vendor specific.

In this implementation, the heterogeneous setup is done with three different sensor node environments such as arduino UNO, arduino MEGA, and Raspberry Pi. To construct sensor coordinate with a blend of equipment stages, Digi XBee is utilized to communicate on the grounds that they are 802.15.4 standard. ZNet 2.5 firmware is preloaded in the XBee-supported modules. The firmware provides the support to execute the ZigBee convention stack. With a specific end goal to integrate all the sensor nodes, various fundamental settings are done in the system design level. Initially, for all the radios to communicate seamlessly, it is determined to set the same Personal Area Network (PAN) ID and synchronize them in the unified channel. In this system for the purpose of testing PAN ID 0x1234 and channel 0x0D are set for all the radios in the network [20]. After customization, the visualization of the interconnected nodes with coordinator nodes and end nodes are shown in Fig. 3.

As mentioned in Sect. 4, an authentication and authorization algorithm based on ECC is configured in the coordinator node. In this implementation, an Arduino MEGA board is used as the coordinator node [21].

When node A sends request to node B, it has been verified by ECC web server which acts as a coordinator in the network. If the request is authorized, then it sends to node B, the same procedure is followed for response also. ECC acts as a coordinator for processing and verifying the request and response in the network.

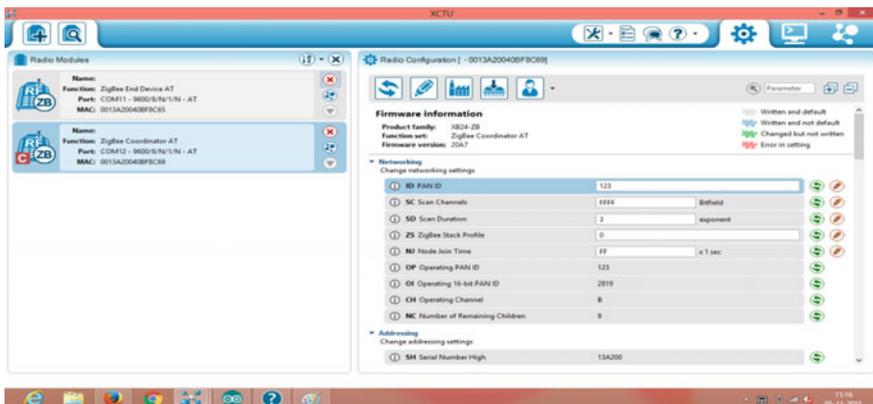


Fig. 3 Coordinator and end device getting ready for network creation

5.2 Authentication and Authorization for a New Node

After the PAN network is setup, when a new device joins the existing network the algorithmic step 4 as mentioned in Sect. 4, a key generation process starts to validate the new device. On execution of the key generation process, the key is generated in json format as shown below [22]. Further, it is transferred to the coordinator node, and the passcode is generated being exchanged with the new device and it authorizes that are entered into the network as shown below.

```
{
  "keys": [
    {
      "Private key":
      "00cc05af300fa0c8d868a2368dae87de289eadba05c0c22a643c9db200eb3410f2",
      "Public key":
      "0373f44bbdf83371cf200c388684fa98d4e6a409db27f344465e52bd0632ecf4b3",
      "Calculated public key":
      "0373f44bbdf83371cf200c388684fa98d4e6a409db27f344465e52bd0632ecf4b3"
    }
  ]
}
```

6 Evaluation

The evaluation is carried out to prove the efficiency of the proposed algorithm[23, 24]. The test was done for all the possible attacks using Soap UI NG Pro. The results are shown in Table 2, and the proposed algorithm is tolerant for DOS attack, Man in the middle. Injection input attacks. Hence, the proposed algorithm is efficient and secure for this environment.

Table 2 Evaluation against different attacks

Attacks	Proposed algorithm (REST with JSON)
XSS attack	Yes
Injection input attack	Yes
DOS attack	Yes
XML attack	No
Replay attack	No
Man in the middle attack	Yes

7 Conclusion

In communication between networks, security is one of the major issues which leads to another major problem such as attacks. In this paper, PAN is created in network layer with heterogeneous devices like arduino and raspberry pi. When a new device enters into an existing network, there occurs an issue of authentication. This can be performed by the coordinator node which is in the network layer within the networks. The role of the coordinator node is a major part in PAN. This CN has an authority to authenticate the new node that enters into the existing network. For nodes in the proposed network, an algorithm is used in the application layer for key generation. The key generated here is in the format of REST with JSON. These algorithms are based on ECC, so using this, attacks in the communication can be reduced and other factors such as lifetime and memory space can be maintained throughout the network establishment.

References

1. Wang, C., Daneshmand, M., Dohler, M., Mao, X., Hu, R.Q., Wang, H.: Guest editorial—special issue on internet of things (IoT): architecture, protocols and services. *IEEE Sens. J.* **13** (10), 3505–3510 (2013)
2. Ludena, R.D.A., Horibe, N., Ahrary, A., Yang, W.S.: IoT-Security approach analysis for the novel nutrition-based vegetable production and distribution system. In: *IIAI 3rd International Conference on Advanced Applied Informatics* (2014)
3. Zhang, Z.-K., Cho, M.C.Y., Wang, C.W., Hsu, C.W., Chen, C.K., Shieh, S.: IEEE Fellow: IoT security: ongoing challenges and research opportunities. In: *IEEE 7th International Conference on Service-Oriented Computing and Applications* (2014)
4. Falk, R., Fries, S.: Managed certificate whitelisting—a basis for internet of things security in industrial automation applications. In: *The Eighth International Conference on Emerging Security Information, Systems and Technologies* (2014) ISBN: 978-1-61208-376-6
5. Kruger, C.P., Hancke, G.P., Implementing the internet of things vision in industrial wireless sensor networks, *IEEE* (2014)
6. Kim, D.H., Cho, J.Y., Kim, S., Lim, J.: A Study of developing security requirements for internet of things (IoT). *Adv. Sci. Technol. Lett.* **87** (Art, Culture, Game, Graphics, Broadcasting and Digital Contents) (2015)
7. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in distributed internet of things. *J. Comput. Netw.* **57**(10) (2013)
8. Pawlowski, M.P., Jara, A.J., Ogorzalek, M.J.: Extending extensible authentication protocol over IEEE 802.15.4 networks. In: *8th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2014)*, pp 340–345 (2014)
9. Marin, L., Pawlowski, M.P., Jara, A.: Optimized ECC implementation for secure communication between heterogeneous IoT devices, *J. Sens.* **15** (2015)
10. Thangarasu, N.: Implementation secure authentication using elliptic curve cryptography. *Int. J. Innovative Res. Adv. Eng. (IJIRAE)* **1**(1) (2014)
11. Park, J., Kang, N.: Lightweight secure communication for CoAP-enabled internet of things using delegated DTLS handshake, 978-1-4799-6786-5/14/\$31.00 ©2014 IEEE ICTC (2014)

12. Sarwar, U., Sinniah, G.R., Suryady, Z., Khosdilniat, R.: Architecture for 6LoWPAN mobile communicator system. In: Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 vol. II, IMECS (2010)
13. Ning, H., Liu, H., Yang, L.T.: Aggregated-proof based hierarchical authentication scheme for the internet of things. *IEEE Trans. Parallel Distrib. Syst.* **26**(3) (2015)
14. Raza, S., Tralbalza, D., Voigt, T.: 6LoWPAN compressed DTLS for CoAP, In: 8th IEEE International Conference on Distributed Computing in Sensor Systems (2012)
15. Yongliang, L., Gao, W., Yao, H., Yu, X.: Elliptic curve cryptography based wireless authentication protocol. *Int. J. Netw. Secur.* **5**(3), 327–337 (2007)
16. Prabhakar, M.: Elliptic curve cryptography in securing networks by mobile authentication, *Int. J. Sci. Res Publ.* **3**(9) (2013) ISSN 2250-3153
17. Somani, N.A., Patel, Y.: Zigbee: a Low power wireless technology for industrial applications. *Int. J. Control Theor. Comput. Model. (IJCTCM)* **2**(3) (2012)
18. Serme, G., de Oliveira, A.S., Massiera, J., Roudier, Y.: Enabling Message Security for RESTful Services. In: IEEE 19th International Conference on Web Services (2012)
19. Lee, S., Jo, J.Y., Kim, Y.: Method for secure RESTful web service IEEE, ICIS 2015 (2015)
20. Alanazi, S., Al-Muhtadi, J., Derhab, A., Saleem, K., AlRomi, A.N., Alholaibah, H.S., Rodrigues, J.J.: On resilience of wireless mesh routing protocol against DoS attacks in IoT-based ambient assisted living applications. In: IEEE 17th International Conference on e-Health Networking, Applications and Services (Healthcom) (2015)
21. Martinez, V.G., Encinas, L.H.: Implementing ECC with java standard edition 7. *Int. J. Comput. Sci. Artif. Intell.* **3**(4), 134–142 (2013)
22. Sheng, Z., (Member, IEEE), Mahapatra, C., (Student Member, IEEE), Zhu, C., (Student Member, IEEE), Leung, V.C.M., (Fellow, IEEE): Recent advances in industrial wireless sensor networks toward efficient management in IoT, vol. 3, 19 May 2015
23. Anggorojati, B., Mahalle, P.N., Prasad, N.R., Prasad, R.: Capability-based access control delegation model on the federated IoT network. IEEE (2012)
24. Hasebe, K., Mabuchi, M., Matsushita, A.: Capability-based delegation model in rbac. In: Proceeding of the 15th ACM symposium on access control models and technologies, ser. SACMAT'10. New York, NY, USA: ACM, pp. 109–118 (2010)

Enhanced Pixel Purity Index Algorithm to Find the Index Position of the Pure Pixels in Hyperspectral Images

S. Graceline Jasmine and V. Pattabiraman

Abstract This paper proposes an enhanced pure pixel index (PPI) algorithm for hyperspectral imaging. The conventional PPI algorithm uses random skewers in the process of finding the pure pixel indexes. The randomness in generating skewers leads to iterate the process multiple number of times to pick the most repeating pixel. The iteration involved in the process increases the computational time of the algorithm. The research contribution in this paper is that, the randomness in generating the skewers has been eliminated in the enhanced PPI algorithm by generating skewer by correlating the skewer with the given hyperspectral image. The proposed algorithm has reduced the computational complexity as well.

Keywords Hyperspectral imaging · Pixel purity index · Endmembers Spectral unmixing · Skewers

1 Introduction

Hyperspectral sensors are used for collecting a set of images. These images are discrete and are captured within the wavelength ranging from 0.4 to 2.5 μm . This collection of images can be viewed as a three-dimensional cube, where the third dimension is the number bands [1]. Hyperspectral imaging (HSI) focuses on analyzing and interpreting the spectra of a particular scene acquired by a hyperspectral sensor. The spatial resolution of the sensor determines the precision of an image which can further used to identify the materials available. The low spatial resolution in HSI leads to a problem known as mixed pixel. Mixed pixel is a pixel where many materials observed through a single pixel. But high spatial resolution HSI produces

S.G. Jasmine (✉) · V. Pattabiraman
School of Computing Science and Engineering,
VIT University, Chennai, India
e-mail: graceline.jasmine@vit.ac.in

V. Pattabiraman
e-mail: pattabiraman.v@vit.ac.in

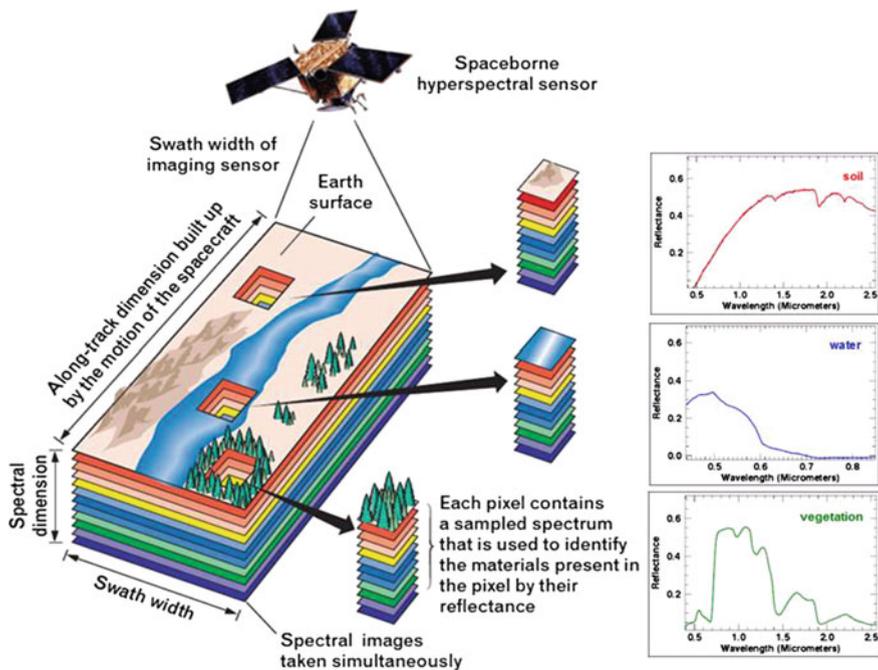
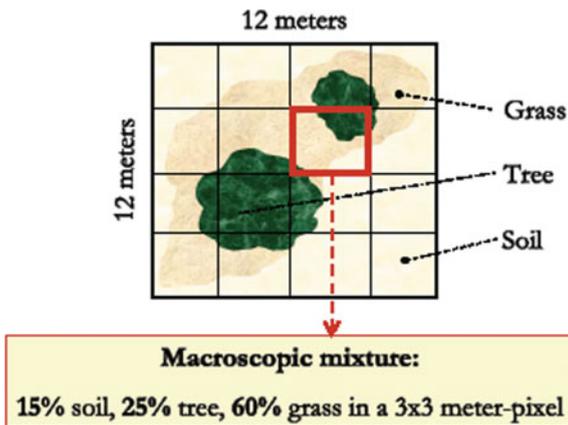


Fig. 1 Hyperspectral imaging

Fig. 2 Representation of mixed pixel



images containing both pure and mixed pixels. The mixed pixels are further classified into a set of classes where each class corresponds to a particular endmember and its abundance. Abundances indicate the proportion of each endmember present in the pixel. The concept of HSI is shown in Fig. 1.

Endmembers can be of any real-world material in the field of view, like leaf, wood, metal, or any living or non-living material. The result is a composite or mixed spectrum as shown in Fig. 2. The mixed pixel can be of linear or nonlinear form. The nonlinear nature of the pixel is the outcome of several reflections within the surface

location [2]. Nonlinear spectral mixture analysis (NLSMA) will increase the unmixing accuracy [3]. This normally occurs when materials are not spatially distinguishable.

2 Hyperspectral Unmixing Chain

Spectral unmixing has to undergo two major processes. Initially the locations of the pure pixels have to be found out. Pure pixels are the pixels containing the spectrum of a single material. Once the pure pixel locations are traced out, it has to be compared with the library spectra of the materials. Apart from these procedures, initially if the data has not done atmospheric correction, which has to be done. Moreover, dimensionality reduction can also be done to reduce the computational complexity. For some data, atmospheric and geometric correction would have done in the data available online in the NASA, Jet Propulsion Laboratory AVIRIS data website. Dimension reduction is not mandatory. The whole hyperspectral image can also be given as an input to the endmember extraction algorithm [4]. The endmember extraction algorithm is used to categorize the material substances.

3 Endmember Extraction

Endmember extraction is a process of finding the materials in the pixel under consideration. In the process of finding the endmembers, there are two common assumptions. The first assumption is that given an image there will be few pixels which contain the spectra signature of only one material. These kinds of pixels are known as pure pixels. The other assumption is that there are no pure pixels. There are many algorithms based on pure pixel assumptions. Some of them are pixel purity index (PPI) [5, 6], N-Finder (N_FINDR) [7], vertex component analysis (VCA) [8], and maximum volume by householder transformation (MVHT) [9]. Few algorithms which are based on no pure pixel assumption are minimum volume simplex analysis (MVSA) [10], the simplex identification via split augmented Lagrangian (SISAL) [11], dependent component analysis (DECA) [12], and convex cone analysis (CCA) [13, 14].

4 Proposed PPI Algorithm

The proposed PPI algorithm is given in the following steps:

1. Skewer generation: Mean is calculated for the original data. Below given is the mean for the first 10 bands. For the dataset with 224 bands, one skewer generated will be a unit vector of size 224×1 . But in the existing PPI algorithm,

k random unit vectors will be generated. Generally, k will be 10^4 and that has been reduced to one in this proposed algorithm.

2. Extreme projection:

All the data vectors, d_i where $i = 314,368$ (614×512), will be projected on the skewer S by a dot product, $(d_i \cdot S)$ to find the extreme points.

3. Purity score calculation:

Initially, the count for all the pixels will be marked as zero and after the step 2, the pixels in the extreme points will be incremented by 1.

4. Endmember selection:

So that pixels whose count marked as 1 will be considered as pure pixels and will be labeled as endmembers.

5 Results and Discussion

PPI algorithm is based on convex geometry which finds the data vectors having minimal or maximal orthogonal projections toward certain direction. As there is no prior information regarding the direction, random skewers were generated. Then the data vectors were projected on these skewers. After that, the pixels repeatedly lying on the extreme points were noted. The pixel having more count on the extreme position will be considered as pure pixel. But there are two issues in the conventional algorithm. The first issue is the number of skewers to be generated and the second issue is the fixing of the threshold value to determine the extreme points. These two issues lead to produce different results in different iterations.

Instead of generating random vectors for skewers, the idea incorporated is to generate mean value of all the layers which will form a vector of size $n \times 1$ where n is the number of bands in the dataset. So the skewer will be $n \times 1$ vector. Table 4 is provided with the skewer values generated by the proposed algorithm. This reduces the iteration involved in the conventional PPI algorithm. Thereby the proposed algorithm reduces the computational complexity. Moreover, as the skewers are generated based on the mean of the given dataset, the skewers are being correlated with the given dataset which removes the randomness in generating skewers. Once the skewer is generated, the data vectors will be projected on the skewer and the pixel purity score is calculated for all the pixels.

The pixel purity score is calculated by incrementing the count for the pixels lying in the extreme points. So the score will get incremented for the pixels if it lies above a particular threshold value. Then the pixels will be arranged in the order of its score. The pixels whose score are above a certain point will be considered as pure pixels. The original value of the pixels in each band is given in Table 1.

The dataset considered is cuprite. The enhanced PPI algorithm has been tested using AVIRIS cuprite scene. The scene selected for experimenting and processing in this paper is 970619t01p02_r02_sc04.a.rfi. This scene has 224 spectral bands, 614 samples, and 512 lines. Figure 3 shows the scene for a band of 150.

Table 1 Projected data of size 10×10 of the original data (224 bands \times 314,368 pixels)

Bands	Pixels									
	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	-0.005	-0.005	-0.005	-0.005	-0.005	-0.005	-0.003	-0.005	-0.005	-0.005
4	0.065	0.062	0.065	0.069	0.062	0.066	0.068	0.072	0.062	0.075
5	0.080	0.081	0.082	0.080	0.077	0.081	0.084	0.085	0.082	0.085
6	0.084	0.085	0.087	0.088	0.083	0.083	0.088	0.082	0.089	0.09
7	0.091	0.092	0.093	0.094	0.088	0.090	0.094	0.096	0.097	0.096
8	0.095	0.096	0.098	0.095	0.093	0.094	0.098	0.099	0.099	0.103
9	0.102	0.104	0.104	0.105	0.101	0.101	0.105	0.106	0.109	0.110
10	0.109	0.109	0.109	0.109	0.105	0.107	0.11	0.113	0.115	0.116

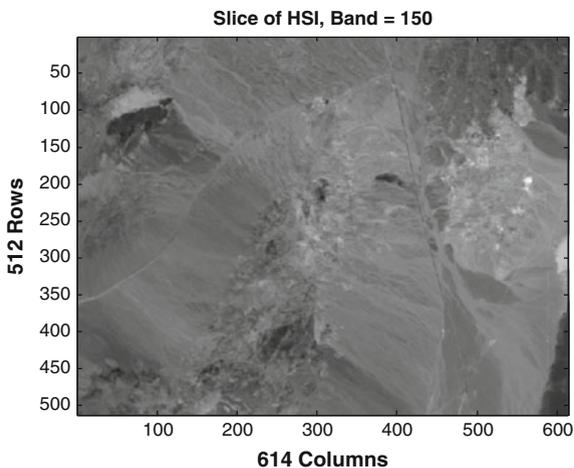


Fig. 3 Cuprite scene image at band = 150

Table 2 Iteration 1:10 skewers of size: 224×10

Skewers									
1	2	3	4	5	6	7	8	9	10
0.59	0.31	0.10	2.27	0.35	1.61	0.01	0.10	0.35	0.15
0.44	0.08	1.31	0.07	0.16	1.69	0.75	0.24	1.41	0.41
1.35	1.01	0.69	1.03	2.00	0.77	0.81	0.53	0.69	0.84
0.00	0.09	0.03	1.23	0.73	1.11	0.35	0.66	0.61	0.95
1.16	1.70	0.17	0.76	1.95	0.14	1.71	0.81	2.20	0.01
0.94	0.69	1.31	0.45	1.41	1.53	1.21	0.56	0.12	0.45
0.92	0.38	0.57	1.10	0.18	0.77	1.58	0.33	0.98	1.51
2.30	1.59	1.18	0.83	1.14	0.40	1.03	0.21	0.72	0.74
0.72	0.16	0.06	0.33	0.12	0.39	0.31	0.84	0.37	0.63
1.74	1.90	3.04	0.99	0.75	0.26	3.29	0.63	0.67	0.69

In the original PPI algorithm, the skewers were generated using rand function in Matlab and the values were converted to absolute values. In order to improve the accuracy of pure pixel index position the data will be repeatedly projected to different random skewers. The skewers generated in the first iteration are given in Table 2 and the skewers generated in the second iteration are given in Table 3.

The pixel purity index calculated in each of the iteration using existing PPI algorithms is given in Table 5. From Table 5 it is clearly visible that the pixel positions which are considered to be pure are getting varied in all the iterations.

The index positions of the pure pixels estimated by the proposed PPI algorithm are listed in Table 6. From Tables 5 and 6 it can be observed that few values are similar for the existing and the proposed PPI algorithms.

Table 3 Iteration 2:10 skewers of size: 224×10

Skewers									
1	2	3	4	5	6	7	8	9	10
0.16	1.89	0.65	1.16	1.62	0.63	0.28	0.49	1.40	1.34
0.04	0.13	0.33	0.93	0.41	0.90	0.51	0.06	0.20	1.21
1.18	0.54	2.42	1.75	0.55	0.46	1.08	0.04	0.44	0.45
0.12	0.36	1.68	0.14	1.61	0.60	0.11	0.54	1.70	1.16
1.44	0.44	0.78	1.11	0.31	0.67	0.01	0.38	1.11	0.19
0.94	0.16	0.67	0.21	1.44	1.22	0.14	0.06	1.61	0.27
3.19	0.04	0.30	0.26	1.53	0.69	0.03	0.17	0.47	0.29
1.71	1.11	1.18	0.71	2.51	0.70	1.80	1.82	1.51	1.16
0.96	0.62	0.23	1.57	0.34	1.15	1.06	0.81	1.35	0.14
1.14	1.27	0.12	0.16	0.26	0.58	0.20	0.85	0.27	0.66

Table 4 Skewer of size 224×1

Band	Skewer mean
1	0
2	0
3	0.001899773
4	0.07091743
5	0.088343226
6	0.095203063
7	0.103256327
8	0.10785971
9	0.115068051
10	0.119411729

Table 5 Index positions of pure pixels listed in decreasing order for five iterations using existing PPI algorithm

Pixel purity index in each iteration				
1	2	3	4	5
107,703	107,703	107,703	107,703	107,703
89,161	79,069	79,069	79,069	89,161
130,751	90,722	90,722	90,722	90,721
134,822	108,216	108,216	108,216	91,208
138,415	113,802	113,802	113,802	13,0751
227,535	124,580	124,580	124,580	131,772
241,078	239,028	239,028	239,028	134,309
256,896	254,843	254,843	254,843	135,333
309,931	1	1	1	1

Table 6 First five index positions of pure pixels using proposed PPI algorithm

Pure pixel index by proposed PPI					
Pixel index	107,703	131,263	113,802	130,751	88,648

6 Conclusion

This paper proposed an enhanced pure pixel index algorithm in order to reduce the complexity of the existing PPI algorithm. The iteration involved in the conventional PPI algorithm proportionally increases the computational complexity of the algorithm. In the proposed enhanced pixel purity index algorithm, the random generation of skewers has been addressed thereby reducing the number of skewers required for finding the pure pixels. Some of the pure pixel index positions returned by the enhanced PPI algorithm are same as the existing one.

References

1. Valero, S., Salembier, P., Chanussot, J.: Hyperspectral image representation and processing with binary partition trees. *IEEE Trans. Image Process.* **22**, 1430–1443 (2012)
2. Nareenart, R., Qian, D.: Nonlinear spectral mixture analysis for hyperspectral imagery in an unknown environment. *IEEE Geosci. Remote Sens. Lett.* **7**, 836–840 (2010)
3. Yoann, A., Nicolas, D., Jean, T.: Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model. *IEEE Trans. Image Process.* **22**, 1267–1275 (2013)
4. Jasmine, G.S., Pattabiraman, V.: Hyperspectral image analysis using end member extraction algorithm. *Int. J. Pure Appl. Math.* **101**, 809–829 (2015)
5. Keshava, N.: A survey of spectral unmixing algorithms. *Lincoln Lab. J.* **14**, 55–78 (2003)
6. Boardman, J.M., Kruse, F.A., Green, R.O.: Mapping target signatures via partial unmixing of AVIRIS data. In: *Summ. JPL Air-borne Earth Sci. Workshop*, pp. 23–26 (1995)
7. Winter, M.E.: N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data. In: *SPIE Conference Imaging Spectrometry*, pp. 266–275 (1999)
8. Nascimento, J., Bioucas-Dias, J.: Vertex component analysis: a fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **43**, 898–910 (2005)
9. Liu, J.M., Zhang, J.S.: New maximum simplex volume method based on householder transformation for endmember extraction. *IEEE Trans. Geosci. Remote Sens.* **50**, 104–118 (2012)
10. Li, J., Bioucas-Dias, J.: Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data. In: *IEEE Geoscience Remote Sensing Symposium*, pp. 2369–2371 (2008)
11. Bioucas-Dias, J.: A variable splitting augmented lagrangian approach to linear spectral unmixing. In: *1st workshop Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pp. 1–4 (2009)
12. Nascimento, J., Bioucas-Dias, J.: Hyperspectral unmixing algorithm via dependent component analysis. In: *IEEE Geoscience Remote Sensing Symposium*, pp. 4033–4036 (2007)
13. Ifarraguerri, A., Chang, C.I.: Multispectral and hyperspectral image analysis with convex cones. *IEEE Trans. Geosci. Remote Sens.* **37**, 756–770 (1999)
14. Junmin, L., Jiangshe, Z., Gao, Y., Chunxia, Z., Li, Z.: Enhancing spectral unmixing by local neighborhood weights. *IEEE J. Select. Topics Appl. Earth Obser. Remote Sens.* **5**, 1545–1552 (2012)

Assamese Character Recognition Using Zoning Feature

Kalyanbrat Medhi and Sanjib Kr. Kalita

Abstract The aim of this paper is to develop a model to recognize basic Assamese characters using feed-forward neural network. The basic characters included in Assamese language are a set of numeral, a set of vowel, and a set of consonant. An algorithm has been designed to segment the line and individual character of the image and zoning features are extracted from the individual character. The network is trained by gradient descent with momentum and adaptive learning rate back-propagation training function. The network consists of two hidden layers with Sum Square Error (SSE). Finally, the unicode value of the recognized character is written in a text file.

Keywords Assamese character recognition · Feed-forward neural network Segmentation · Zoning · Back propagation

1 Introduction

In pattern recognition it is still a challenging task to produce an editable document from handwritten or machine-written characters. There are two ways to recognize characters: recognize while writing or recognize after writing. The former is called online and the latter is offline recognition technique. This experiment tries to implement offline technique. Character recognition system is used in bank for cheque processing and sorting, where the digits of cheque numbers are special font printed with magnetic ink that is prone to noise. Moreover, mobile applications are available that allows user to extract information automatically from visiting card of a person to their mobile by taking a snapshot of the card. In post office it can be

K. Medhi (✉) · S.Kr. Kalita
Department of Computer Science, Gauhati University, Guwahati, India
e-mail: kalyanbratmedhi@gmail.com

S.Kr. Kalita
e-mail: sanjib959@rediffmail.com

used for postal number reading and sorting. Publication house can rapidly produce editable document from old books for publishing using CR system.

In India, OCR system has been developed for Bangla, Devanagari, Gurumukhi, Kannada, Malayalam, Tamil, and Telugu [1]. Thus, OCR systems for Indian scripts have just started appearing. Literature review reveals that research on Assamese language is very tiny and that is why this is an attempt to design a CR system for this language.

2 Properties of Assamese Scripts

The script used in Assamese language is originated from an ancient Indian script called Brahmi [2]. Most of the characters in Assamese script have a horizontal line from where it hang, called the head line or matra, e.g., ক, ঘ, অ, আ, ই, etc. The vowel can present independently in a word or they may act as modifier with consonant. The consonant can be independent, but for pronunciation they took help of vowel. Combination of two or three consonants can produce another compound character.

3 Literature Review

Sharma and Gupta [3] proposed a system with a k-nn classifier with $k = 1$ for classification. The 4×4 zonings are used as features. The dataset used in the experiment contains 10,000 and 5400 training and test data, respectively. They have claimed that the maximum recognition rate is 99.89%.

In an experiment, Shanthiand and Duraiswamy [4] proposed a support vector machine-based handwritten Tamil character recognition system. After collection of images they are converted into binary using Otsu's method and skeletonization is done using Hilditch algorithm. All the segmented character images are normalized into 32×32 using bilinear interpolation technique. Different zones of size 2×2 , 4×4 , and 8×8 were used as features. It is noted that 4×4 zone yields best accuracy for test data. They have claimed that recognition accuracy is 82.04% for all the 34 characters.

An offline Arabic handwritten text recognition system is designed by Parvez et al. [5], based on structural techniques. In the preprocessing phase, from the segmented text lines the words/sub-words are extracted and slant correction is done. Each segmented sub-word is corrected and verified by a recognizer to improve any over- or under-segmentations. A nearest neighbor (NN) is used as the classifier which gives an average accuracy of 97.18% for ADBase database of Arabic numeral. 79.58% accuracy is reported for IfN/ENIT database of isolated characters with fuzzy-attributed turning function (FATF) classifier.

Isolated handwritten characters recognition in Gurumukhi script is done by Sharma and Jhaji [6]. Initially, original image is normalized to 48×48 and created 64 zones to find zonal densities that are taken as features. The K-nearest neighbors (K-NN) classifiers and support vector machine (SVM) with linear kernel, polynomial kernel, and radial basis kernel have been used for classification process. It has been claimed when 2050 images are used; 72.68, 73.02, and 72.83% accuracies found by SVM with linear, polynomial, and radial basis function (RBF) kernel are used, respectively. The K-nn classifier gives 72.54% recognition accuracy.

4 Methodology

4.1 Data Collection and Preprocessing

In this experiment characters are captured through a scanner with resolution of 300 dpi. The scanned image is first converted into grayscale and then to binary image using well-known global thresholding and Otsu's method. After this, noise is removed by opening the image morphologically. Figure 1a, b shows a binary image with noise and after removing noise, respectively.

The preprocessed character needs to be cropped before feature extraction.

4.2 Character Segmentation

In the present study an algorithm is designed to segment the individual character from the image based on pixel value of the image. The algorithm can successfully crop non-touching character from an image. Two-line segmentation algorithm for Assamese and MeeteiMayek Script is proposed by Kumar and Kalita [8, 9]. Our algorithm differs from it as there are two steps in the algorithm: line segmentation

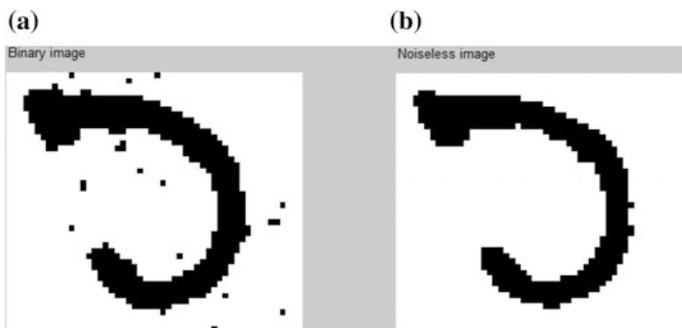


Fig. 1 a Binary image with noise. b After removing noise

and character segmentation. First it is necessary to segment the lines followed by character segmentation. The line segmentation algorithm is as follows:

- Step1. Read the binary image to variable I.
- Step2. Find the width and height of the image and put the value in variable x and y, respectively.
- Step3. Set $x1 = 0$, $y1 = 0$, $x2 = 0$, $y2 = 0$, $cell_s = 1$.
- Step4. Assume height of text line height = 0 and width of text line width = 0.
- Step5. Starting with $i = 1$.
- Step6. $white = 1$.
- Step7. Starting with $j = 1$.
- Step8. If $I(i,j) = 0$, set $white = 0$, and go to next step. If $I(i,j) \neq 0$, increment the value of j by 1 and continue this step till $j < y$. When $j = y$, go to step 9.
- Step9. If $white = 0$ && $x1 == 0$ && $y1 == 0$, set $x1 = 1$, $y1 = i-1$.
- Step10. If $white = 1$ && $x1 == 1$ && $y1 > 0$, set $width = y$, $y2 = i$, $height = y2 - y1$.
- Step11. If $x1 > 0$ && $y1 > 0$ && $width > 0$ && $height > 0$, crop the image $IL1 = \text{imcrop}(I, [x1 \ y1 \ width \ height])$, store the line image in cell array $ILarray\{cell_s\} = IL1$ and set $x1 = 0$, $y1 = 0$, $width = 0$ and $height = 0$, $cell_s = cell_s + 1$
- Step12. Increment the value of i by 1 and go to step 6 if $i < x$ otherwise stop.

After the completion of all line segmentation in the image the character segmentation process begins by reading image one by one from the cell array. The algorithm for character segmentation is as follows:

- Step1. Find maximum subscript of cell array 'ILarray' and put the value in maxc variable and set $cell_s = 1$.
- Step2. Read the line image to variable IL from the cell array 'ILarray' with subscript c = 1.
- Step3. Find the width and height of the image and put the value in variable x and y, respectively.
- Step4. Set $x1 = 0$, $y1 = 0$, $x2 = 0$, $y2 = 0$.
- Step5. Assume height of text line height = 0 and width of text line width = 0.
- Step6. Starting with $i = 1$.
- Step7. $white = 1$.
- Step8. Starting with $j = 1$.
- Step9. If $IL(j,i) = 0$, set $white = 0$, and go to step 10. If $IL(i,j) \neq 0$, increment the value of j by 1 and continue this step till $j < x$. When $j = x$, go to step 10.
- Step10. If $white = 0$ && $x1 == 0$ && $y1 == 0$, set $y1 = 1$, $x1 = i-1$.
- Step11. If $white = 1$ && $y1 == 1$ && $x1 > 0$, set $height = x$, $x2 = i$, $width = x2 - x1$.
- Step12. If $x1 > 0$ && $y1 > 0$ && $width > 0$ && $height > 0$, crop the image $ILC1 = \text{imcrop}(IL, [x1 \ y1 \ width \ height])$, store the character image in cell array $ILCarray\{cell_s\} = ILC1$ and set $x1 = 0$, $y1 = 0$, $width = 0$ and $height = 0$.

Step13. Increment the value of i by 1 and go to step 7 if $i < y$ otherwise go to step 14.

Step14. Increment the value of c by 1 and go to step 2 if $c \leq \text{maxc}$ otherwise stop.

After executing this algorithm, all the segmented character can be obtained from cell array 'ILCarray' and required feature can be extracted from the image.

4.3 Feature Extraction

In this experiment we used static zoning method of size 4×4 , 6×6 , and 8×8 . After the segmentation of character we normalized the image to 60×48 sizes so that shape of each character is preserved. An outstanding literature on zoning feature for handwritten character recognition can be found in [7].

4.4 The Classifier

The feed-forward neural network consists of a series of layers and each layer has a lot of neurons. The inputs to every neuron come from the output of previous layer's neurons. The first layer receives the feature vector and the last layer gives the result of classification. In the first layer number of neuron must be same as the length of feature vector and on the other hand in last layer number of neurons must be equal to number of pattern class. The number of neurons in each of the hidden layers is a choice of network designer, but in between the number of neuron of output layer and input layer.

All the actual processing is performed in hidden layers and output of layer is calculated by a transfer function from its net input.

The mathematical equation to calculate output at every neuron using sigmoid function is defined as follows:

$$a_{jm} = \frac{1}{1 + e^{-S_{jm}}}, \tag{1}$$

where

$$S_{jm} = \sum_{x=0}^n w_{ijx} a_{ix}, \tag{2}$$

where a_{jm} is the activation code of receiving neuron m in layer j , S_j is the sum of the products of the activations of all relevant "emitting" nodes by their respective weights, and w_{ij} is the set of all weights between layers i and j that are associated with vectors that feed into neuron m of layer j .

5 Result and Discussion

Experiment is carried out in three different architectures for digits, vowels, and consonants. In all experiment the feed-forward neural network has two hidden layers with one input and one output layer. The network consists of 10, 11, and 38 input nodes for recognizing digits, vowels, and consonant, respectively.

5.1 Digits Recognition

Figure 2 shows a sample of Assamese handwritten digits.

The numbers of neurons in first hidden layer are 25 and in second hidden layer are 10 for digits recognition. After preprocessing and cropping the image size is normalized to 60×48 and feature are extracted from each 10×10 zone of the image. A feature vector for digit 'zero' is depicted in Fig. 3.

This feature vector is provided as input into the input layer of the network. The following Table 1 depicts the parameters used in experiment for digits.

The back propagation algorithm adjusts the weights of the links and finally the pattern is put in appropriate pattern class.



Fig. 2 Handwritten Assamese digits

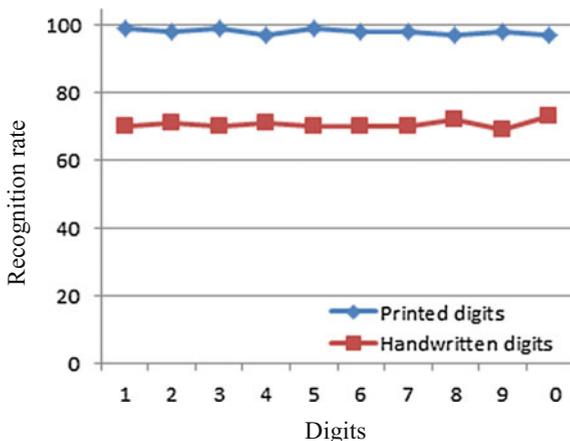
[0.11	0.55	1.00	0.95	0.12	0.44	1.00	0.55	0.49	0.80	0.96	0.62	0.00
0.10	1.00	1.00	0.10	0.00	0.10	1.00	0.96	0.10	0.00	0.62	1.00	0.49
0.50	0.55	1.00	0.80	0.12	0.95	1.00	0.55	0.11]				

Fig. 3 Feature vector for digit zero

Table 1 Parameter use in feed-forward neural network for digit recognition

Parameter	Value	Parameter	Value
Learning rate	0.01	Maximum number of epochs to train	2000
Momentum constant	0.95	Minimum error exist in the network	0.1
Performance goal	0.1	Initial weights and biased term values	Randomly generated values between 0 and 1

Fig. 4 Recognition rate of printed and handwritten Assamese digits



The training set of digits contains total of 500 digits with 400 handwritten and 100 printed digits and the testing set consists of 600 new digits together with 500 training digits and in total 1100 digits. It is seen that for printed digits recognition accuracy is 98% and for handwritten digits recognition rate is 70.6%.

Figure 4 shows recognition rate of printed and handwritten Assamese digits.

The numbers of neurons in output layer for vowel recognition process are 11. The numbers of neurons in first hidden layer are 25 and in second hidden layer are 10.

5.2 Vowels Recognition

Figure 5 shows a sample of Assamese handwritten vowels.

After preprocessing and cropping the image is normalized to 60×48 and features are extracted from each 10×10 zone of the image. A feature profile for vowel ঔ (Ou) is depicted in Fig. 6.

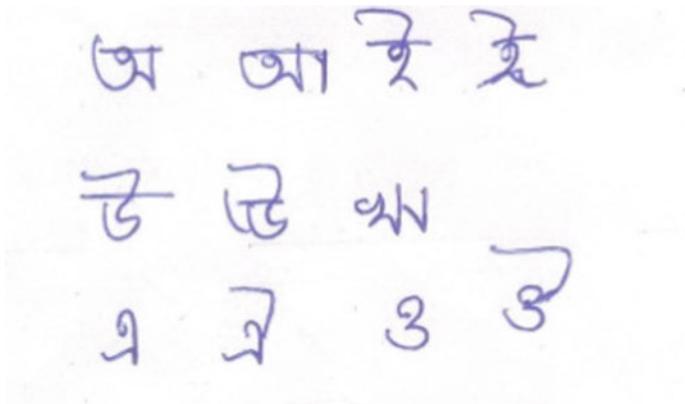
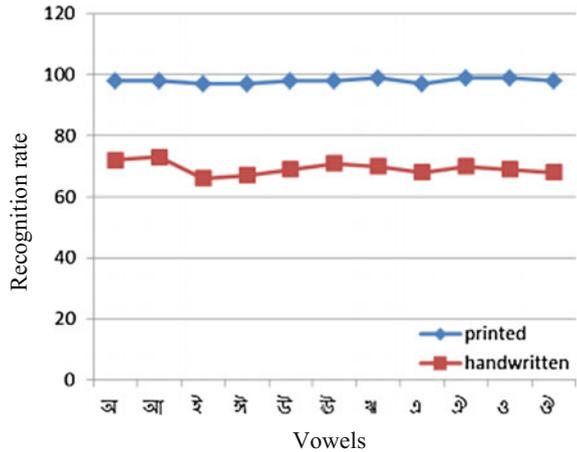


Fig. 5 Handwritten Assamese vowels

[0.00	0.42	0.02	0.00	0.00	0.00	0.21	0.42	0.12	0.00	0.00	0.00	0.09	0.31	0.32	0.18
	0.26	0.33	0.49	0.41	0.33	0.27	0.10	0.53	0.29	0.19	0.21	0.00	0.37	0.13	0.00	0.36
	0.40	0.50	0.03]												

Fig. 6 Feature vector for vowel ঔ(Ou)

Fig. 7 Recognition rate of printed and handwritten Assamese vowels



In case of vowel recognition the learning rate is reduced to 0.001 in order to improve performance. The training set of vowels contains total 550 vowels with 440 handwritten and 110 printed vowels and the testing set contains 660 new vowels together with 550 training vowels and in total 1210 vowels. It is seen that for printed vowels recognition accuracy is 98% and for handwritten vowels recognition rate is 69.63%. The best performance during training phase is achieved at 160 epochs.

Figure 7 shows recognition rate of printed and handwritten Assamese vowels (Fig. 7).

5.3 Consonants Recognition

In the experiment for consonant recognition, the learning rate is again reduced to 0.0001 in order to improve performance. The training set of consonants contains total 2050 consonants with 1640 handwritten and 410 printed consonants and the testing set contains 2460 new consonants together with 2050 training consonants and in total 4510 consonants. It is seen that for printed consonants recognition accuracy is 98% and for handwritten consonants recognition rate is 71.23%. The best performance during training phase is achieved at 160 epochs. Figure 8 shows a sample Assamese consonants (Fig. 8)

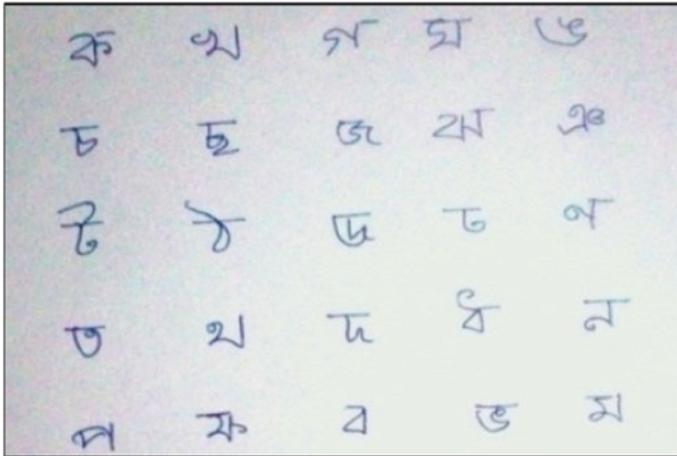


Fig. 8 A sample of handwritten Assamese consonants

Table 2 Recognition rate of digits, vowels, and consonants

Character	Printed (%)	Handwritten (%)
Digits	98	70.6
Vowels	98	69.62
Consonants	98	71.23

6 Recognition Rate

The recognition accuracies for digits, vowels, and consonants are shown in Table 2

7 Conclusion

The segmentation algorithm successfully segments lines and non-touching characters. The efficiency of the recognizer has been examined using zoning features. It is observed that for printed characters the recognition rate is 98% irrespective of digits, vowels, and consonant. Recognition rates for digits, vowels, and consonants are 70.6, 69.62, and 71.23%, respectively. The recognition rate for handwritten consonant characters is found to be maximum.

This work can be extended for other features such as diagonal, slice zoning, etc; moreover, these features could be implemented in fuzzy or neuro fuzzy to increase the recognition rate.

References

1. http://tdil-dc.in/index.php?option=com_vertical&parentid=77. Accessed on 1 June 2014
2. Medhi, K., Kalita, S.K.: Recognition of assamese handwritten numerals using mathematical morphology. In: IEEE International Advance Computing Conference (IACC), pp. 1076–1080 (2014)
3. Sharma, D., Gupta, D.: Isolated handwritten digit recognition using adaptive unsupervised incremental learning technique. *Int. J. Comput. Appl.* **7**, 27–33 (2010)
4. Shanthi, N., Duraiswamy, K.: A novel SVM-based handwritten Tamil character recognition system. *Pattern Anal. Appl.* **13**, 173–180 (2010)
5. Parvez, M.T., Mahmoud, S.A.: Arabic handwriting recognition using structural and syntactic pattern attributes. *Pattern Recogn.* **46**, 141–154 (2013)
6. Sharma, D., Jhaji, P.: Recognition of isolated handwritten characters in Gurumukhi script. *Int. J. Comput. Appl.* **4**, 9–17 (2010)
7. Impedovo, D., Pirlo, G.: Zoning methods for handwritten character recognition: a survey. *Pattern Recogn.* **47**, 969–981 (2014)
8. Kumar, C.J., Kalita, S.K.: Line segmentation in handwritten Assamese and Meetei Mayek script using algorithm. In: ISTE Sponsored International Conference on Research and Innovations in Engineering and Technology, Amritsar College of Engineering and Technology (2014)
9. Kumar, C.J., Kalita, S.K.: Line segmentation in handwritten Assamese and Meetei Mayek script using algorithm. In: Optical Society of India Sponsored International Conference on Opto Electronics and Applied Optics, IEM Engineering Campus, Springer Proceedings in Physics (2014)

Improving Convergence in iBGP Route Reflectors

Shipra Shukla and Mahesh Kumar

Abstract Convergence is a crucial challenge for Border Gateway Protocol (BGP). The lower convergence time can improve network performance attributes such as end-to-end delay and throughput. In literature, various studies suggested that MRAI timers improve the convergence in iBGP. The effect of MRAI timer examined extensively in past. However, we found very less literature on the effect of MRAI timer in iBGP route reflectors. Route reflectors avoid too many iBGP sessions in the internal BGP. However, it creates many routing paths in iBGP. Moreover, when there is a change in the topology, the sharing of updates very frequently degrades convergence in the route reflectors. Therefore, this paper measures the effect of adjustable MRAI timer and suggests for better convergence in iBGP route reflectors. We measured the performance parameters convergence time, the number of exchanged updates, throughput, and delay using adjustable MRAI timer and compared with standard BGP.

Keywords BGP · Intra-domain routing · Convergence · MRAI timer
Route reflectors

1 Introduction

The Internet consists of several self-administered autonomous systems. All autonomous systems (ASes) follow a rule-based policy to share the inside and outside path information. It is required to use a robust path vector routing protocol

S. Shukla · M. Kumar (✉)

Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna, Madhya Pradesh, India
e-mail: mahesh.chahar@gmail.com

S. Shukla

e-mail: ershiprashukla88@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_40

for sharing the path information. In ASes, Border Gateway Protocol (BGP) is a widely used inter-domain routing protocol.

An AS is a set of routers. In external BGP, the selected BGP speakers spread path information among ASes. Similarly, in internal BGP, routers spread path information learned by eBGP. In intra-domain, earlier concepts used fully mesh iBGP topologies, but it suffers from several limitations such as partial scalability, complex configurations, etc. In a fully mesh iBGP topology, all routers connect with each other and build approximately $n * (n - 1)/2$ iBGP sessions within an AS [1]. These topologies are very complex and useful for small ASes only. The concept of route reflector came into existence to avoid too many iBGP sessions inside the AS.

In intra-domain, BGP selects some special routers to share the path information of routers (either learned from inside or outside router) within the AS. These special routers are known as route reflectors that spread path information of routers using iBGP sessions. Each route reflector has a unique origin id and cluster id. The origin id of route reflector helps to avoid the loop. The route reflector client and non-client connect with routers reflector in a hierarchy. In a cluster, the route reflectors and the route reflector clients share the same cluster id.

The route reflectors share the best path and follow certain rules to share the route information [2]. A route reflector can share the route received from a route reflector client with connected route reflector client and non-client. A route learned by eBGP peer can send to route reflector client, route reflector non-client, and eBGP neighbor. A route learned by route reflector client can send to eBGP peer, route reflector, and route reflector non-client. A route learned by route reflector non-client can send to eBGP peer and route reflector. A route learned by route reflector client cannot send to route reflector non-client.

The routing keeps on changing within the AS. It causes some issues such as loop prevention [3], convergence time, reduction of exchanged updates, routing lags on routing failures [4], route oscillation [3], etc.

According to path selection policy, a route reflector selects a path to a destination and if it finds a better path then replaces the current path in route information base (RIB). Moreover, the sharing the information of changed paths very frequently degrades networks performance. Hence, there is a requirement to hold the updates for a definite time to select a path when updates increase rapidly. Researchers [5–8] suggested the use of MRAI (Minimum Route Advertisement Interval) for addressing BGP convergence problem.

MRAI timer holds the updates for certain time interval. Meanwhile, the router selects the best path to a destination and shares fewer updates with other routers. In this way, MRAI timer gets better convergence time, but MRAI timer can exacerbate convergence [6] also if there is an unnecessary hold of updates. Hence, the implementation of MRAI timer matters a lot. The MRAI timer should be adjustable to improve the convergence.

The authors implemented the different types of MRAI timers. Different ASes uses diverse MRAI timer such as per-destination [1], and per-peer that can affect the convergence and network performance attributes [9]. The use of MRAI timers can be significant in iBGP, and that motivated us to examine the effect of MRAI timer in iBGP route reflectors. Therefore, this work measures the effect of adjustable MRAI timer in route reflectors. The performance metrics convergence time, throughput, the number of exchanged updates, and end-to-end delay has been recorded and compared with standard iBGP MRAI timer.

The remaining sections are structured as follows. In Sect. 2, the related describes the techniques present in the literature. The internal BGP convergence problem on route reflectors has been discussed in Sect. 3. Section 4 proposes the solution for decreasing the convergence time in BGP using adjustable MRAI timer. Section 5 explains the simulation setup and Sect. 6 evaluates the performance of the approach. In Sect. 7, we conclude the paper.

2 Related Work

Convergence time has always been an important concern in BGP. RFC 1771 suggested the value of MRAI should be 5 s in internal BGP and 30 s in external BGP.

The authors of [7] studied the effect of MRAI in BGP and found that as the MRAI value increased the convergence time also increased initially and then went down. Hence, a constant value of MRAI should not apply to all types of networks. The MRAI timer should be used according to topology and traffic conditions.

Vissicchio et al. [3] studied different cases of iBGP policies and concluded that iBGP policies might exacerbate the convergence of BGP. It is important to optimize the iBGP routing policy correctly.

In literature, there are different kinds of MRAI timer in BGP such as per-node [5], per-destination [1], and per-peer [4] that affect convergence time significantly.

The authors of [4] implemented per-peer and per-prefix MRAI timer for removing routing lag after routing failures occur in the backbone networks.

The authors of [8] suggested the adaptive MRAI timers. The product of the number of MRAI timers and granularity is constant. Granularity is a constant value that determines the number of MRAI timers. However, the approach does not characterize the traffic load and not performs well under high traffic load.

3 Convergence Problem in Route Reflectors

In an iBGP fully mesh topology, there is a direct connection between two nodes. If any change occurs in the networks, then node gets the updates directly from each other. In hierarchical route reflectors, route reflectors clients and route reflectors

connect with each other in a hierarchy and route reflectors clients get an update via route reflector [10]. The hierarchy increases the number of hops. Furthermore, there is an increase in the number of paths for path selection, and if any change occurs in the topology, then the number of updates enhances rapidly.

MRAI timer limits the advertisement of updates. In past, both per-destination and per-peer MRAI timers are used. The standard iBGP uses per-destination MRAI timer [1]. However, per-destination MRAI timer has partial scalability, while per-peer MRAI timer has potential to delay the updates of connected nodes of peers.

The above approaches are not providing flexibility to change the duration of MRAI timer when traffic load increases. Therefore, we propose adjustable MRAI timers that use features of both per-destination and per-peer MRAI timer.

4 Adjustable MRAI Timers

The objective of the presented algorithm is to reduce the convergence time of iBGP in networks using MRAI timers.

The adjustable MRAI timers limit the updates of different destinations. In default MRAI [1], the value of MRAI timer is 5 s. In proposed algorithm 1, the route reflector router starts X MRAI timers. The difference between starting time of two MRAI timers is X/p seconds. The router identifies the prefix load, and if it is greater than the upper bound U_{prefix} , then the algorithm uses adjustable MRAI timer; otherwise use default MRAI timer. Prefix load is the number of prefixes that are advertised by the peers. The value of p can be adjusted to change the start time of MRAI timer. 'm' is a predefined time value that is assigned by the network operator.

Per-destination approach starts an MRAI timer for each destination [1]. In contrast to a per-destination MRAI timer, adjustable MRAI timer associates all updates advertised between the time interval K_0 and $K_0 + X/p$ with timer X_0 where $\{p = 1, 2 \dots i\}$. Adjustable MRAI timer X_0 starts at time K_0 . Similarly, next MRAI timer X_1 starts at $K_0 + 2 * X/p$, X_2 starts at $K_0 + 3 * X/p$, and so on.

The adjustable MRAI timer maintains a queue similar to per-peer MRAI timer. All non-converged routes associate with the recent running MRAI timer in a queue. Route reflector router holds the updates unless MRAI timer expires. It deletes the prefix from the queue when MRAI expires. An expired timer can be restart. The algorithm resets the timer X_0 when all X_i adjustable MRAI timers expire.

Algorithm 1

```

1: Begin-Procedure
2: Route Reflector Router starts
3: // Identify the prefix load ( $U_{\text{prefix}}$ ):
4: If prefix load is high then
5:   Starts X MRAI timers with the difference of p seconds
6:     Use the adjustable MRAI timer
7:     Set  $p=m$ 
8:     If change occurs in the networks then
9:       Associate the prefix with the Recent-
         Running MRAI timer in a queue
10:      If timer expires then
11:        Share the Update
12:        Delete the prefix
13:        Reset the timer
14:      End if
15:    End if
16: End if
17: Else if
18:   Use default MRAI timer
19: End if
End Procedure

```

5 Simulation Setup

We measured the results in network simulator (NS2) version 2.34 [11]. The performance parameters, convergence time, end-to-end delay, the number of exchanged updates, and throughput of the topology, have been computed. The proposed approach adjustable MRAI timer is compared with standard iBGP MRAI timer [1] in route reflectors.

The BGP extension NS-BGP 2.34 [12] is used for simulation of iBGP route reflectors. Figure 1 shows the simulation topology. This topology is inspired by topology described in [13]. There are 15 nodes in the topology including six route reflectors and seven route reflector clients. The IP addresses range from 10.0.0.1 to 10.0.14.1.

All links in the topology are duplex each having 1 Mbps bandwidth and 1 ms delay. In an autonomous system, there are three clusters and each cluster has a route reflector. All route reflectors have a cluster id. We have selected route reflector randomly. The simulation runs for 1000 s.

All links are up at the same time. The network prefixes are advertised at 0.25 s. The prefix advertisement rate increases from 4 to 16. In adjustable MRAI timers,

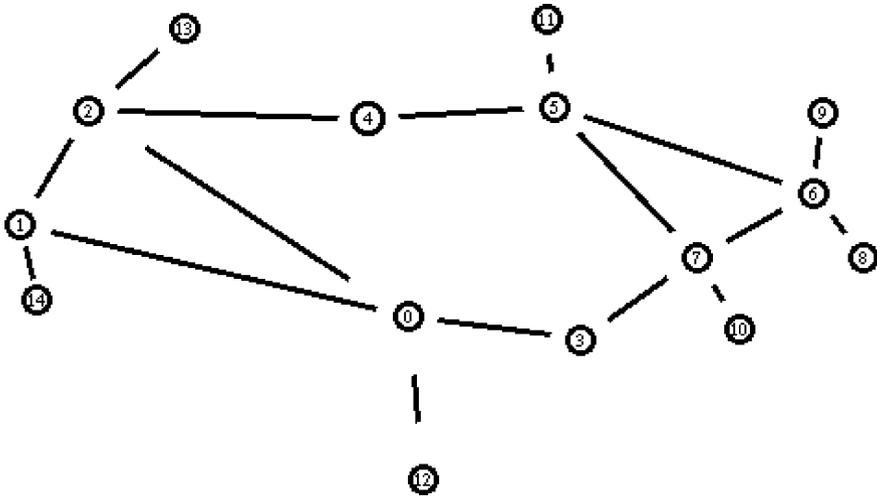
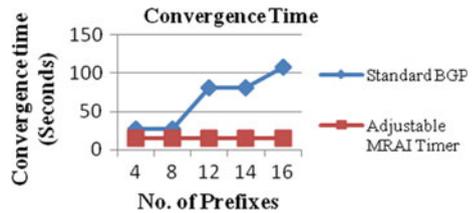


Fig. 1 Simulation topology

Fig. 2 Convergence time of the route reflectors



the value of X , p , and m is 5, 5, and 0.5, respectively. The CBR (Constant Bit Rate) of traffic is 1.0 Mbps. The traffic moves from the node 4 to node 3. In standard internal BGP [1], the value of MRAI timer is 5 s.

6 Performance Evaluation

To evaluate our approach, we recorded the performance metrics such as convergence time, number of exchanged updates, end-to-end delay, and throughput of the topology (presented in Fig. 1) using adjustable MRAI timers. Line charts compare the performance of adjustable MRAI timers with standard iBGP [1] in route reflectors.

Figure 2 compares the convergence time of the topology using both approaches. Adjustable MRAI timer converges earlier than standard BGP. The convergence time presents the time interval between starting of the time when any change take place in the topology until all routers get an update of modification.

In standard BGP route reflector, the convergence time increases with increase in the number of prefixes, while it remains almost constant and much lower using

adjustable MRAI timers. The convergence time increases from 27.25 to 108.25 in standard BGP which is approximately 44.5–85.9% higher than adjustable MRAI timer.

Figures 3, 4, and 5 show the graphs of the number of exchanged updates, throughput (kbps), and end-to-end delay (ms) in the networks. The number of exchanged updates denotes the communication overhead of the routing protocol.

In comparison to standard BGP, adjustable MRAI timer shows the minor increase in the number of exchanged updates in Fig. 3, but it puts a very nominal effect on other performance parameters as illustrated in Figs. 4 and 5. The throughput of the networks is higher in the adjustable MRAI timer approach in comparison to standard iBGP approach. In our approach, the average time spend in the queue of the packets is lower than the standard BGP that contributes in throughput and guaranteed end-to-end delay of the networks.

Figure 5 presents that the end-to-end delay (ms) that vary with the increase of prefix advertisements. In comparison to standard BGP, the end-to-end delay of adjustable MRAI timer is almost similar. Hence, our approach performs better than the compared approach.

Fig. 3 Exchanged updates of both approaches

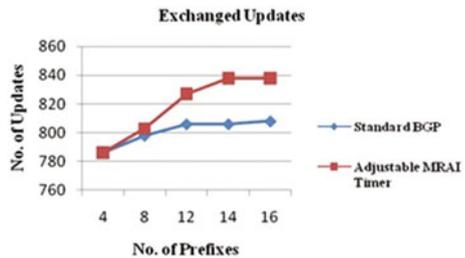


Fig. 4 Comparison of throughput on increment of number of prefixes

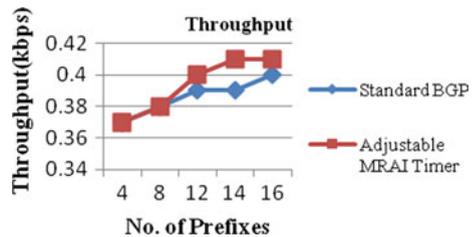
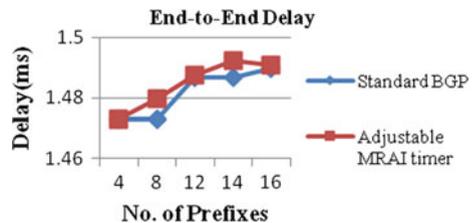


Fig. 5 Comparison of end-to-end delay with respect to increment of number of prefixes



7 Conclusion

MRAI timer has a noteworthy impact on the network performance. This paper improves the convergence in the internal BGP route reflectors using adjustable MRAI timers. The results signify the importance of MRAI timers in route reflectors. The performance attributes such as convergence time, number of the exchanged updates, throughput, and end-to-end delay has been evaluated and compared with standard iBGP. Results confirm that the adjustable MRAI timer gets better convergence in internal BGP route reflectors.

References

1. Rekhter, Y., Li, T.: Border Gateway Protocol 4, RFC 4271 (2006)
2. Chandra, R.: BGP Route Reflection: An alternative to full mesh IBGP, RFC 4456 (2006)
3. Rawat, A., Shayman, M.A.: Preventing persistent oscillations and loops in IBGP configuration with route reflection. *Comput. Netw.* **50**(18), 3642–3665 (2006)
4. Wang, F., Gao, L.: Impact of routing lags on internet routing failures. In *Proceedings of 4th International IFIP-TC6 Networking Conference*, pp. 1334–1338. Springer, Berlin, Heidelberg (2005)
5. Abuzneid, A., Stark, B.J.: Improving BGP convergence time via MRAI timer. In *Proceedings of the Novel Algorithms and Techniques in Telecommunications and Networking*, pp. 105–110. Springer, Netherlands, December 2009
6. Fabrikant, A., Syed, U., Rexford, J.: There’s something about MRAI: timing diversity can exponentially worsen BGP convergence. In *Proceedings of IEEE International Conference on Computer Communications INFOCOM*, pp. 2975–2983. IEEE (2011)
7. Griffin, T.G., Premore, B.J.: An experimental analysis of BGP convergence time. In *Proceedings of International Conference of Network Protocols*, pp. 53–61. IEEE (2001)
8. Laskovic, N., Trajkovic, L.: BGP with an adaptive minimal route advertisement interval. In *Proceedings of Performance Computing and Communications Conference (IPCCC)*, pp. 136–142. IEEE (2006)
9. Vissicchio, S., Cittadini, L., Battista, G.Di: On iBGP routing policies. *IEEE Trans. Netw.* **23**(1), 227–240 (2015)
10. Park, J.H., Oliveira, R., Amante, S., Mcpherson, D.: BGP route reflection revisited. *IEEE Commun. Mag.* **50**(7), 70–75 (2012)
11. Issariyakul, T., Hossain, E.: *Introduction to Network Simulator NS2* (2nd ed), pp. 1–18. Springer, US (2012).
12. Sahraei, R.: ns-BGP 2.0. <http://www2.ensc.sfu.ca/~ljljja/cnl/projects/BGP-ns-2.34/ns-2.34-BGP.html> (2009)
13. Pelsser, C., Uhlig, S., Takeda, T., Quoitin, B., Shiomoto, K.: Providing scalable NH-diverse iBGP route re-distribution to achieve sub-second switch-over time. *Comput. Netw.* **54**, 2492–2505 (2010)

Performance Analysis of Various Eigenvalue-Based Spectrum Sensing Algorithms for Different Types of Primary User Signals

Pankaj Verma and Brahmjit Singh

Abstract Spectrum sensing plays a very essential role in the implementation of cognitive radio networks. Eigenvalue-based spectrum sensing algorithms have been comprehensively discussed in the literature, for detection of primary user signal in the case of uncertain noise. For detection of signals, the test statistics of these algorithms depend on the eigenvalues of the covariance matrix of the received signal. Eigenvalues generally capture the correlation among the signal samples. In this context, we have examined the sensing performance of various eigenvalue-based spectrum sensing techniques for different types of primary user signals having different levels of correlation. In results, it has been noticed that the sensing performance of the algorithms relies on the type of primary user signal transmitted.

Keywords Cognitive radio · Spectrum sensing · Energy detection
Eigenvalue-based detection · Random matrix theory

1 Introduction

The accelerated growth in the wireless services and their applications has induced the requirement of more and more spectrum, which is a scarce resource. The prevailing spectrum allocation policy allots spectrum to some users (also known as licensed users) for exclusive access. However, it has been revealed in a survey conducted under the supervision of the Federal Communications Commission (FCC) [1] that the part of the spectrum, which is assigned to the licensed users, remains unexploited maximum of the time. This underutilization of the radio

P. Verma (✉) · B. Singh
Department of Electronics and Communication Engineering,
National Institute of Technology Kurukshetra, Kurukshetra, India
e-mail: pankaj@nitkkr.ac.in

B. Singh
e-mail: brahmjit@nitkkr.ac.in

spectrum motivated researchers to search for some technology that can handle this issue and use the radio spectrum effectively.

Cognitive radio (CR) is a recent archetype, which attempts to solve this issue of spectrum underutilization through opportunistic spectrum access [2, 3]. Under this technology, secondary users (SUs) or unlicensed users can transmit in a frequency band of licensed users (or primary users (PUs)) provided the band is not being used. CR has the ability to alter its transmission specifications like modulation scheme, transmit power, operating frequency, and other parameters, as per the environment in its vicinity. To enable this technology, spectrum sensing plays a very imperative role. SUs identify the vacant spectrum with the help of different spectrum sensing algorithms. If the PU is found absent, then the band can be used by SUs, otherwise sensing is done again to identify other vacant bands. There are numerous spectrum sensing techniques debated in the literature, each having its own advantages and disadvantages [4, 5]. Among all the sensing algorithms, energy detection (ED)-based spectrum sensing is used extensively because of its simple implementation and non-requirement of prior information about PU signal, but its performance deteriorates due to uncertainty in the noise power. To overcome this problem of noise uncertainty, eigenvalue-based spectrum sensing schemes are introduced in the literature. In eigenvalue-based sensing, eigenvalues of the covariance matrix of the received signal are used to make the test statistics. In [6], the authors proposed maximum eigenvalue (ME) as the test statistics for the detection of PUs. In [7, 8], the ratio of maximum to the minimum eigenvalue (MME) and the ratio of the observed energy of the received signal to the minimum eigenvalue (EME) of the covariance matrix of the received signal are used as the test statics. In [6–8], the performance of all the algorithms mentioned above is analyzed only for captured ATSC DTV and wireless microphone signals. Therefore, there is a need to analyze the performance for different kinds of PUs signals. In this paper, we have carried out the performance analysis of various spectrum sensing methods such as ED, ME, MME, and EME for the different types of signals transmitted by PUs like Additive White Gaussian Noise (AWGN), cosine, Frequency Modulated (FM), and Binary Phase Shift Keying (BPSK) signals.

The rest of this paper is arranged as follows: Sect. 2 describes the model of the system under consideration and different types of sensing algorithms. Simulation results are analyzed in Sect. 3 and concluding remarks are presented in Sect. 4.

2 System Model and Various Spectrum Sensing Algorithms

The detection of PU can be devised as a testing of binary hypotheses, i.e., H_0 and H_1 . The hypothesis H_0 indicates the absence of PU signal and hypothesis H_1 denotes the presence of PU signal. The received signal $y(m)$ under both hypotheses can be represented as

$$H_0 : y(m) = u(m) \quad (1)$$

$$H_1 : y(m) = x(m) + u(m), \quad (2)$$

where $u(m)$ is AWGN signal assumed to be independent and identically distributed (i.i.d.) random process having mean zero and variance σ_u^2 , and $x(m)$ is the signal transmitted by PU.

The two probabilities associated with the performance of any sensing scheme are detection probability (P_d) and false alarm probability (P_{fa}). P_d indicates the presence of PU when it is present and P_{fa} indicates the presence of PU when it is absent.

For L successive samples, the following vectors can be represented as [6]:

$$\bar{y}(m) = [y(m) \quad y(m-1) \dots y(m-L+1)]^T \quad (3)$$

$$\bar{x}(m) = [x(m) \quad x(m-1) \dots x(m-L+1)]^T \quad (4)$$

$$\bar{u}(m) = [u(m) \quad u(m-1) \dots u(m-L+1)]^T, \quad (5)$$

where L is the smoothing factor. Thereafter, the statistical covariance matrix can be defined as

$$R_y = E[\bar{y}(m)\bar{y}^T(m)] \quad (6)$$

$$R_x = E[\bar{x}(m)\bar{x}^T(m)] \quad (7)$$

$$R_u = E[\bar{u}(m)\bar{u}^T(m)]. \quad (8)$$

It is known that [6]

$$R_y = R_x + \sigma_u^2 I_L. \quad (9)$$

For hypothesis H_0 , $R_x = 0$ and for hypothesis H_1 , $R_x \neq 0$. Assuming E_{\max} and E_{\min} as the maximum and minimum eigenvalue of R_y , respectively, and ρ_{\max} as the maximum eigenvalue of covariance matrix R_x . It is known that $\rho_{\max} = 0$, if $R_x = 0$. A statistical covariance matrix can be determined using a limited number of samples. The sample autocorrelations of the received signal can be defined as

$$E(l) = \frac{1}{M} \sum_{p=0}^{M-1} y(p)y(p-l), \quad l = 0, 1, \dots, L-1 \quad (10)$$

where M is the total number of samples used for detection. The covariance matrix R_y is approximately represented by a sample covariance matrix as

$$\widehat{R}_y(M) = \begin{bmatrix} E(0) & E(1) & \dots & E(L-1) \\ E(1) & E(0) & \dots & E(L-2) \\ \vdots & \vdots & \vdots & \vdots \\ E(L-1) & E(L-2) & \dots & E(0) \end{bmatrix}. \tag{11}$$

When hypothesis H_0 is true, sample covariance matrix $R_y(M)$ turns to $R_u(M)$ which is approximately a Wishart random matrix. For large M , maximum and minimum eigenvalues of $R_u(M)$ converge to $\frac{\sigma_u^2}{M}(\sqrt{M} + \sqrt{L})^2$ and $\frac{\sigma_u^2}{M}(\sqrt{M} - \sqrt{L})^2$, respectively [6].

Let

$$B(M) = \frac{M}{\sigma_u^2} R_u(M) \tag{12}$$

and

$$\mu = (\sqrt{M-1} + \sqrt{L})^2 \quad \text{and} \quad \nu = \left(\sqrt{M-1} + \sqrt{L}\right) \left(\frac{1}{\sqrt{M-1}} + \frac{1}{\sqrt{L}}\right)^{\frac{1}{3}}. \tag{13}$$

Further, assuming $\lim_{M \rightarrow \infty} \frac{L}{M} = z(0 < z < 1)$, $\frac{A_{\max}(B(M)) - \mu}{\nu}$ converges (with probability one) to the order 1 Tracy-Widom (TW) distribution. The TW distribution function can be described as

$$F_1(x) = \exp\left(-\frac{1}{2} \int_x^\infty (q(t) + (t-x)q^2(t)) dt\right), \tag{14}$$

where $q(t)$ is the solution of the nonlinear Painlevé II differential equation

$$q''(t) = tq(t) + 2q^3(t). \tag{15}$$

Generally, it is very tedious to evaluate it. Some values of the F_1 are given in Table 1 [9].

2.1 Different Algorithms

Based upon the eigenvalues of the covariance matrix, several detection algorithms have been recommended in the literature, and some of them are listed below:

Table 1 Numerical values of TW distribution of order 1

x	-3.90	-2.78	-1.27	-0.59	0.45	0.98	2.02
$F_1(x)$	0.01	0.10	0.50	0.70	0.90	0.95	0.99

Algo 1: Maximum–minimum eigenvalue (MME)

Step 1: Estimate the sample covariance matrix of the received signal.
 Step 2: Calculate the minimum (E_{\min}) and maximum (E_{\max}) eigenvalue.
 Step 3: If $E_{\max}/E_{\min} > \gamma_1$, then the signal is considered as present, otherwise absent. Here, γ_1 is the threshold for MME algorithm given as [7]

$$\gamma_1 = \frac{(\sqrt{M} + \sqrt{L})^2}{(\sqrt{M} - \sqrt{L})^2} \left(1 + \frac{(\sqrt{M} + \sqrt{L})^{-2/3}}{(ML)^{1/6}} F_1^{-1}(1 - P_{\text{faMME}}) \right). \quad (16)$$

Algo 2: Energy detection (ED)

Step 1: Calculate the energy of the received signal as

$$T(M) = \frac{1}{M} \sum_{m=0}^{M-1} |y(m)|^2. \quad (17)$$

Step 2: Decision: if $T(M) > \gamma_2$, then PU is considered as present, otherwise absent. γ_2 is the threshold used for ED method [10]:

$$\gamma_2 = \sigma_u^2 \left(1 + \frac{Q^{-1}(P_{\text{faED}})}{\sqrt{M/2}} \right). \quad (18)$$

Algo 3: Maximum eigenvalue (ME)

Step 1: Similar to Algo 1.
 Step 2: Calculate the maximum (E_{\max}) eigenvalue of the sample covariance matrix.
 Step 3: If $E_{\max} > \gamma_3 \sigma_u^2$, then signal is present, otherwise absent. Here, γ_3 is the threshold corresponding to the ME algorithm [6]

$$\gamma_3 = \frac{(\sqrt{M} + \sqrt{L})^2}{M} \left(1 + \frac{(\sqrt{M} + \sqrt{L})^{-2/3}}{(ML)^{1/6}} F_1^{-1}(1 - P_{\text{faME}}) \right). \quad (19)$$

Algo 4: Energy to minimum eigenvalue (EME)

Step 1: Similar to Algo 1.
 Step 2: Obtain the energy of the received signal and minimum eigenvalue (E_{\min}) of the sample covariance matrix.

Step 3: If $T(M)/E_{\min} > \gamma_4$, then PU is considered as present, otherwise absent. γ_4 is the threshold used in EME method [7]

$$\gamma_4 = \left(1 + \frac{Q^{-1}(P_{fa_{EME}})}{\sqrt{M/2}} \right) \frac{M}{(\sqrt{M} - \sqrt{L})^2}. \quad (20)$$

2.2 Different Types of PU Signals

The performance of the various methods mentioned above is analyzed for various types of signals transmitted by PUs such as BPSK signal, FM signal, cosine signal, and AWGN signal.

The parameters used for the generation of the signal are listed as follows:

- Carrier frequency $f_c = 100$ kHz
- Sampling frequency $f_s = 10f_c$
- Frequency deviation $\Delta f = 25$ kHz.

3 Simulation Results and Analysis

In this section, we have analyzed the performance of various spectrum sensing algorithms. For smoothing factor, $L = 8$, number of samples $N = 100,000$ and fixed false alarm probability, $P_{fa} = 0.1$, 10,000 Monte Carlo simulations are performed to assess the performance. Figure 1 shows the detection probability evaluated at distinct values of signal-to-noise ratio (SNR) for MME method for different types of signals transmitted by PU. It has been observed that this method works well for BPSK and cosine signal, but does not detect the AWGN signal. Figure 2 shows the same plot, but for ED method. It can be observed that this method of detection gives the same performance irrespective of the type of signal transmitted by PU. Similarly, Figs. 3 and 4 show the same results, but for ME and EME method, respectively. Overall, this can be concluded that, except ME and ED method, non-other works when the PU signal is considered as random in nature (i.e., AWGN).

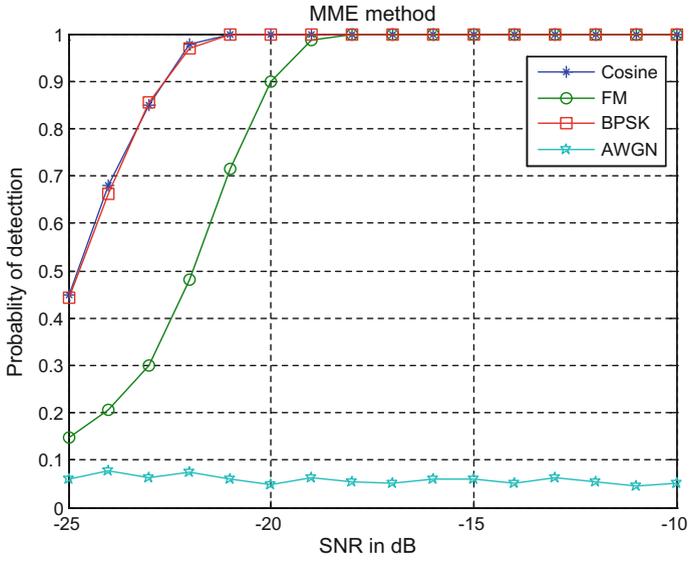


Fig. 1 Performance analysis of MME method

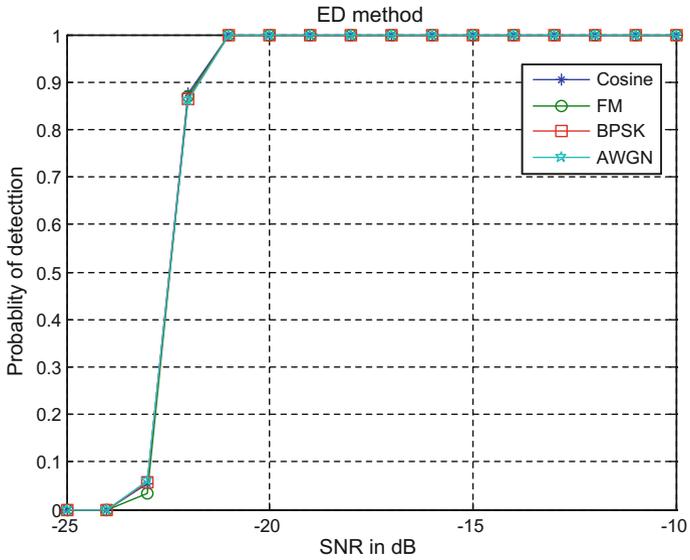


Fig. 2 Performance analysis of ED method

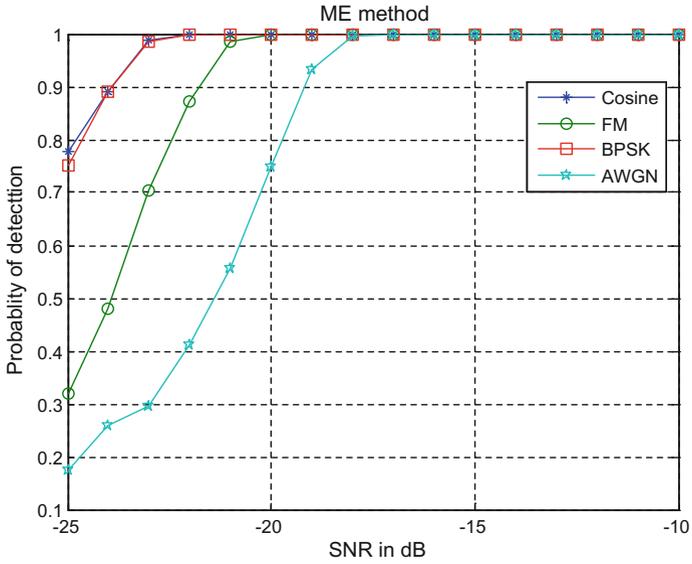


Fig. 3 Performance analysis of ME method

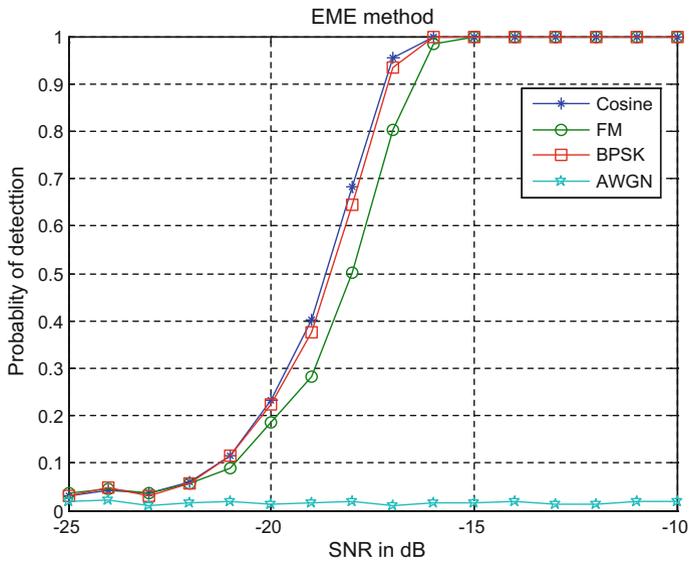


Fig. 4 Performance analysis of EME method

4 Conclusion

In this study, we have carried out the performance analysis of ME, MME, EME, and ED methods for different types of PU signals, viz., Cosine, BPSK, FM, and AWGN. By simulations, it has been observed that MME and EME method fails to give satisfactory performance when the signal is assumed to be AWGN, and on the other side, the performance of ED method is independent of the type of PU signal. Overall, it can be concluded that the type of PU signal should be considered while selecting an eigenvalue-based spectrum sensing algorithm.

References

1. Federal Communications Commission: Spectrum policy task force report. FCC, pp. 02–155 (2002)
2. Mitola, J., Maguire, G.Q.: Cognitive radio: making software radios more personal. *IEEE Pers. Commun.* **6**(4), 13–18 (1999)
3. Haykin, S.: Cognitive radio: brain-empowered wireless communication. *IEEE J. Sel. Areas Commun.* **23**(2), 201–220 (2005)
4. Ariananda, D.D., Lakshmanan, M.K., Nikookar, H.: Survey on spectrum sensing techniques for cognitive radio. In: 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management (2009)
5. Ghasemi, A., Sousa, E.S.: Spectrum sensing in Cognitive radio networks: requirements, challenges and design trade-offs cognitive radio communication and networks. *IEEE Commun. Mag.* 32–39 (2008)
6. Zeng, Y., Koh, C.L., Liang, Y.C.: Maximum eigenvalue detection: theory and application. In: *Proceeding of IEEE International Conference*, pp. 4160–4164 (2008)
7. Zeng, Y., Liang, Y.: Maximum-minimum eigenvalue detection for cognitive radio. In: *Proceeding IEEE 18th International Symposium on Personal, Indoor, Mobile Radio Communication (PIMRC)*, pp. 1–5 (2007)
8. Zeng, Y., Liang, Y.C.: Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Trans. Commun.* **57**, 1784–1793 (2009)
9. Tracy, C.A. Widom, H.: The distribution of the largest eigenvalue in the Gaussian ensembles. In: *Calogero-Moser-Sutherland Models*, pp. 461–472. Springer, New York (2000)
10. Verma, P., Singh, B.: On the decision fusion for cooperative spectrum sensing in cognitive radio networks. *Wirel. Netw.* 1–10 (2016)

A Crown-Shaped Microstrip Patch Antenna for Wireless Communication Systems

Atanu Nag, Kousik Roy and Debika Chaudhuri

Abstract This paper describes the design and analysis of a modified E-shaped and a crown-shaped microstrip patch antenna with excitation frequency of 1.2 GHz. Both the antennas are derived from a Rogers RT/Duroid 6202 laminate substrate (dielectric constant = 2.94 and loss tangent = 0.001). The simulation is achieved using the software Ansoft Nexxim. From the simulation results it is observed that the crown-shaped antenna exhibits 20.15 dB better return loss and 50% more gain than the modified E-shaped antenna. Furthermore, better impedance matching is observed for the crown-shaped antenna exhibiting that the crown-shaped antenna finds better applications than the modified E-shaped antenna to be effectively used for wireless application.

Keywords Microstrip antenna · Wireless communication · Return loss Gain · Ansoft

1 Introduction

This paper elucidates the design, simulation, and analysis of a modified E-shaped and a crown-shaped microstrip patch antenna [1–5]. The excitation frequency for both the antennas has been chosen as 1.2 GHz. The antennas are derived from a laminate substrate having a loss tangent of 0.001. Both the antennas are made compact and thin with the substrate material having low value of dielectric constant

A. Nag (✉) · D. Chaudhuri
Modern Institute of Engineering & Technology, Bandel, Hooghly 712123, WB, India
e-mail: tnnag79@gmail.com

D. Chaudhuri
e-mail: debika.chaudhuri@gmail.com

K. Roy
Asansol Engineering College, Asansol 713305, WB, India
e-mail: kousikroy002@gmail.com

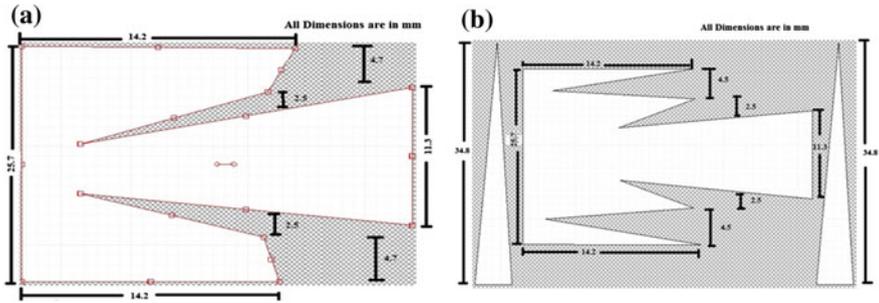


Fig. 1 Geometry of **a** modified E-shaped antenna and **b** crown-shaped antenna

(2.94). Simulation is consummate utilizing the software Ansoft and Nexxim [6] which helps providing good insight on the effects of various parameters.

2 Antenna Configuration

The geometries of the modified E-shaped antenna [7–9] and the crown-shaped antenna are shown in Fig. 1a, b, respectively. The excitation frequency for both antennas has been chosen as 1.2 GHz. The antennas are derived from a Rogers RT/Duroid 6202 laminate substrate (loss tangent = 0.001), having dielectric constant 2.94. The simulation is completed utilizing the software Ansoft Nexxim.

3 Simulated Results for Modified E-Shaped and Crown-Shaped Antenna

3.1 $S(1, 1)$ and Return Loss

$S(1,1)$ signifies the total amount of power reflected from the antenna and hence it is also known as the reflection coefficient.

From Fig. 2a it implies that the modified E-shaped antenna radiates best at 4.8 GHz, where $S(1, 1) = -22.5$ dB and the observed bandwidth for the modified E-shaped antenna is found to be 830 MHz, whereas from Fig. 2b it implies that the crown-shaped antenna radiates best at 5.2 GHz, where $S(1, 1) = -43$ dB and the observed bandwidth for the crown-shaped antenna is found to be 2.4 GHz, viewing that the crown-shaped antenna exhibits 20.15 dB better return loss as compared to modified E-shaped antenna design.

3.2 Radiation Pattern and Smith Chart

Figure 3a, b, respectively, shows the radiation pattern of the modified E-shaped antenna and the crown-shaped antenna indicating that both the antenna are highly

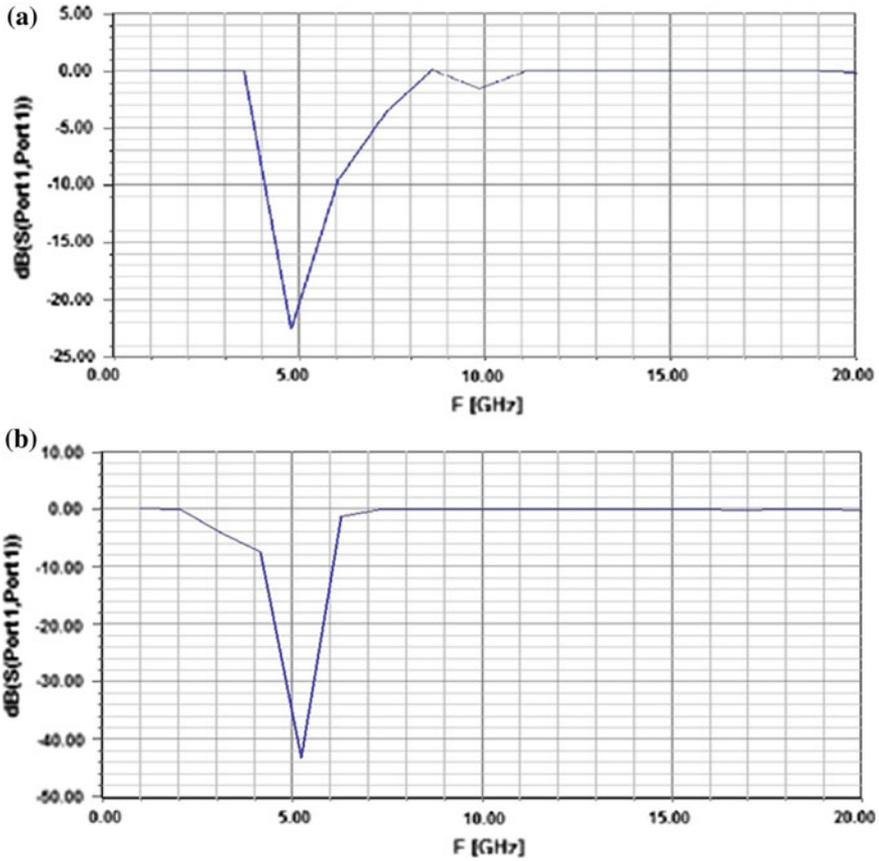


Fig. 2 Return loss of **a** modified E-shaped antenna and **b** crown-shaped antenna

directional, whereas Fig. 4a, b, respectively, shows the smith chart of the modified E-shaped antenna and the crown-shaped antenna.

3.3 Swept Gain and 3-D Input Gain

Figure 5a, b, respectively, shows the swept gain of the modified E-shaped antenna and the crown-shaped antenna, whereas Fig. 6a, b, respectively, shows the 3-D input gain of the modified E-shaped antenna and the crown-shaped antenna. The gains of the modified E-shaped antenna and the crown-shaped antenna are, respectively, 10 and 15 dBi approximately indicating that the crown-shaped antenna exhibits 50% more gain compared to modified E-shaped antenna design.

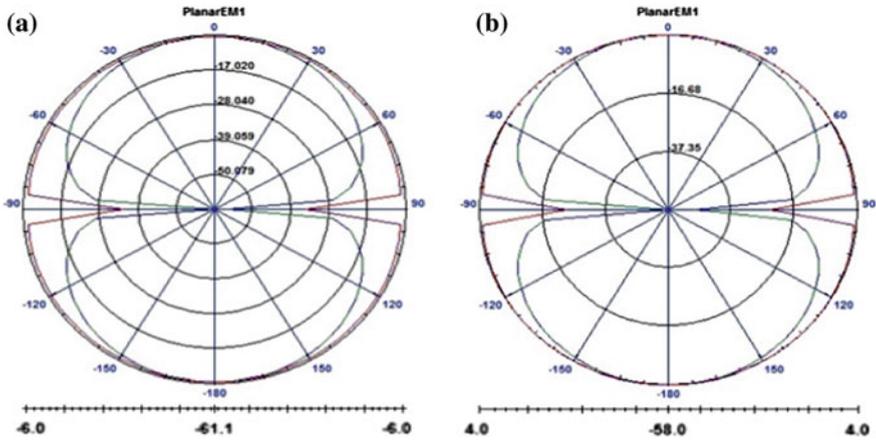


Fig. 3 Radiation pattern of a modified E-shaped antenna and b crown-shaped antenna

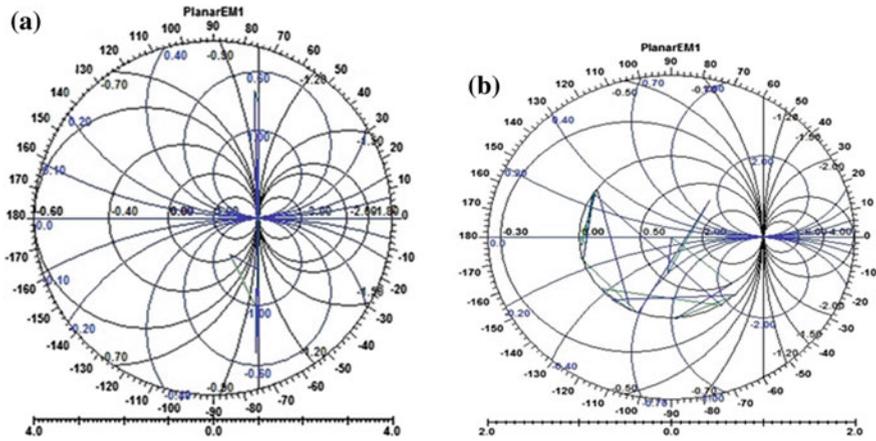


Fig. 4 Smith chart of a modified E-shaped antenna and b crown-shaped antenna

3.4 Far-Field Region

Figure 7a, b, respectively, shows the far-field region of the modified E-shaped antenna and the crown-shaped antenna.

A comparison, on the basis of simulation results, for understanding the parameters of both the antennas is shown in Table 1.

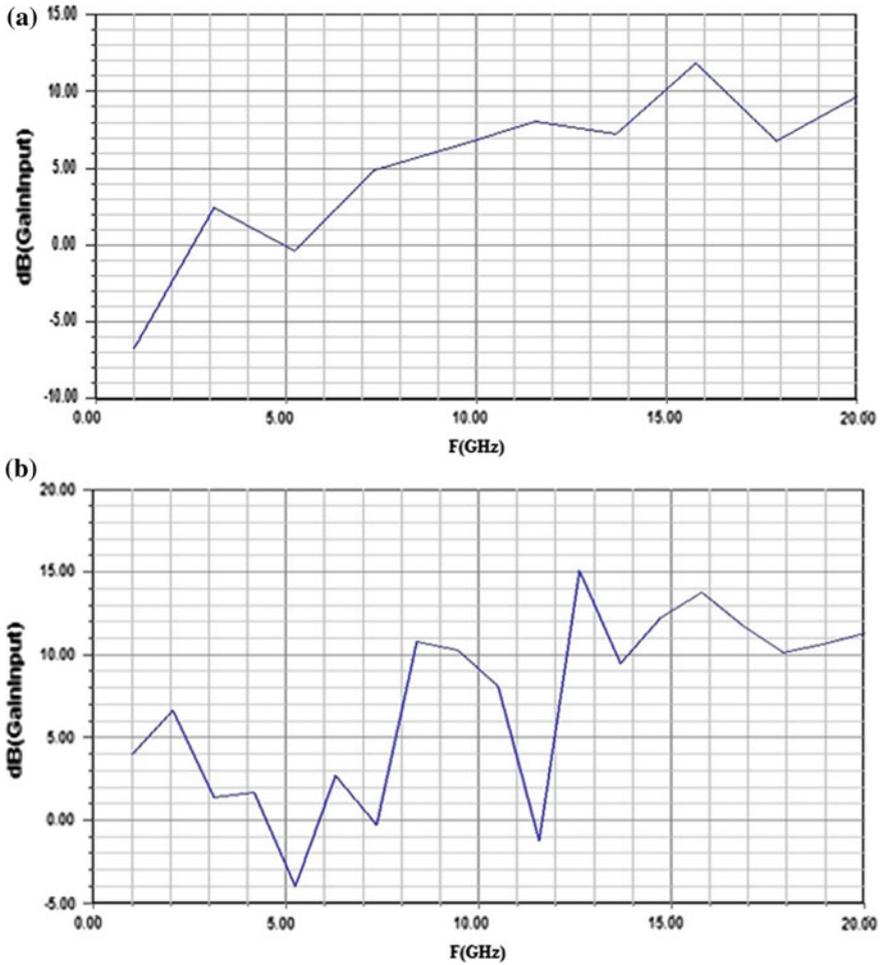


Fig. 5 Swept gain of **a** modified E-shaped antenna and **b** crown-shaped antenna

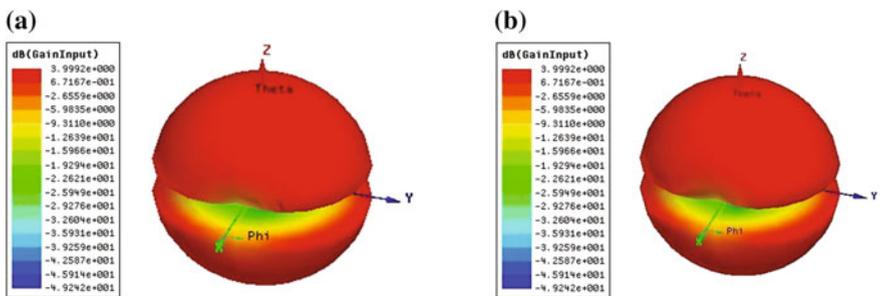


Fig. 6 3-D input gain of **a** modified E-shaped antenna and **b** crown-shaped antenna

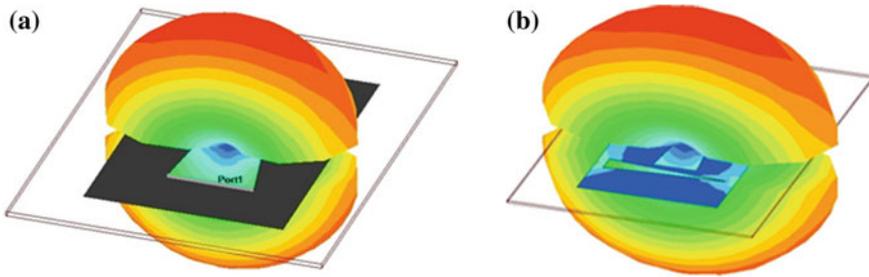


Fig. 7 Far-field region of **a** modified E-shaped antenna and **b** crown-shaped antenna

Table 1 Comparison of modified E-shaped and crown-shaped microstrip patch antenna

Antenna type/parameters	Modified E-shaped microstrip patch antenna	Crown-shaped microstrip patch antenna
Radiation pattern	Highly directional	Highly directional
Gain	High (10 dBi)	Higher (approx 15 dBi)
Bandwidth	830 MHz	2.4 GHz
Input impedance	52.77 + j92.48	52.18 + j2.58
Return loss	-22.5 dB	-43 dB
Applications	<ol style="list-style-type: none"> 1. Wireless video links 2. Wireless security cameras 3. Wireless video transmitters 4. Wireless security systems 	<ol style="list-style-type: none"> 1. Wireless video links 2. Wireless security cameras 3. Wireless video transmitters 4. Wireless security systems

4 Conclusion

From the simulation results of our designed modified E-shaped microstrip patch antenna and the crown-shaped microstrip patch antenna it is obvious that the reflection coefficient for the modified E-shaped antenna is -22.5 dB at 4.8 GHz and for the crown-shaped antenna is -43 dB at 5.2 GHz, showing that the crown-shaped antenna exhibits 20.15 dB better return loss and 50% more gain compared to modified E-shaped antenna design. Better impedance matching is also achieved in the crown-shaped antenna than the modified E-shaped antenna exhibiting that the crown-shaped antenna finds better applications than the modified E-shaped antenna to be effectively used for high-speed applications like video links, video transmitters, and security purpose in wireless communication systems.

References

1. Kraus, J.D.: *Antennas for All Applications*, 2nd edn. McGraw Hill Inc., New York (1997)
2. Balanis, C.A.: *Antenna Theory, Analysis and Design*, 2nd edn. Wiley, New York (2005)
3. Yang, F., Zhang, X.X., Ye X., Rahmat-Samii, Y.: Wide-band E shaped patch antennas for wireless communications. *IEEE Trans. Antennas Propag.* **49** (2001)
4. Pandey, V.K., Vishvakarma, B.R.: Analysis of an E-shaped patch antenna. *Microwave Optical Technol. Lett.* **49** (2007)
5. Kim, B.: Novel single-feed circular microstrip antenna with reconfigurable polarization capability. *IEEE Trans. Antenna Propag.* **56** (2008)
6. Ansoft Designer and Nexxim. <http://ansoft-designer-and-nexxim.software.informer.com/>
7. Pal, S., Roy, K., Nag, A., Tiwary, A.K.: A Typical compressed wide band and high gain microstrip patch antenna for GSM applications. *Int. J. Innov. Res. Sci. Eng. Technol.* **3**, 37–40 (2014). (Spl 6)
8. Roy, K., Nag, A., Pal, S., Tiwari, A., Sinha, A.: Analysis of E-shapped geometry microstrip antenna. *Int. J. Innov. Res. Sci. Eng. Technol.* **4**, 170–173 (2015). (Spl 9)
9. Roy, K., Chaudhuri, D., Bose, S., Nag, A: A novel dual band antenna for radar application. In: *Proceedings of 3rd International Conference on Advanced Computing Networking and Informatics*, vol. 2, Springer, India, vol. 44, pp. 643–650 (2015). doi:[10.1007/978-81-322-2529-4](https://doi.org/10.1007/978-81-322-2529-4). ISBN: 978-81-322-2528-7

Atrophy Measure of Brain Cortex to Detect Alzheimer's Disease from Magnetic Resonance Images

Dulumani Das and Sanjib Kumar Kalita

Abstract In medical science, diagnosis of Alzheimer's disease is mainly done manually by expert radiologist. In this paper, an automatic approach to detect Alzheimer's disease using cortex thickness is analyzed. The cortex of brain is extracted and thickness is measured from magnetic resonance images. For experiment, 20 images of control subjects, 20 with mild cognitive impairment and 20 with Alzheimer's disease, are taken from Alzheimer's Disease Neuroimaging Initiative database. Initially, segmentation of cortex, from T1-weighted coronal magnetic resonance image, is done using genetic algorithm-based region growing technique. Then the thickness of the cortex is measured using distance transform. The experiment gives 100, 80, and 85% recognition accuracy for normal, mild cognitive impairment, and Alzheimer's disease, respectively.

Keywords Segmentation · MRI · Alzheimer's disease · Mild cognitive impairment · Cortex · Medical image processing

1 Introduction

Alzheimer's disease is a progressive degenerative disease, characterized by atrophy of different parts of brain. It is a challenging task in medical science to detect Alzheimer's disease (AD) at an early stage. Mild cognitive impairment (MCI) is considered as an early stage of AD. Human brain is a complex nonlinear structure. It requires high skill in radiology to understand structural change and location of

First author receives Inspire Fellowship from Department of Science and Technology, New Delhi.

D. Das (✉) · S.K. Kalita
Department of Computer Science, Gauhati University, Guwahati, Assam, India
e-mail: dulumonids1@gmail.com

S.K. Kalita
e-mail: sanjib959@rediffmail.com

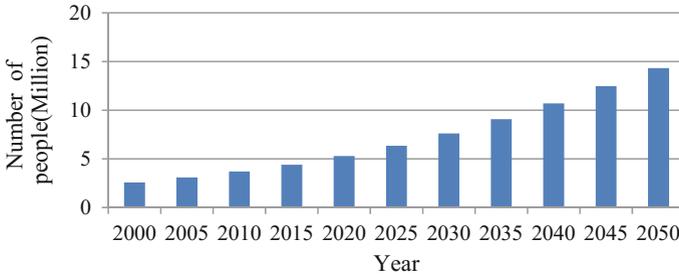


Fig. 1 Estimation of number of peoples with Alzheimer's disease in India

disorder for proper diagnosis. This makes it very important, to detect any structural change or disorder of brain automatically. The automatic detection of AD requires true segmentation of different structures of brain. A survey of the disease reveals that gray matter loss of cortex is an early sign of AD [1–3]. In this paper total gray matter loss of cortex for whole brain is measured to detect AD and MCI. This work is experimented, using structural MRI. MRI is a noninvasive imaging technique of medical science. In MRI machine, images can be taken in multiple planes, i.e., axial, coronal, and sagittal plane without changing the position of the patient inside the machine. High soft tissue differentiation is an advantage of MRI over other medical imaging techniques and it also provides no harmful effects on human body. Commonly, MRI scans can be T1-weighted and T2-weighted. In this work, coronal T1-weighted MRI is considered for experiment.

1.1 Alzheimer's Disease and Motivation

Alzheimer's disease is the main cause of dementia in old age [4]. It is commonly seen after 60/65 years of age. AD is a neurodegenerative disease; it starts slowly but gradually affects the whole brain. This causes loss of brain cells and tissues. There is no cure for AD, but early detection can help to slow down the progression of the disease and increase the lifespan [4]. There is an estimation of number of people with Alzheimer disease in India between 2000 and 2050 published by Alzheimer's and Related Disorders Society of India 2010 [4] as in Fig. 1. It is clear from Fig. 1 that the number of AD affected people is increasing rapidly. However, literature review reveals that there is a lack of proper model to automatically detect AD at an early stage. This motivated us to work in the field of AD and provide some solution using image processing techniques.

2 Theoretical Background

The algorithm that is used in this work for segmenting the cortex from brain is genetic algorithm. The main objective of genetic algorithm is to find the best solution from all available possible solutions for a problem. Chromosome

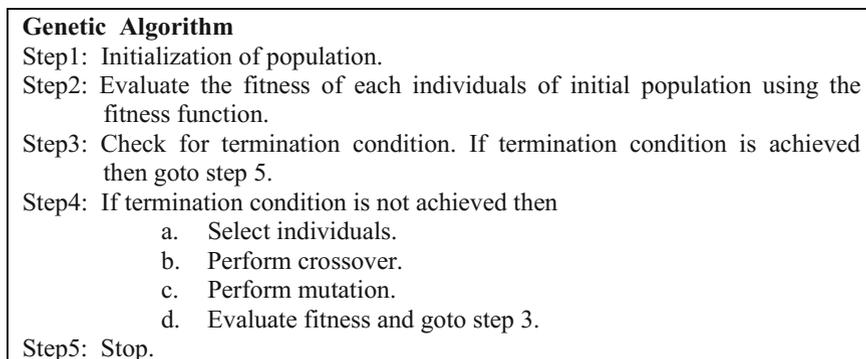


Fig. 2 Standard genetic algorithm

representation of each possible solution and a fitness function are the two basic requirement of genetic algorithm. The standard genetic algorithm [5] is shown in Fig. 2.

3 Related Works

The gray matter volume change of brain is an absolute neuroimaging marker for AD detection. Region-based analysis of gray matter volume change gives high accuracy than the analysis of brain as a whole. The volume change of hippocampus and other cortical regions are useful marker for AD detection. Researchers have been working on AD detection from last few decades using different approaches such as deformation-based approach, atlas-based approach, data mining approach, etc. Table 1 shows some of the related works on AD.

4 Proposed Method

In this paper, an automatic method to recognize MCI, AD, and normal subject is proposed. The main objective of our work is to segment out the cortex of brain from coronal T1-weighted MRI and then find the thickness of cortex, as the loss of gray matter in cortex is a predictive marker for AD and MCI. Segmentation of cortex is done using genetic algorithm-based region growing technique. Initially, the image pixels are grouped into some clusters by randomly selecting some points from the image as cluster center, and using euclidean distance. Each cluster represents a chromosome. Genes of the chromosome represents the cluster number. Variable length chromosome representation is used in the proposed genetic algorithm-based method. The population size is considered as 25. Initialization of population is done

Table 1 Related works on AD

Author	Method	Feature	Accuracy
Plant et al. [6]	Support vector machine (SVM), bayes statistics, and voting feature intervals (VFI)	Entropy of voxel	75% for the prediction of MCI to AD conversion
Querbes et al. [7]	Coregistration and Partitioning	Cortical thickness	85% for discrimination between AD and normal and 76% in prediction of MCI to AD conversion
Yang et al. [8]	Support vector machine	ICA (independent component analysis) based	97.7% for discrimination between AD and normal and 87.8% between MCI and normal
Cho et al. [9]	Individual subject classification method based on incremental learning	Cortical thickness	Discrimination of AD from normal with 82% sensitivity and 93% specificity
Lerch et al. [10]	Neural net classifier, deformation algorithm, and discriminant analysis	Cortical thickness	90% accuracy obtained in the medial temporal lobes and other limbic structures and 100% with multivariate discriminant analysis
Hamalainen et al. [11]	Voxel-based morphometry (VBM)	Gray matter atrophy	–
Desikan et al. [12]	Division of cerebral cortex into gyral-based regions of interest using intraclass correlation coefficients (ICC), and mean distance maps (atlas-based method)	Anatomic curvature information of cerebral cortex	Average ICC of 0.835 across all of the ROIs, and a mean distance error of less than 1 mm

by randomly selecting 25 chromosomes. The steps of the proposed method are outlined in Fig. 3. The fitness function defined here is based on region homogeneity and intensity value. The fitness function f for a chromosome x is defined as in (1)

$$f(x) = n(x) + m(x), \quad (1)$$

where $n(x)$ = Number of similar neighboring pixels and $m(x)$ = Number of pixels having intensity value greater than a threshold value.

For counting number of similar neighboring pixels, four-connected neighborhood pixels are considered. The threshold value is obtained using standard deviation and average intensity. Selection of chromosome from initial population is done using roulette wheel selection algorithm. In the present work one-point adaptive crossover is performed. After selection operation the average fitness of selected chromosomes is calculated by dividing the total fitness of all selected chromosomes by total number of selected chromosomes. Then crossover probability of each chromosome is calculated as defined in (2) using the concept used in [13]:

- Step1: Read the MRI.
 Step2: Remove the skull and unwanted portions of brain using morphological operation (erosion and dilation).
 Step3: Perform Contrast enhancement in spatial domain.
 Step4: Reduce noise from MRI.
 Step5: Find seed points in cortex using genetic algorithm.
 Step6: Extract cortex by region growing technique using the seed point obtained in step 5.
 Step7: Find the thickness of cortex.
 Step8: Recognition of AD, MCI and normal MRI using thickness value.

Fig. 3 Proposed method for distinction of AD, MCI, and normal MRI

$$pc = \begin{cases} \frac{k_1(f)}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}}, \\ k_2, & f < f_{\text{avg}} \end{cases}, \quad (2)$$

where f = fitness of chromosome, f_{\max} = maximum fitness obtained from selected chromosomes, f_{avg} = average fitness, $k_1 = 0.9$, and $k_2 = 0.3$.

After calculating probability, a random number r between 0 and 1 is generated and the chromosome for which crossover probability is greater than r is selected for crossover. Similar to crossover, mutation operation performed here is also adaptive. The mutation probability of each chromosome is calculated as defined in (3) using the concept used in [13]:

$$pm = \begin{cases} \frac{k_1(f)}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}}, \\ k_2, & f < f_{\text{avg}} \end{cases}, \quad (3)$$

where f = fitness of chromosome, f_{\max} = maximum fitness obtained from chromosomes after crossover, f_{avg} = average fitness after crossover, $k_1 = 0.6$, and $k_2 = 0.3$.

After calculating probability, a random number r between 0 and 1 is generated and the chromosome for which mutation probability is greater than r is selected for mutation. The chromosomes obtained after mutation forms the new population. The termination condition considered for the algorithm is the number of iteration. After termination of genetic algorithm we get some chromosomes and the chromosome with highest fitness gives the seed points of cortex. Then region growing operation is performed with the obtained seed points to extract the cortex region. Thickness of extracted cortex is measured using distance transform to discriminate AD, MCI, and normal MRI.

5 Result and Discussion

The experiment is performed on 20 control subjects, 20 MCI, and 20 AD MRI of 65–85 age group of ADNI (Alzheimer’s disease neuroimaging initiative) screening database (www.loni.ucla.edu/ADNI). T1-weighted MRI with coronal view is observed. The preprocessing of images includes removal of unwanted portions such as brain skull and brain stem, noise reduction, and contrast enhancement. Figure 4

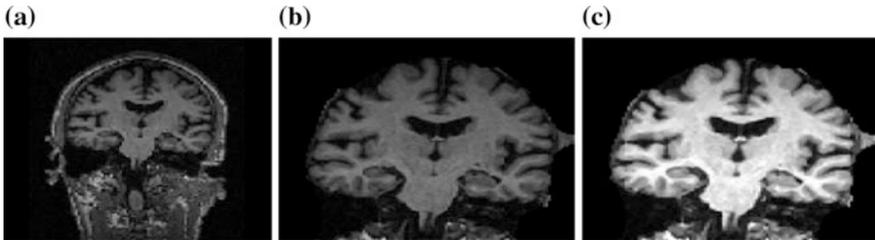


Fig. 4 a Original normal image, b after removal of unwanted portions, c after noise reduction and contrast enhancement

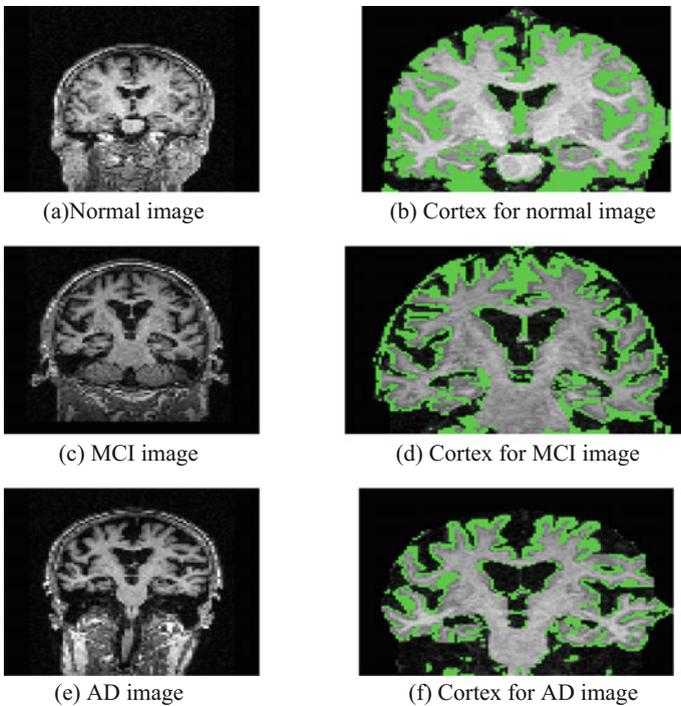


Fig. 5 b, d, f are images after extraction of cortex from a, c, and e, respectively

Table 2 Recognition of images

Normal	MCI	AD
100%	80%	85%

shows an image after preprocessing step. After preprocessing step, the cortex from the preprocessed image is extracted using the proposed region growing technique. Figure 5 shows the extracted cortex of brain with green color. From Fig. 5b, d, f it is observed that the green portion of images that indicates cortex is gradually reduced in MCI and AD images, and the amount of reduction is more in AD images than MCI images.

The maximum thickness found is 25, 18, and 15 for normal, MCI, and AD MRI, respectively. Table 2 shows the recognition of normal, MCI, and AD images, respectively.

6 Conclusion and Future Work

In this paper an automatic approach for detection of Alzheimer's disease (AD) and mild cognitive impairment (MCI) based on cortex thickness is performed. Thickness of brain cortex is measured globally in this work, but cortex atrophy that happens due to Alzheimer's disease is different in different cortical and subcortical structures of brain. So we can also measure structure-wise thickness to detect AD and MCI at an early stage. Due to different sizes and shapes of brain and intensity inhomogeneity of MRI, the preprocessing steps should be most robust to extract the cortex accurately.

References

1. Frisoni, G.B., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., Laakso, M. P.: Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. *J. Neurol. Neurosurg. Psychiatry* **73**, 657–664 (2002)
2. Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W.: Cortical change in Alzheimer's disease detected with a disease specific population based brain atlas. *Cereb. Cortex* **11**(1), 1–16 (2001)
3. Thompson, P.M., Hayashi, K.M., Zubicaray, G.D., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittner, S.S., Doddrell, D.M., Toga, A.W.: Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* **23**(3), 994–1005 (2003)
4. Shaji, K.S., Jotheeswaran, A.T., Girish, N., Bharath, S., Amit, D., Pattabiraman, M., Varghese, M.: The dementia India report. [online] http://www.alzheimer.org.in/dementia_2010.pdf (2010)
5. Pujari, K.A.: Data Mining Techniques. Universities Press

6. Plant, C., Teipel, S.J., Oswald, A., Bohm, C., Meindl, T., Miranda, M.J., Bokde, A.W., Hampel, H., Ewers, M.: Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage* **50**, 162–174 (2009)
7. Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Demonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P.: Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. *Brain* **132**, 2036–2047 (2009)
8. Yang, W., Li, Y., Chen, X.: ICA image feature extraction for improving diagnosis of Alzheimer's disease and mild cognitive impairment. 10th International Conference on Natural Computation, IEEE, Xiamen, pp. 802–806 (2014)
9. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y.: Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. *NeuroImage* **59**(3), 2217–2230 (2012)
10. Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A. C.: Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol. Aging* **29**(1), 23–30 (2008)
11. Hamalainen, A., Tervo, S., Olivares, M.G., Niskanen, E., Pennanen, C., Huuskonen, J., Kivipelto, M., Hanninen, T., Tapiola, M., Vanhanen, M., Hallikainen, M., Helkala, E.L., Nissinen, A., Vanninen, R., Soininen, H.: Voxel-based morphometry to detect brain atrophy in progressive mild cognitive impairment. *NeuroImage* **37**(4), 1122–1131 (2007)
12. Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**(3), 968–980 (2006)
13. Li, A.: The operator of genetic algorithms to improve its properties. *Mod. Appl. Sci.* **4**(3), 60–62 (2010)

Study of Effect of High-k Dielectric Gate Oxide on the Performance of SB-GNRFETs

Bhubon Chandra Mech and Jitendra Kumar

Abstract A Schottky barrier GNRFET with asymmetric and symmetric dielectric layer with high-k oxides is presented. The device considered is a double-gate MOSFET structure. We have studied the effect of various high-k gate dielectrics on the performance of graphene nanoribbon FETs (GNRFET). In addition to that, we have studied the effect of using two different dielectrics on top and bottom gate oxide. The study is carried out within the non-equilibrium Greens function formalism (NEGF). Results show that the use of high-k gate dielectric improves GNRFET characteristics like on-current, on-to-off current ratio. We obtained the value of 2.2 μA and 22 for drain current and on-to-off current ratio, respectively, for SiO_2 , whereas for La_2O_3 we obtained the value of 6.33 μA and 63.3, respectively for the same.

Keywords NEGF · Graphene · Graphene nanoribbon FET · $I_{\text{on}}/I_{\text{off}}$
High-k dielectrics

1 Introduction

Today, the CMOS technology is fascinated by the introduction of 2D material. With the introduction and the development of technique to extract the 2D materials like graphene, the IC industry has been able to cope up with the Moore's law. Graphene is expected to be the future material with potential to replace the silicon in IC industry because of its light weight, higher carrier mobility, and most significantly because of its ability to exist as a single atomic layer. With the scaling of channel using 2D material, the gate dielectric scaling is also needed. Even though the 2D graphene has stretched the limits of scaling to a far extent, the limitation in

B.C. Mech (✉) · J. Kumar

Indian Institute of Technology (ISM), Dhanbad 826004, Jharkhand, India
e-mail: bhubon1126@gmail.com

J. Kumar

e-mail: jitenkg@rediffmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_44

scaling of conventional SiO_2 gate dielectric layer is a major setback. As the dielectric layer is scaled down, the gate leakage current increases, and oxide breakdown occurs. As an alternative to SiO_2 , various high-k dielectrics are explored as an option so that low gate leakage current and high on-to-off current ratio are obtained [1, 2].

This paper deals with the study of a ballistic double-gate Schottky barrier GNRFET with different gate dielectrics. The mathematical tool used is the NEGF formalism which is used in conjunction with self-consistent solution of Schrödinger–Poisson equation [3]. All the results are obtained using NanoTCAD_ViDES which is based on NEGF formalism [4].

The GNR-based FET considered in this paper is described in Sect. 2. In the following sections simulation approach and the results are discussed. The paper is concluded in Sect. 5 which is the summary of results obtained.

2 Structure Model

The SB-GNRFET discussed in this paper is as shown in Fig. 1. The GNR considered is armchair GNR with chirality of (6, 0) which is shown in Fig. 2. The other related dimensions are as shown in Table 1. The GNR of chirality (6, 0) has a band gap of 0.602 eV and width of 1.37 nm [5, 6].

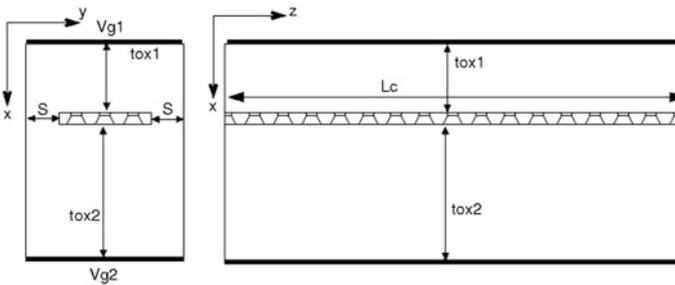


Fig. 1 The Schottky barrier GNRFET structure used in simulation

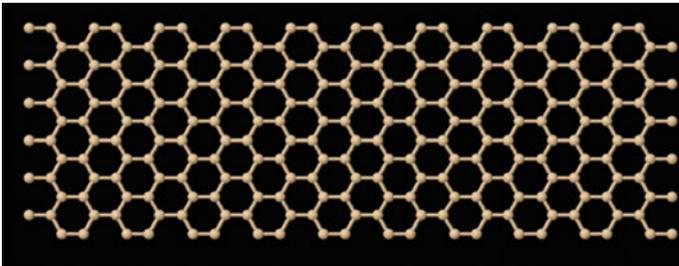


Fig. 2 The (6, 0) armchair GNR used in the GNRFET

Table 1 Dimensions of the device used in the simulation

Sl. No.	Symbol	Parameters	Value (nm)
1	$t_{\text{ox}2}$	Bottom oxide thickness	2
2	$t_{\text{ox}1}$	Top oxide thickness	2
3	L_c	Channel length	15
4	S	Lateral spacing	1

Two different types of devices are considered: (i) symmetric gate oxide GNRFET (SG-GNRFET) with same gate oxide in top and bottom oxide layer (ii) asymmetric gate oxide with one dielectric as top oxide and a different dielectric as bottom oxide (AG-GNRFET).

3 Simulation Approach

NEGF formalism is used to calculate the current–voltage characteristics. The Hamiltonian for the channel part is obtained using tight-binding approach. Real space approach has been implemented for the GNR with open boundary conditions [4, 7].

The most important parameter in NEGF is the Green's function which is given by

$$G(E) = [EI - H - \Sigma_S - \Sigma_D]^{-1} \quad (1)$$

where H is the tight-binding Hamiltonian of GNR and Σ_S and Σ_D are the source and drain self-energy matrices, respectively. In order to determine the electron and hole concentration, the Greens function is obtained by solving self-consistently the Poisson equation. Once the Greens function is obtained the drain current is calculated by the following formula [4]:

$$I = \frac{2q}{h} \int_{-\infty}^{+\infty} dET(E)[f(E - E_{F_S}) - f(E - E_{F_D})] \quad (2)$$

where f is the Fermi-Dirac function and $T(E)$ is the transmission probability [3].

4 Results and Discussion

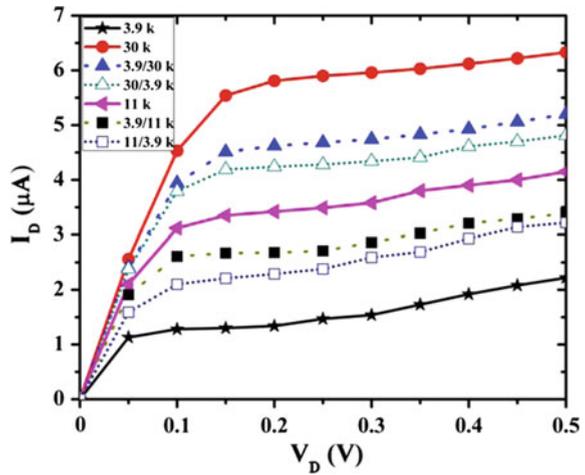
The various dielectrics that we have considered in our simulation are listed with their chemical formula, relative dielectric constant, and the energy band gap in Table 2 [8, 9].

The drain current versus drain voltage (I_D - V_D) characteristics are plotted for different gate dielectrics in Fig. 3. The I_D - V_D characteristics are plotted at the gate

Table 2 Different gate dielectric materials with their relative dielectric constants and energy band gap

Sl. No.	Gate dielectric material	Dielectric constant (k)	Energy band gap
1	SiO ₂	3.9	9
3	HfSiO ₄	11	6.5
7	La ₂ O ₃	30	5.8

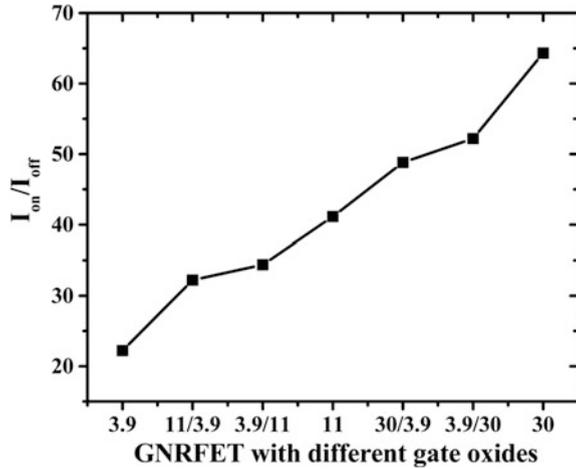
Fig. 3 Current–voltage characteristics of GNRFET with different gate dielectrics



voltage $V_G = 0.5$ V while varying the drain voltage V_D from 0 to 0.5 V. The on-current is obtained at $V_G = V_D = 0.5$ V. The plots are obtained by considering (i) top and bottom oxides same as SiO₂ which are indicated by ‘3.9 k’ curve; (ii) top and bottom oxides both having La₂O₃ as oxide layer which are shown by ‘30 k curve’; (iii) SiO₂ as top oxide and La₂O₃ as bottom oxide which are shown by ‘3.9/30 k’ curve; (iv) with La₂O₃ as top oxide and SiO₂ as bottom oxide which are shown by ‘30/3.9 k’ curve; (v) with SiO₂ on top and HfSiO₄ on bottom which are shown by ‘3.9/11 k’ curve; (vi) HfSiO₄ and SiO₂ on top and bottom, respectively, are indicated by ‘11/3.9 k’ curve; and (vii) with both top and bottom oxides as HfSiO₄ which are indicated by ‘11 k’ curve.

From Fig. 3, we have seen that the output characteristic of the GNRFET improves with the use of high-k dielectrics. The drain current for GNRFET with SiO₂ as gate dielectric is 2.2 μ A and it improves with the higher values of dielectric constant. With HfSiO₄ as the gate dielectric, the I_D increases to 4.15 μ A. The drain current further increases to 6.33 μ A with the use of La₂O₃ as gate dielectric. It is observed that the drain current increases with the higher dielectric constant because of higher gate capacitance in high-k dielectric material. The graph of asymmetric gate dielectrics with two different gate dielectrics on top and bottom oxide indicated by ‘3.9/30 k’, ‘30/3.9 k’, ‘11/3.9 k’, and ‘3.9/11 k’ also shows a rising trend in drain current. However, the graph also shows that the GNRFET with higher relative

Fig. 4 Variation of I_{on}/I_{off} ratio with combination of different gate dielectrics under consideration



dielectric constant at bottom oxide as compared to top oxide has higher drain current.

The on-to-off current ratio is shown in Fig. 4 with various gate dielectrics and their combination as considered in Fig. 3. The I_{on}/I_{off} ratio is an important parameter. It determines the switching speed and it is expected to be higher. From the graph we have seen that the I_{on}/I_{off} for GNRFET with SiO_2 has the lowest value with 22, while it has highest value of 63.3 in case of GNRFET with La_2O_3 as the gate dielectric. The rising trend in the I_{on}/I_{off} ratio is further confirmed by the GNRFET with other dielectrics and their combination indicated by the curves of '3.9/30 k', '30/3.9 k', '11/3.9 k', and '3.9/11 k' curve. The increase in the I_{on}/I_{off} ratio is due to the increase of drain current.

5 Conclusion

In this paper we have studied the performance of Schottky barrier GNRFET in the ballistic regime with different gate dielectric materials. From the results obtained we can conclude that the drain current improves with the use of high-k dielectrics in place of SiO_2 . We have seen the highest drain current is obtained with La_2O_3 as both top and bottom oxide layers. It is also observed that the relatively higher value of dielectric constant on bottom oxide than on top oxide has higher drain current as compared to the one having high-k dielectric on top gate oxide than on bottom oxide. Moreover, the I_{on}/I_{off} ratio also improves with high-k dielectric oxide which is clearly observed in the results.

References

1. Nagaraju, P.V., DasGupta, A.: Study of gate leakage current in symmetric double gate MOSFETs with high-k/stacked dielectrics. *Thin Solid Films* **504**, 317–320 (2006)
2. Frank, M.M., Kim, S., Brown, S.L., Bruley, J., Copel, M., Hopstaken, M., Chudzik, M., Narayanan, V.: Scaling the MOSFET gate dielectric: from high-k to higher-k? *Microelect. Eng.* **86**, 1603–1608 (2009)
3. Datta, S.: *Quantum Transport: Atom to Transistor*. Cambridge University Press, UK (2005)
4. Venugopal, R., Ren, Z., Lundstrom, M.S., Jovanovic, D.: Simulating quantum transport in nanoscale transistors: real versus mode-space approaches. *J. Appl. Phys.* **92**, 3730–3739 (2002)
5. Fiori, G., Iannaccone, G.: Simulation of graphene nanoribbon field-effect transistors. *IEEE Electron Device Lett.* **28**, 760–762 (2007)
6. Yoon, Y., Fiori, G., Hong, S., Guo, J., Iannaccone, G.: Performance comparison of graphene nanoribbon FETs with Schottky contacts and doped reservoirs. *IEEE Trans. Electron Devices* **55**, 2314–2323 (2008)
7. Datta, S.: Nanoscale device modeling: the Green's function method. *Superlattices Microstruct.* **28**, 253–278 (2000)
8. Robertson, J.: High dielectric constant oxides. *Eur. Phys. J. Appl. Phys.* **28**, 265–291 (2004)
9. Das, S., Kundu, S.: Simulation to study the effect of oxide thickness and high-k dielectric on drain-induced barrier lowering in n-type MOSFET. *IEEE Trans. Nanotechnol.* **12**(6), 945–947 (2013)

A Simple and Efficient Higher Order Finite Element Scheme for Helmholtz Waveguides

Tattwa Darshi Panda, K.V. Nagaraja and V. Kesavulu Naidu

Abstract This paper presents a simple and efficient finite element scheme for computing the cutoff wave numbers of arbitrary-shaped waveguides using higher order triangular elements. The waveguide geometry is divided into a set of triangular elements and each of these elements is mapped to a standard isosceles triangle by discretizing with subparametric finite elements. For waveguides containing arbitrary cross sections, the transformation is done using a series of higher order parabolic arcs. In this case, the curve boundaries are approximated by curved triangular finite elements and then transformed to an isosceles triangle. Numerical results are illustrated to validate the present approach. The obtained results have converged very well with the existing literature with minimum number of triangular elements, degree of freedoms, order of computational matrix, etc.

Keywords Helmholtz equation · Finite element method · Eigenvalue problem
Subparametric transformation · Waveguides · Parabolic arcs

1 Introduction

There are many numerical methods developed for computing accurate and optimal numerical solutions to waveguide eigenvalue problem. Sensale et al. have presented a unique approach for microwave waveguides based on hypersingular boundary

T.D. Panda (✉)

Department of Electronics and Instrumentation Engineering, Amrita School of Engineering,
Amrita Vishwa Vidyapeetham, Amrita University, Bangalore 560035, India
e-mail: tattwadarshipanda1994@gmail.com

K.V. Nagaraja · V. Kesavulu Naidu

Department of Mathematics, Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Amrita University, Bangalore 560035, India
e-mail: nagarajaitec123@yahoo.com

V. Kesavulu Naidu

e-mail: kesa78@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_45

element formation and can be found [1]. Dong et al. [2] have derived some solutions to waveguide eigenvalue problem based on generalized differential quadrature method. Numerical approximation to the two-dimensional Helmholtz equation by a meshfree approach can be found in [3, 4]. This paper is concerned with higher order finite element solution to the Helmholtz equation in waveguides using straight-sided and curved triangular elements. In the present case, solutions are determined for TM modes. For waveguides with rectangular boundaries, the geometry is discretized with a series of higher order finite elements and then suitably transformed to a standard isosceles right-angled triangle. Certain point transformation formulae are used for ensuring highest level of mapping. For waveguides with irregular cross sections, a series of higher order parabolic arcs are used to discretize the geometry. Rathod et al. [5] have derived a mapping procedure of mapping the curved boundary by a series of parabolic arcs. The point transformation formula in this case gives an interpolating polynomial that ensures approximation of the curved cross section by a parabolic arc in which two points lie along the arc and the other two points serve as the end points of the arc. The detailed analysis of this transformation procedure and its advantages is discussed by Nagaraja and Naidu [6]. Nagaraja et al. [7] have used the same approach for determining optimal numerical solutions to elliptic partial differential equations. In the present context, numerical illustration of the present approach is done for a single-ridge waveguide and vaned rectangular waveguide. The obtained computational values have converged very well to those existing in literature with minimum computational parameters. Later, a waveguide with a highly irregular cross section is considered. A good convergence is observed in this case with increasing the order of the transformation which validates the computational efficiency of the proposed method.

2 Finite Element Formulation

The waveguide geometry is first divided into a set of triangles and then each of the triangles is mapped by a subparametric finite element. Subparametric refers to higher order finite elements [5–7]. The transformation using subparametric finite elements is given by the following transformation formula [6, 7]:

$$u = \sum_{i=1}^{\frac{(m+1)(m+2)}{2}} N_i^{(m)}(r, s) u_i^e. \quad (1)$$

In Eq. (1), $N_i^{(m)}(r, s)$ denotes the shape functions for each triangular element and is different for all the order of transformation. The transformation of the waveguide geometry is shown in Figs. 1 and 2.

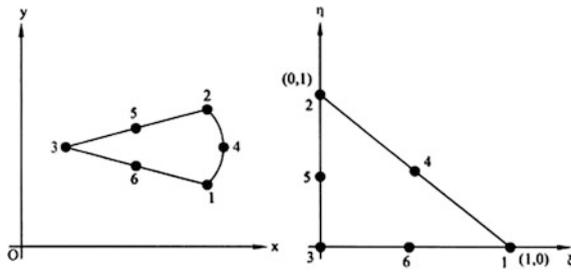


Fig. 1 Quadratic order transformation

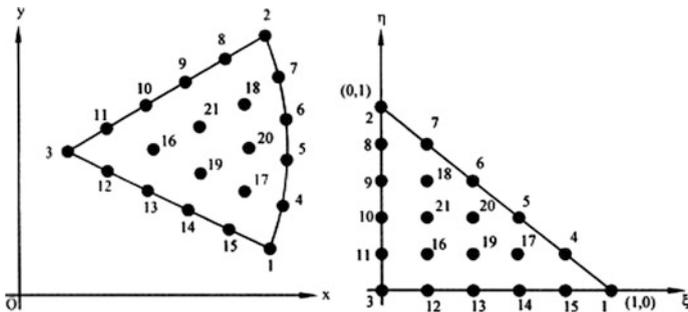


Fig. 2 Cubic order transformation

The above point transformation formula can be reduced to a standard interpolating polynomial by the following a symmetric procedure given in [5] and is reported below:

$$x = x_3 + (x_1 - x_3)r + (x_2 - x_3)s \tag{2}$$

$$y = y_3 + (y_1 - y_3)r + (y_2 - y_3)s. \tag{3}$$

The governing differential equation in the present case is the two-dimensional Helmholtz equation and is given below:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + k_c^2 u = 0. \tag{4}$$

TM mode of wave propagation is defined by the following equation [8]:

$$E_x = -z_0 \frac{\partial u}{\partial x}, \quad E_y = -z_0 \frac{\partial u}{\partial y}. \tag{5}$$

On applying the Galerkin finite element formulation, the Helmholtz equation is converted to a matrix form. The detail analysis of this conversion is done in our previous work [8]:

$$[F + G]_{NP \times NP} \times \{V\}_{NP \times 1} = 0. \tag{6}$$

The unknowns in the matrix form are clearly defined in [8] and are listed below for convenience:

$$F_{x,x}^{i,j} = \int_0^1 \int_0^{1-\xi} \left(\frac{1}{J}\right) \left(\frac{\partial N_i}{\partial \xi} \frac{\partial y}{\partial n} - \frac{\partial N_i}{\partial \eta} \frac{\partial y}{\partial \xi}\right) \left(\frac{\partial N_j}{\partial \xi} \frac{\partial y}{\partial n} - \frac{\partial N_j}{\partial \eta} \frac{\partial y}{\partial \xi}\right) d\eta d\xi \tag{7}$$

$$F_{y,y}^{i,j} = \int_0^1 \int_0^{1-\xi} \left(\frac{1}{J}\right) \left(-\frac{\partial N_i}{\partial \xi} \frac{\partial x}{\partial n} + \frac{\partial N_i}{\partial \eta} \frac{\partial x}{\partial \xi}\right) \left(-\frac{\partial N_j}{\partial \xi} \frac{\partial x}{\partial n} + \frac{\partial N_j}{\partial \eta} \frac{\partial x}{\partial \xi}\right) d\eta d\xi \tag{8}$$

$$G_{i,j}^e = \iint_{\Omega_e} k_c^2 N_i N_j dX dY = \int_{\xi=0}^1 \int_{\eta=0}^{1-\xi} k_c^2 N_i N_j J^e d\eta d\xi. \tag{9}$$

The above integrations are computed by an efficient numerical procedure developed by Rathod et al. [5].

3 Results and Discussions

3.1 Single-Ridge Waveguide

The domain of a single-ridge waveguide considered here is the same as in [9]. In the present case the domain is divided into 12 and 24 triangular elements for the quadratic case and 12 triangular elements for the cubic case. A good convergence is observed with the numerical results in [9] and is reported in Table 1. In [9], the geometry is discretized into 96 subregions comparably to the present case where the geometry is discretized into 16 subregions. In [9], the boundaries of the ridge waveguide are discretized using 112 subcontours. In the present case, the boundary is divided into 42 subcontours. The size of the computational matrix in [9] is

Table 1 Cutoff wavenumbers of a single-ridge waveguide

TM mode No.	[9]	Present case (cubic)
1	12.2338	12.2393
2	12.4106	12.5294
3	14.2152	14.1633
4	15.8221	15.7512

Fig. 3 Convergence rate of eigenvalues

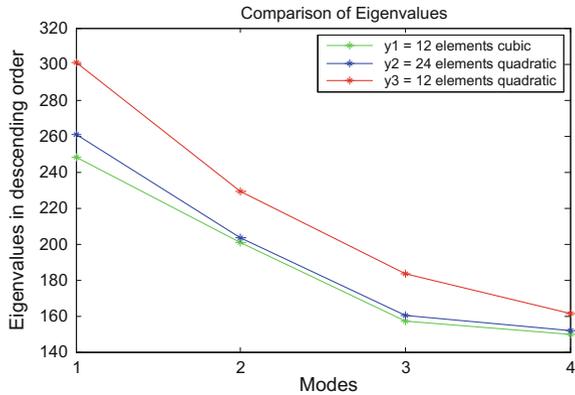


Table 2 Cutoff wavenumbers of a vaned rectangular waveguide

[1] 56 elements	[1] 112 elements	[10]	[11] 19 × 19	[11] 23 × 23	Present case (cubic)
3.6827	3.6881	3.6770	3.6939	3.6932	3.6548
4.9425	4.9583	4.9279	4.9745	4.9726	4.9826
6.4726	6.4761	6.4151	6.4747	6.4721	6.5943
7.0096	7.0231	7.0220	7.0248	7.0248	7.0346

96 × 96. In the present case the size of the computational matrix for the cubic case is 43 × 43. Similarly, the size of the computational matrix for 24 elements quadratic case is 35 × 35 and 21 × 21 for 12 elements quadratic case. The difference between the cutoff wavenumber obtained in the present case and [9] is 0.04%. For the second, third, and fourth modes, the differences are 0.95, 0.36, and 0.44%, respectively. The cutoff wavenumber obtained for 24 elements quadratic case is 12.3269. The maximum difference of this case with [9] is 0.76%. For 12 elements quadratic case, the cutoff wavenumber obtained is 12.7015. The convergence rate among the eigenvalues is given in Fig. 3. A good convergence is observed when the order of discretization is observed from quadratic to cubic.

3.2 Vaned Rectangular Waveguide

The domain of a vaned rectangular waveguide can be found in [1]. In the present case the domain is divided into eight elements for both the quadratic and cubic cases. The results are reported in Table 2. It can be observed that a good convergence is obtained in the present case. In [1], the accuracy is obtained using 56 and 112 elements comparably to present method with just eight elements. The difference (%) of the cutoff wavenumber obtained in the present method with 56 elements and 112 elements case of [1] is 0.75 and 0.90%. Similarly, the difference (%) with [10] is 0.60%. A comparison of the meshes used in the present case and [11]

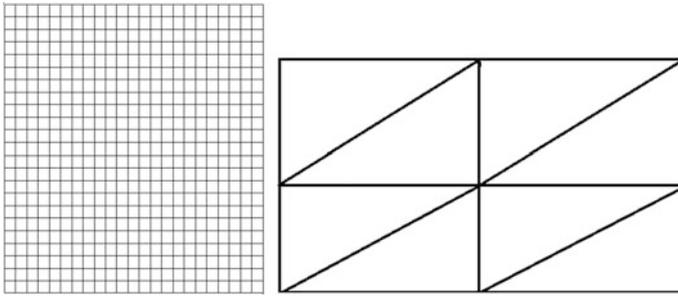


Fig. 4 Mesh used in [11] and present case (cubic)

Table 3 Cutoff wavenumbers of a waveguide with irregular geometry

Quadratic	15.411	12.545	10.544	9.6332	6.982
Cubic	12.210	11.433	10.689	8.932	6.271
Quartic	11.410	10.647	10.253	8.681	6.217
Quintic	11.254	10.538	10.232	8.596	6.215

is done in Fig. 4. A very refined mesh is used in the present case and hence the computational effort and time will be fairly less. More the number of elements used in the computation process, more will be the computational cost.

4 Waveguide with Curved Cross Sections

A waveguide with highly irregular cross sections is been considered for demonstrating the accuracy of the proposed method over curved boundaries and the results obtained for all order of discretization are reported in Table 3. For this case the rate of cutoff wavenumbers is compared for various orders of discretization. Also it should be noted that, for irregular geometries, it is required to go for higher order discretization since the result does not converge well at lower orders. The domain of the waveguide considered here is shown in Fig. 5. This domain is extracted from the works of Nagaraja et al. in [7]. The domain is divided into eight triangular elements as shown in Fig. 6. In the first case, the domain is discretized using cubic order curved triangular elements. The curve boundaries of elements 5, 6, 7, 8 are predominantly replaced by cubic arcs according to the transformation formula given in [6]. Similarly for the second and third cases, the discretization is done using quartic and quintic order curved triangular elements (Figs. 8 and 9). Hence, in both these cases, the curve boundaries of elements 5, 6, 7, 8 are implicitly replaced by quartic and quintic order parabolic arcs. The interpolating polynomial for these transformation are derived in the works of Rathod et al. [5] and the advantages of these transformation techniques are discussed in the works of Naidu and Nagaraja in [6, 7]. The results have converged very well when the order of discretization is increased from quadratic to quintic. The greatest accuracy is obtained at quintic level (Fig. 9).

Fig. 5 Waveguide geometry

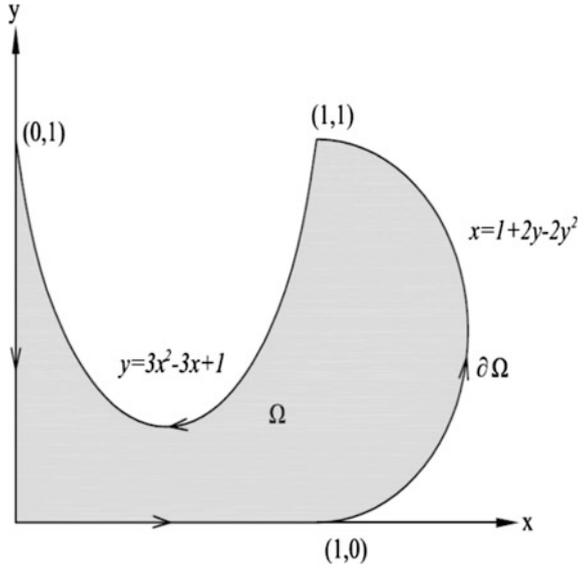
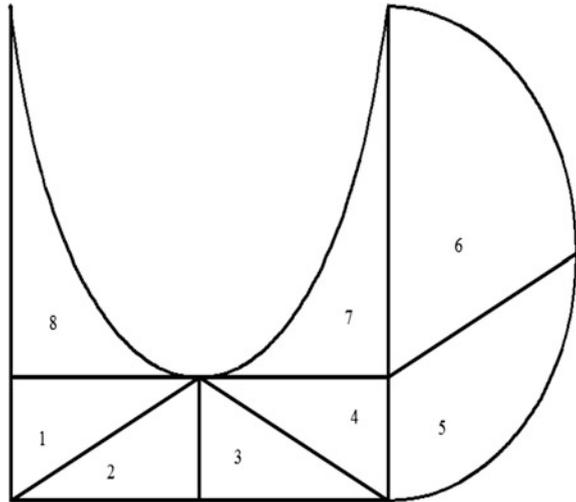


Fig. 6 Eight elements division



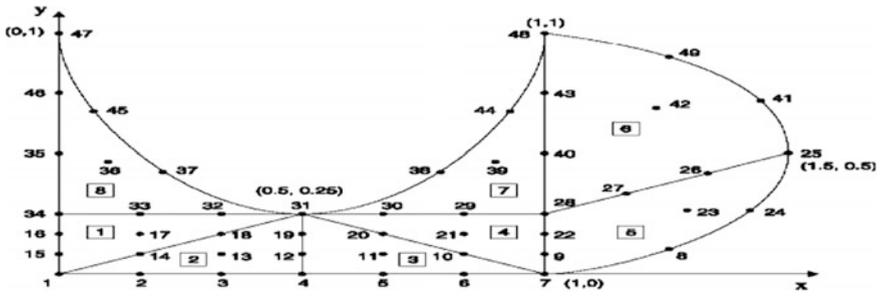


Fig. 7 Cubic order discretization

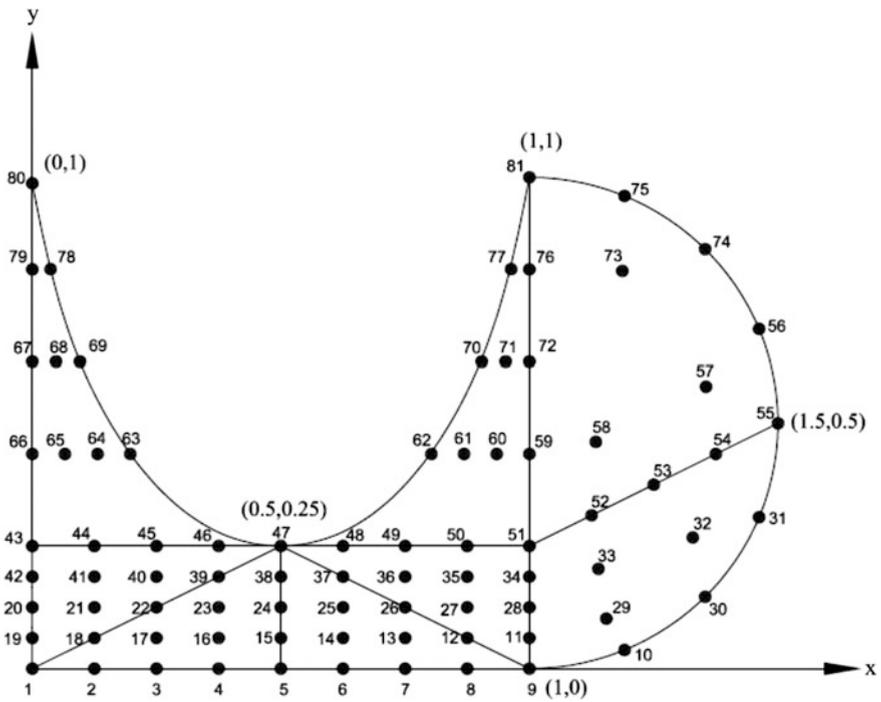


Fig. 8 Quartic order discretization

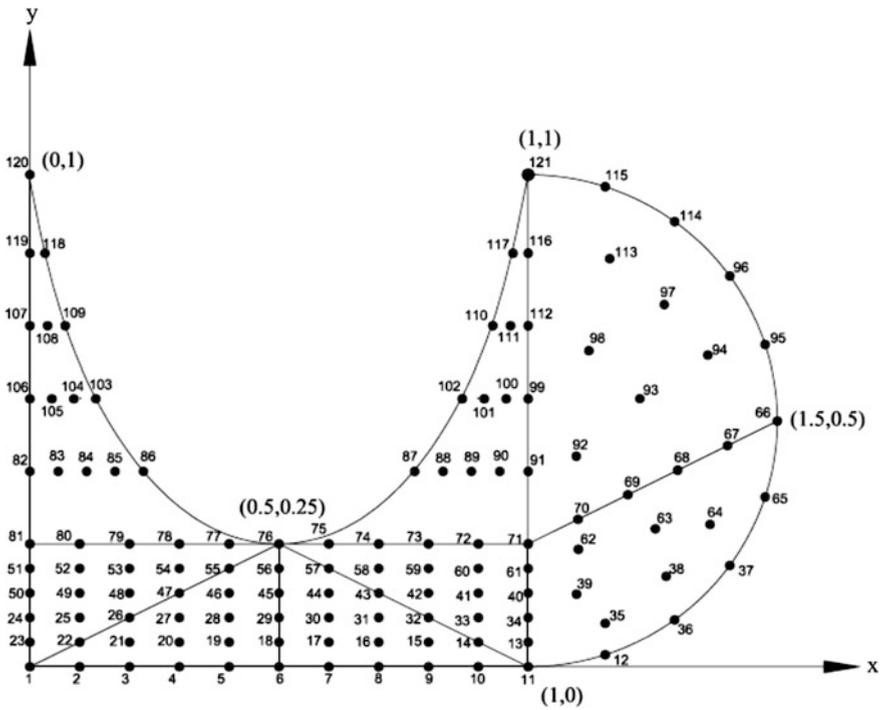


Fig. 9 Quintic order discretization

5 Conclusion

In the present work, the concept of mapping waveguides geometries using higher order triangular element is been introduced. The principle is further extended to map-curved boundaries of a waveguide by a series of curve triangular elements rather than straight-sided elements. A robust code is developed and can be applied to any kind of geometries. The results have converged very well with the existing numerical values in literature.

References

1. Sensale, B.S., Sensale, B., Leitao, V.: Determination of the TE and TM modes in arbitrarily shaped waveguides using a hypersingular boundary element formulation. *Int. J. Electron. Commun.* **62**, 576–581 (2008)
2. Dong, Y.C., Leong, M.S., Kooi, P.S., Lam, K.Y., Shu, C.: Computation of the propagation characteristics of TE and TM modes in waveguides with the use of the generalized differential quadrature method. *Microw. Opt. Technol. Lett.* **14**, 39–44 (1997)

3. Wang, L.: Convex meshfree solutions for arbitrary waveguide analysis in electromagnetic problems. *Prog. Electromagn. Res. B* **48**, 131–149 (2013)
4. Kowalczyk, P., Mrozowski, M.: Mesh-free approach to Helmholtz equation based on radial basis functions. *J. Telecommun. Inf. Technol.* **2**, 71–74 (2005)
5. Rathod, H.T., Nagaraja, K.V., Naidu, V.K., Venkatesudu, B.: The use of parabolic arcs in matching curved boundaries by point transformations for some higher order triangular elements. *Finite Elem. Anal. Des.* **44**, 920–932 (2008)
6. Naidu, V.K., Nagaraja, K.V.: Advantages of cubic arcs for approximating curved boundaries by transformation for some higher order triangular elements. *Appl. Math. Comput.* **219**, 6893–6910 (2013)
7. Nagaraja, K.V., Naidu, V.K., Siddheshwar, P.G.: Optimal subparametric finite elements for elliptic partial differential equations using higher-order curved triangular elements. *Int. J. Comput. Meth. Eng. Sci. Mech.* **15**, 83–100 (2014)
8. Panda, T.D., Nagaraja, K.V.: Finite element method for eigenvalue problem in waveguide modes. *Global J. Pure Appl. Math.* **11**, 1241–1251 (2015)
9. Balagangadhar, M., Sarkar, T.K., Rejeb, J., Boix, R.R.: Solution of general Helmholtz equation in homogeneously filled waveguides using a static Green's function. *IEEE Trans. Microw. Theory Tech.* **46**, 302–307 (1998)
10. Swaminathan, M., Arvas, E., Sarkar, T.K., Djordjevic, A.R.: Computation of cutoff wavenumbers of TE and TM modes in waveguides of arbitrary cross sections using a surface integral formulation. *IEEE Trans. Microw. Theory Tech.* **38**, 154–159 (1990)
11. Shu, C., Chew, Y.T.: Application of multi-domain CDQ method to analysis of waveguides with rectangular boundaries. *Prog. Electromagn. Res.* **21**, 1–19 (1999)

Point Feature Based Recognition of Handwritten Meetei Mayek Script

Chandan Jyoti Kumar and Sanjib Kumar Kalita

Abstract In this article we are proposing a hybrid point feature based recognition of Meetei Mayek optical handwritten characters. Four different interesting point features namely Harris corner detector, Laplacian-of-Gaussian (Log) detector, Harris-Laplacian detector, and Gilles feature are used to detect different fixed number of interesting feature points. The $[X, Y]$ coordinates of these interesting points are used as feature vector. Four different feature vectors are generated out of these and we have developed a fused feature out of it. Feature reduction is done by PCA, and accuracy calculations for all these features are performed using support vector machine. A comparative analysis is done to get a feature that shows consistency in accuracy before and after applying feature reduction. Experimental result shows that fused feature FV5 shows better accuracy in all different situations. Before reduction it shows 97.16% accuracy, on reduction by probabilistic PCA it shows 94.15% accuracy, and PCA-based reduction 97.16% accuracy, which is best in all situations as compared to other features.

Keywords Point feature · Meetei Mayek · Optical character recognition
Feature reduction · SVM

1 Introduction

Northeast India consists of eight states called eight sisters, the states of Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura. Except Assamese and Bengali, most of the languages used by the native speakers

C.J. Kumar (✉)

Department of Computer Science & IT, Cotton University, Guwahati, India
e-mail: chandan14944@gmail.com

S.K. Kalita

Department of Computer Science, Gauhati University, Guwahati, India
e-mail: sanjib959@reddiffmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_46

431

are of Sino-Tibetan and Austro-Asiatic origin. Meetei is one of the major northeast Indian regional languages. It is primarily spoken by majority of the native speaker in the state of Manipur. Meetei Mayek script is used for writing Meetei (Manipuri) language. The script has a long history behind it. The Meetei Mayek script was used in earlier days for writing Meetei Mayek till eighteenth century, and most of the historical documents found are written using Meetei Mayek script. However, there was a mass destruction of documents written in Meetei Mayek in eighteenth century, and Bengali script was adopted for writing Meetei thereafter. Good news is that researchers worked over the script and finally in 1976, they have finalized the characters in the script having 18 characters and added 9 more letters which are derivatives of previous letters and suggested the government to make the script popular in the state of Manipur [1]. In 1983, the education department of Manipur Government starts publishing text book written in Meetei Mayek.

1.1 Meetei Mayek Script

Meetei Mayek script has 18 characters in original which are called Eeyek Eepee and 9 additional characters incorporated are called Lom Eeyek. Figures 1 and 2 show the character set of Meetei Mayek script. The most interesting fact about the script is some of the letters used in Meetei Mayek refer to human body part.

1.2 Motivation of the Work

According to the website ‘ethnologue.com’, which does a lot of survey and statistical analysis on languages, it has mentioned that currently 7102 living languages are available on earth. Based on socioeconomical and political scenario, this value fluctuates a little. Recent trend is that the number of living languages is always going down, which is becoming an alarming matter. An article published by Unesco in 2009, says that most of the endangered languages belong to India. India is ahead of all in the list of languages in endangered state. A language is said to be endangered when the number of people using that language drastically goes down. There are many possible reasons for this to happen. One of the major causes is economical,

Fig. 1 Eeyek Eepee characters



Fig. 2 Lom Eeyek characters



social, and cultural interference of one community over native speakers. Another cause may be the ignorance toward the language by the natives. There are cases where some people feel ashamed of using regional language in public gatherings. There are people who think they will lose their status if they use their own language and that is why they start speaking a different language. We must respect and learn national and International languages, but it does not imply to sacrifice the native language. A lot of effort is put on developing the OCR and speech processing technology in some parts of India [2, 3]. However, such initiative is very less in northeast India [4, 5]. That is why major northeast Indian languages are facing this problem. Keeping in mind the need of hour, we are trying to develop an OCR system that can digitize the optical image written in regional language. This application will be of ample use and it will certainly help sustainability of the language and the script as well. If we can make our historical documents digitize, it will bring the attraction of the natives. If we can process handwritten optical characters of the local scripts, then people will definitely be encouraged to use the script [6].

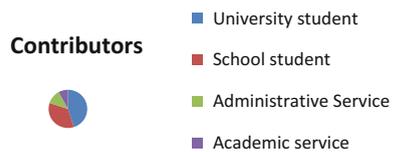
The rest of the paper is organized as follows: Sect. 2 discusses about dataset preparation and preprocessing; Sect. 3 discusses about algorithms; and Sect. 4 discusses about results related with feature extraction and reduction and finally we conclude.

2 Dataset Preparation and Preprocessing

Standard data collection sheet of A4 size was provided to the contributors for writing the characters of Meetei Mayek script. Each of the characters was written ten times in a row. Separate characters are written in separate row. One page per contributor was collected. The data samples are collected from people with various professional backgrounds and of various age groups. Figure 3 shows the diverse distribution of contributors. However, most of the contributors are from S.S.N. school, Cotton University, and Gauhati University.

The individual characters are segmented and noise removal and morphological-based operations are performed to make these samples compatible with feature extraction techniques.

Fig. 3 Data sample contributor's distribution



3 Feature Extraction Algorithms

Feature extraction is the most crucial phase of optical character recognition system. Proper feature extraction and selection has the direct impact on accuracy and efficiency of any classification technique [7]. Literature review reveals that researchers have investigated and come up with different feature set efficient for different script recognitions [8, 9]. For Meetei Mayek handwritten character recognition, we have investigated four different types of feature point extraction methods. The Harris corner detector [10, 11], Laplacian-of-Gaussian (Log) [12], Harris-Laplacian [13], and Gilles [14] point feature based feature vectors and combinations of them are used. Algorithms 3.1, 3.2, 3.3, 3.4, and 3.5 explain how these features are extracted.

Algorithm 3.1 To compute Harris corner vector FV1

Input: Optical image of isolated handwritten Meetei Mayek character.

Output: Harris corner vector feature vector with maximum dimension 1×100 .

- Step 1. Compute Gaussian derivatives at each pixel
- Step 2. Compute second moment matrix M in a Gaussian window around each pixel. Using a Gaussian smoothing kernel of standard deviation s , calculate the sampled means: $\langle I_x^2 \rangle, \langle I_y^2 \rangle, \langle I_x I_y \rangle$.
- Step 3. Compute corner response function,

$$R = \frac{\langle I_x^2 \rangle + \langle I_y^2 \rangle}{\langle I_x^2 \rangle \langle I_y^2 \rangle - \langle I_x I_y \rangle^2}$$

- Step 4. Set criteria to decide threshold
- Step 5. Find all local maxima of response function greater than threshold

$$\text{Points} = \text{find}(\text{max_local} > = R)$$

- Step 6. FV1=combine(X coordinates of Points, Y coordinates of Points).

Algorithm 3.2 To compute Laplacian-of-Gaussian (Log) vector FV2

Input: Optical image of isolated handwritten Meetei Mayek character.

Output: Log vector feature vector with maximum dimension 1×400 .

- Step 1. Set Laplacian-of-Gaussian parameters
- Step 2. Calculate scale-normalized Laplacian operator

$$L(x, \sigma) = \sigma^2(I_{xx}(x, \sigma) + I_{yy}(x, \sigma)).$$

Step 3. Search for local maxima points

Step 4. FV2=combine(X coordinates of Points, Y coordinates of Points).

Algorithm 3.3 To compute Harris-Laplacian vector FV3

Input: Optical image of isolated handwritten Meetei Mayek character.

Output: Log vector feature vector with maximum dimension 1×120 .

Step 1. Set derivative masks

Step 2. Calculate sum of the auto-correlation matrix

Step 3. Find interest point response

Step 4. Search for local maxima points

Step 5. Find local maxima greater than threshold

$$Hpoints = \text{find}(\max_local \geq t)$$

Step 6. Compute scale-normalized Laplacian operator.

Step 7. Points= Hpoints for which LoG attains extrema.

Step 8. FV3=Combine(X coordinates of Points, Y coordinates of Points).

Algorithm 3.4 To compute Gilles detector vector FV4

Input: Optical image of isolated handwritten Meetei Mayek character.

Output: Log vector feature vector with maximum dimension 1×50 .

Step 1. Set circular filter mask

Step 2. Compute local entropy with that filter

Step 3. Find the local maxima based on entropy and mask

Step 4. Keep only points above a threshold

Step 5. FV4=Combine(X coordinates of Points, Y coordinates of Points).

Algorithm 3.5 To compute Hybrid interesting point detector vector FV5

Input: Optical image of isolated handwritten Meetei Mayek character (Img).

Output: Feature vector with maximum dimension 1×670 .

Step 1. PTH=Algorithm3.1(Img)

Step 2. PTL=Algorithm3.2(Img)

Step 3. PTHL=Algorithm3.3(Img)

Step 4. PTG=Algorithm3.4(Img)

Step 5. FV5=Combine(PTH, PTL, PTHL, PTG) .

4 Results and Analysis

In the present study, the feature vectors **FV1** (Harris corner detector), **FV2** (Laplacian-of-Gaussian detector), **FV3** (Harris-Laplacian detector), **FV4** (Gilles feature detector), and **FV5** (Fused feature of FV1, FV2, FV3, and FV4) are extracted using Algorithms 3.1, 3.2, 3.3, 3.4 and 3.5, respectively. These five different feature vectors are extracted for 6750 image samples, out of which 5062 image features are used for training the support vector machine and rest is used for testing the SVM classifier recognition accuracy. Table 1 shows the size and recognition accuracy corresponding to the five features before applying any feature reduction technique.

The point features and their combinations are showing good accuracy and the best thing about this feature is that they are scale invariant. However, from Fig. 4 it is observed that the feature vector size is very large and fusion of features makes the situation worse. Therefore, we have applied some standard feature reduction technique over these features and experiment is done on the reduced feature set.

After applying probabilistic PCA reduction technique, we get the following features: **R1FV1**: Probabilistic PCA-reduced **FV1**, **R1FV2**: Probabilistic PCA-reduced FV2, **R1FV3**: Probabilistic PCA-reduced FV3, **R1FV4**: Probabilistic PCA-reduced FV4, and **R1FV5**: Probabilistic PCA-reduced FV5. Table 2 shows the size and recognition accuracy corresponding to the five features after applying probability PCA feature reduction technique. From Fig. 5, it is observed that after applying probabilistic PCA-based reduction, even though we are able to minimize the feature vector size, only the fused feature FV5 is able to show good recognition accuracy.

After applying PCA reduction technique, the features we get are as follows: **R2FV1**: PCA-reduced FV1, **R2FV2**: PCA-reduced FV2, **R2FV3**: PCA-reduced

Table 1 Size and accuracy of non-reduced features

Feature vector	Feature vector size	Accuracy (in %)
FV1	100	97.16
FV2	400	97.16
FV3	120	92.97
FV4	50	79.66
FV5	670	97.16

Fig. 4 Size and accuracy of non-reduced features

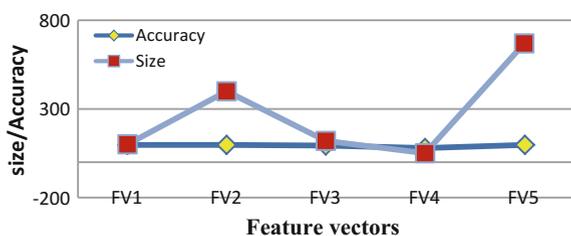


Table 2 Size and accuracy of probability PCA-reduced features

Feature vector	Feature vector size	Accuracy (in %)
R1FV1	18	56.53
R1FV2	18	54.25
R1FV3	18	0
R1FV4	18	41.55
R1FV5	70	94.15

Fig. 5 Size and accuracy of probability PCA-reduced feature

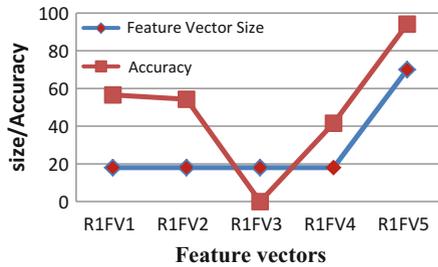
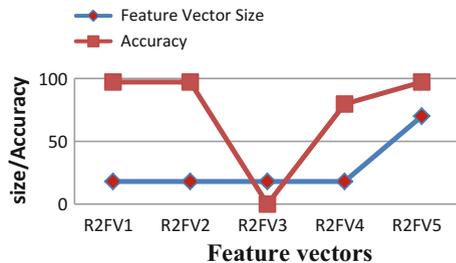


Table 3 Size and accuracy of PCA features

Feature vector	Feature vector size	Accuracy (in %)
R2FV1	18	97.16
R2FV2	18	97.16
R2FV3	18	0
R2FV4	18	79.66
R2FV5	70	97.16

Fig. 6 Size and accuracy of PCA-reduced features



FV3, **R2FV4**: PCA-reduced FV4, and **R2FV5**: PCA-reduced fused FV5. Table 3 shows the size and recognition accuracy corresponding to the five features after applying PCA-based feature reduction technique. From Fig. 6 it can be observed that PCA-based reduction is very helpful for dimensionality reduction. Accuracy of only few of the feature gets altered.

Combining all the results, we have drawn the graph representing the feature versus accuracy after and before reduction, in Fig. 7. From Fig. 7, it can be observed that for feature FV5, SVM classifier shows highest accuracy for normal as well as reduced feature space. Even though some features show good accuracy for

Fig. 7 Accuracy of non-reduced, PCA-reduced and ProbPCA-reduced features

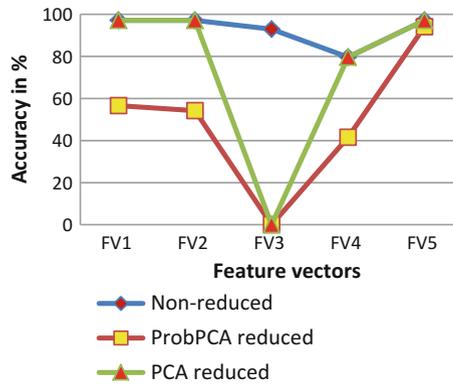
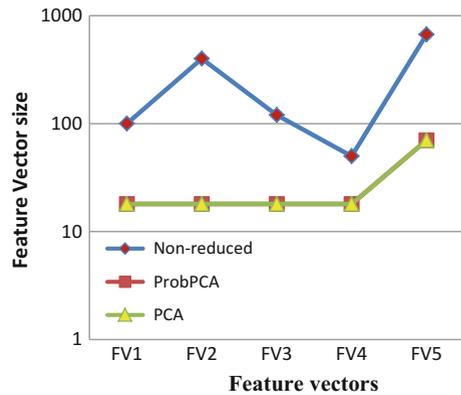


Fig. 8 Size of non-reduced, PCA-reduced and ProbPCA-reduced feature



normal and PCA-reduced feature vector, they are not consistent with probabilistic PCA. Again feature reduction leads to a very small-sized feature with good accuracy. From Fig. 4, it is quite obvious. So, using the FV5, which is a fused corner feature, Meetei Mayek handwritten script can be recognized with better accuracy. Figure 8 shows that after applying feature reduction technique the size of the fused feature gets so reduced that its size is smaller than most of the non-reduced features.

5 Conclusion

In this work, SVM classification accuracy with various feature vectors and their different combinations is investigated. However, some feature vectors are quite large, and fusion of two or more of them leads to a huge vector size. It leads us to think about reducing the dimensionality of the feature. The experiment is performed with PCA and probabilistic PCA to reduce the feature vector size. Some point features show good accuracy without reduction, but when reduced by a technique, it

does not give the same performance. Consistency of accuracy is much required, before and after reduction. Fused feature FV5 shows better accuracy in all different situations, hence preferable to use for recognition of Meetei Mayek script.

The entire work till now is focused on different possible point features, their combinations, reduction of feature vector, and SVM-based classification [15]. Different feature vectors may show different behaviors toward different classifiers. Introduction of feature descriptor may also give better recognition rate at various orientations. In future combination of various classifiers may be investigated.

References

1. Thokchom, T., Bansal, P.K., Vig, R., Bawa, S.: Recognition of Handwritten Character of Manipuri Script. *J. Comput.* **5**(10) (2010)
2. Gupta, V., Lehal, G.S.: A survey of common stemming techniques and existing stemmers for Indian languages. *J. Emerg. Technol. Web Intell.* **5**(2), 157–161 (2013)
3. Arora, T., Dhir, R.: Correlation-based feature selection and classification via regression of segmented chromosomes using geometric features. *Med. Biol. Eng. Comput.* 1–13 (2016)
4. Saharia, N., Sharma, U., Kalita, J.: Stemming resource-poor Indian languages. *ACM Trans. Asian Lang. Inform. Process.* **13**(3) (2014)
5. Roy, P., Das, P.K.: A hybrid VQ-GMM approach for identifying Indian languages. *Springer Int. J. Speech Technol.* **15**(2) (2012)
6. Bhagvati, C., Ravi, T., Kumar, M., Negi, A.: On developing high accuracy OCR systems for Telugu and other Indian scripts. *Lang. Eng. Conf. Proc.* (2002)
7. Aggarwal, A., Singh, C.: Zernike moments-based Gurumukhi character recognition. *Appl. Artif. Intell.* **30**(5), 429–444 (2016)
8. Sinha, G., Dhir, R., Rani, R.: Handwritten Gurumukhi numerals recognition using zone based hybrid feature extraction techniques. *Int. J. Comput. Appl.* (0975-888) **47**(21), 24–29 (2012)
9. Zhao, G., Ahonen, T., Matas, J., Pietikäinen, M.: Rotation-invariant image and video description with local binary pattern features. *IEEE Trans. Image Process.* **21**(4) (2012)
10. Harris, C.G., Stephens, M.J.: A combined corner and edge detector. In: *Proceedings Fourth Alvey Vision Conference, Manchester*, pp. 147–151 (1988)
11. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. *Int. J. Comput. Vis.* **37**(2), 151–172 (2000)
12. Lindeberg, T.: Feature detection with automatic scale selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 77–116 (1998)
13. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. *Int. J. Comput. Vis.* **60**(1), 63–86 (2004)
14. Gilles, S.: Robust description and matching of images. Ph.D. thesis, University of Oxford (1998)
15. Richarz, J., Vajda, S., Grzeszick, R., Fink, G.A.: Semi-supervised learning for character recognition in historical archive documents. *Pattern Recogn.* **47**, 1011–1020 (2014)

Design of Circularly Polarized Antenna with Different Iterations for UWB Applications

Piyush Sharma, Ashok Yadav and Vinod Kumar Singh

Abstract A microstrip patch antenna with the strip line feed along with partial ground is anticipated for circularly polarized characteristics. The antenna resonates at 7 GHz frequency. It comprises a ground, a substrate, and patch. The simulation is performed in HFSS software and the simulated results of proposed antenna such as reflection coefficient and axial ratio are presented. The circular polarization (CP) is achieved by inserting slots of the different shapes and sizes, i.e., rectangle and the circle of the different dimensions in the structure of the antenna. The antenna is very compact (40 mm × 30 mm × 1.6) in size and simple to design. The outcomes reflection coefficient, axial ratio, gain, and radiation pattern of the proposed antenna remain valid for operating frequency band. The proposed antenna has been designed, simulated, and fabricated. The prototype antenna has also been characterized with -31 dB return loss. Measured result shows a good agreement with simulated and theoretical ones.

Keywords Return loss · VSWR · Axial ratio · HFSS software

1 Introduction

The microstrip antenna has numerous useful properties which includes minute size, low-cost of the fabrication, light weight, and ease of installation but the main limitations of printed antennas remain their narrow bandwidth features which limit

P. Sharma · A. Yadav · V.K. Singh (✉)
Krishna Engineering College, Mohan Nagar, Ghaziabad, UP, India
e-mail: singhvinod34@gmail.com

P. Sharma
e-mail: piyush.sharma5589@gmail.com

A. Yadav
e-mail: ashok.biet@gmail.com

P. Sharma · A. Yadav · V.K. Singh
Department of Electrical Engineering, S R Group of Institutions, Jhansi, UP, India

the range of frequency over which the antenna can work effectively. Microstrip antenna comprises three most important parts which are substrate, patch, and ground. The thickness of patch antenna is less than the free space wavelength; it means $t \ll \lambda_0$, where λ_0 is the wavelength in the free space. One side is radiating patch and other side is ground plane and a dielectric substrate sandwiched between it. The conducting patch is placed on the dielectric substrate which is used as radiating element. On other side of dielectric substrate, there is conducting layer used as ground part [1–4]. Nowadays, the fast development of modern communication systems is required for transportable devices for some important features which include easy designing, light weight, small in size, compatible with micro-wave, millimeter wave integrated circuits, less production cost, and easy fabrication of microstrip antennas.

High data rate operating ability in the multipath fading environment and a large bandwidth capability of supporting numerous applications have become the unavoidable feature of any modern wireless communication system. Therefore, there remains continuous increasing demand of the performance objectives of the printed antenna mostly used in such systems. In this regard, antennas which support circular polarization (CP) are being preferred and their properties are continuously investigated by researchers around the world [5–9]. Also used in radar and satellite navigation applications where the receiver or transmitter orientation is not clearly known, CP antennas have been found to be very effective in overcoming the harmful effects of multipath fading and thus give high data transmission rates. Another recent application of CP antennas is in RFID systems where a continuously emitted rotating field has the ability of tagging any object coming in its way. In principle, circular polarization is achieved when the generated electric field has two orthogonal modes with equal amplitude and 90° phase difference between them [10–16]. The usual techniques to achieve CP in a microstrip antenna include insertion of asymmetries in regular geometries, and deployment of different patches in order to rotate the electric field vector. CP antennas designed on microstrip patch structure suffer from low impedance bandwidth due to the inherent high quality (Q) factor [17–21].

In this article, a prototype of novel microstrip line-fed patch antenna with partial ground is proposed for circular polarization. Circular polarization mainly comes in existence where reception is required by moveable receivers or receivers in vehicles and also for fixed reception.

2 Antenna Configuration

The prototype of proposed antenna has been fabricated with FR4 dielectric substrate having relative permittivity of 4.4, and thickness of 1.6 mm as shown in Fig. 1. The radiating element of the proposed antenna prototype is patch structure printed over the substrate which comprises circular and rectangular slots. The rectangular dielectric substrate with the overall dimensions of 40×30 mm and the

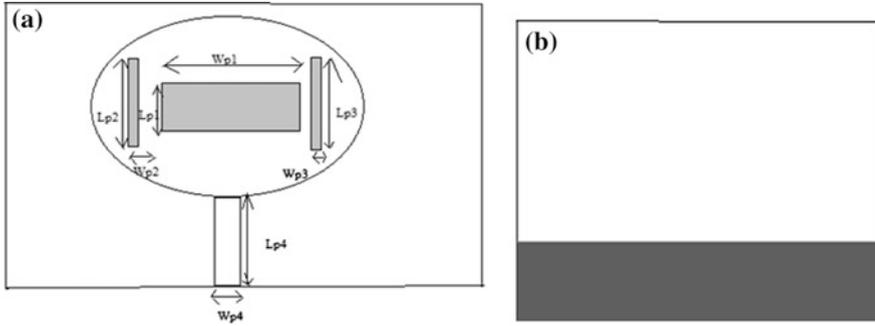


Fig. 1 Proposed antenna geometry **a** front view **b** back view

antenna structure comprises circles and rectangles which are printed over the FR4 substrate such that antenna shows the optimum axial ratio required to achieve the circular polarization and the other parameters such as the return loss, Gain, VSWR, etc. Patch of proposed antenna is designed with the rectangles and the circles and several geometries have been inserted to achieve the optimum return loss and the required axial ratio for the circular polarization as the axial ratio below 3-dB is required for the circular polarization; hence, several changes have to be made to attain the required axial ratio with the accepting return loss and the gain. The proposed antenna came into existence with several iterations in order to achieve the desired results which are essential for the circular polarization. Initially, basic structure is made using the circle and the rectangle and then several slots have been inserted in it, the rectangular and the circular shape of the different dimensions as discussed in Table 1. The slots are inserted in order to achieve the desired return loss and the axial ratio which is necessary for the circular polarization. Copper served as the conductive material for the slots of the antenna in order to achieve the desirable conductivity.

The following equations are used to calculate the dimensions of rectangular microstrip patch antenna such as length and the width of the patch and length and width of ground plane:

$$W = \frac{c}{2f\sqrt{(\epsilon_r + 1)/2}} \tag{1}$$

$$\epsilon_{r\text{eff}} = \frac{(\epsilon_r + 1)}{2} + \frac{(\epsilon_r - 1)}{2} \left[1 + 10 \frac{h}{W} \right]^{-\frac{1}{2}} \tag{2}$$

Table 1 Antenna design specifications

Parameters	L_{p1}	L_{p2}	L_{p3}	L_{p4}	L
Unit (mm)	4	12	12	10	30
Parameters	W_{p1}	W_{p2}	W_{p3}	W_{p4}	W
Unit (mm)	10	1	1	3.4	40

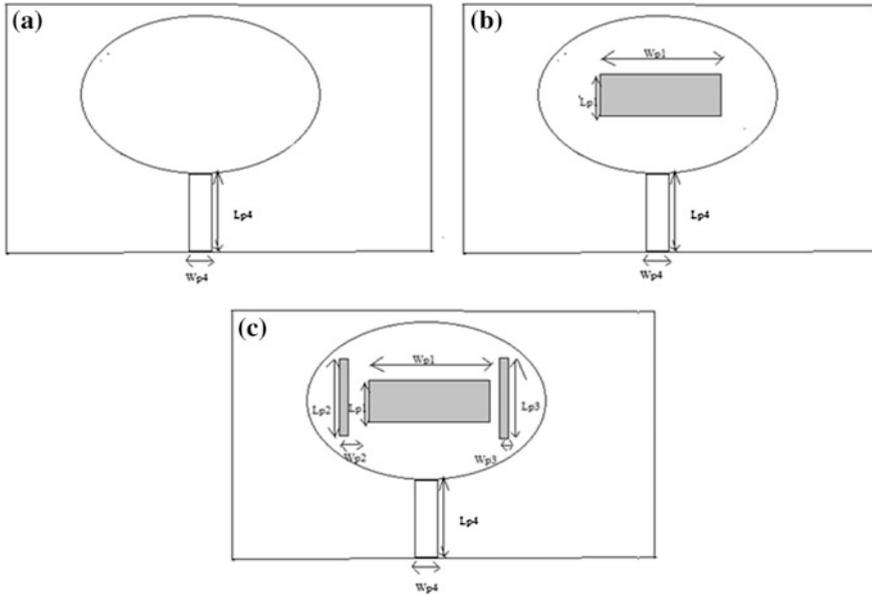


Fig. 2 Geometry of the proposed antennas **a** antenna1, **b** antenna2, **c** antenna3

$$\frac{\Delta l}{h} = 0.412 \frac{(\epsilon_{reff} + 0.300) \left(\frac{W}{h} + 0.262\right)}{(\epsilon_{reff} - 0.258) \left(\frac{W}{h} + 0.813\right)} \tag{3}$$

$$L = \frac{c}{2f\sqrt{\epsilon_{reff}}} - 2\Delta l \tag{4}$$

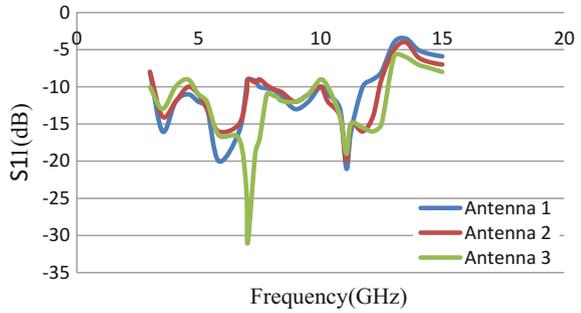
$$L_g = L + 6h \tag{5}$$

$$W_g = W + 6h.$$

3 Optimization of Antenna Geometries

Figure 2 shows the geometry of the proposed antenna1, antenna2, and antenna3 and Fig. 3 shows the simulated reflection coefficient versus frequency plot of presented antenna1, antenna2, and antenna3. The optimized antenna is resonating at 7 GHz and gives the -31 dB return loss at resonating frequency. A comparison of the simulated performance of the return loss of three antennas is presented here. The reflection coefficient versus frequency plot in the figure below clearly shows that with the increase in the slots in the primary design the return loss increases considerably in the negative direction.

Fig. 3 Return loss versus frequency of antenna1, antenna2, and antenna3



4 Fabricated Proposed Antenna and Experimental Results

Figure 4 shows the top and bottom views of the fabricated antenna3. The proposed antenna is fabricated using the FR4 as substrate and copper as the conductive patch of the antenna. The fabricated antenna is tested and the results are recorded. The recorded results are plotted on the graph which shows quite compromising results with the simulated one. The results of fabricated antenna show the maximum return loss near 7 GHz frequency which is -18.26 dB. The comparison of the return loss of both the measured and the simulated antenna3 is shown in Fig. 5. Figure 5 shows that the proposed antenna works as expected as both the simulated and the fabricated antennas resonate near the given resonating frequency (7 GHz) and the desired results are obtained below the -10 dB, which shows that the antenna has the good directivity. Figure 6 shows the comparison of axial ratio versus frequency of simulated and measured results of proposed antenna3.

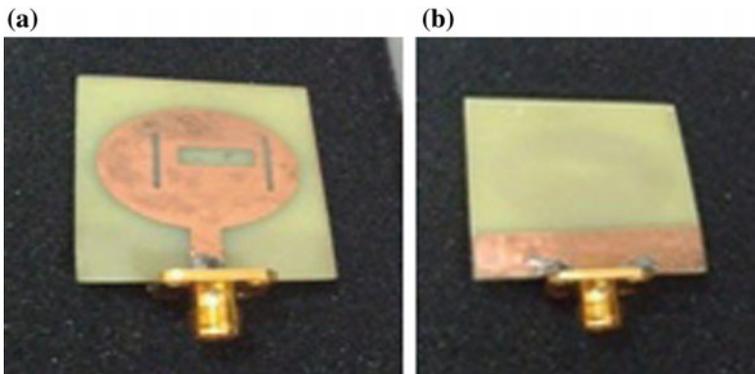


Fig. 4 Hardware of proposed antenna **a** top view **b** bottom view

Fig. 5 Comparison of return loss versus frequency plot of the simulated and measured results of proposed antenna3

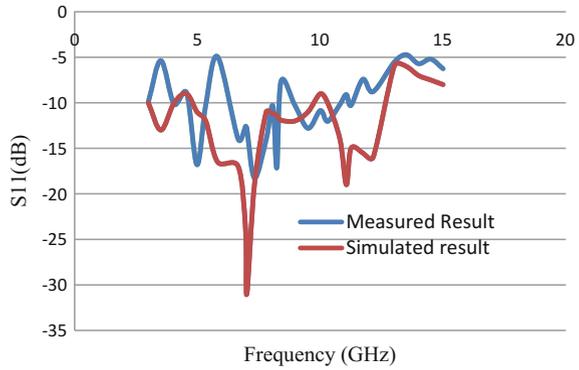
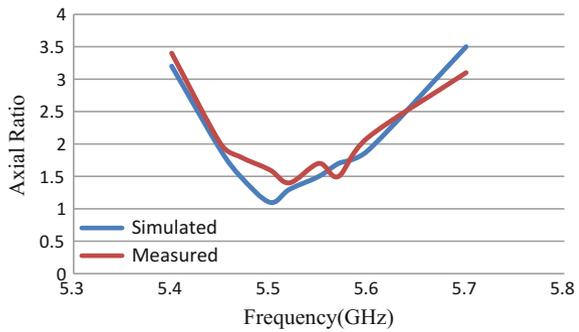


Fig. 6 Comparison of axial ratio versus frequency of simulated and measured results of proposed antenna3



5 Summary

A microstrip patch antenna has been designed, simulated, and fabricated for the circular polarization using HFSS software. The antenna resonates at the frequency of 7 GHz as per wireless communication standard. The corresponding simulated value of reflection coefficient is -31 dB which suggests that there is good impedance matching at the frequency point below the -10 dB region. The antenna also shows quite good axial ratio and reflection coefficient. The axial ratio approximately equal to 1 clearly shows that antenna is circularly polarized one. The proposed antenna is compact in size and simple in design, and also finds application for satellite communication.

References

1. Ansari, J.A., Verma, S., Verma, M.K., Agrawal, N.: Wide band microstrip-line-fed antenna with defected ground for CP operation. *Prog. Electromagn. Res. C* **58**, 169–181 (2015)
2. Raghupatruni, V., Krishna, R., Kumar, R.: Design of temple shape slot antenna for ultra wideband applications. *Prog. Electromagn. Res. B* **47**, 405–421 (2013)
3. Balanis, C.A.: *Antenna Theory: Analysis and Design*. Wiley, New York (2004)
4. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: A wide band compact microstrip antenna for GPS/DCS/PCS/WLAN applications. *Intell. Comput. Networking Inf.* **243**, 1107–1113 (2014). (ISBN: 978-81-322-1664-3)
5. Srivastava, R., Singh, V.K., Ayub, S.: Comparative analysis and bandwidth enhancement with direct coupled C slotted microstrip antenna for dual wide band applications. *Frontiers Intell. Comput. Theory Appl.* **328**, 449–455 (2015). (ISBN: 978-3-319-12011-9)
6. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: Bandwidth optimization of compact microstrip antenna for PCS/DCS/Bluetooth application. *Cent. Eur. J. Eng.* **4**(3), 281–286 (2014)
7. Singh, N., Singh, A.K., Singh, V.K.: Design & performance of wearable ultra wide band textile antenna for medical applications. *Microw. Opt. Technol. Lett.* **57**(7), 1553–1557 (2015)
8. Srivastava, S., Singh, V.K., Ali, Z., Singh, A.K.: Duo triangle shaped microstrip patch antenna analysis for WiMAX lower band application. In: *International Conference on Computational Intelligence: Modelling Techniques and Applications (CIMTA-2013)*, *Procedia Technology Elsevier* **10**, pp. 554–563 (2013)
9. Singh, N.K., Singh, V.K., Naresh, B.: Textile antenna for microwave wireless power transmission. In: *International Conference on Computational Modelling and Security (CMS 2016)*, *Procedia Computer Science*, vol. 85, pp. 856–861 (2016)
10. Gupta, N., Singh, V.K., Ali, Z., Ahirwar, J.: Stacked textile antenna for multi band application using foam substrate. In: *International Conference on Computational Modelling and Security (CMS 2016)*, *Procedia Computer Science*, vol. 85, pp. 871–877 (2016)
11. Rawat, A.K., Singh, V.K., Ayub, S.: Compact wide band microstrip antenna for GPS/WLAN/WiMax applications. *Int. J. Emerg. Trends Eng. Dev.* **7**(2), 140–145 (2012)
12. Baudh, R.K., Kumar, R., Singh, V.K.: Arrow shape microstrip patch antenna for WiMax application. *J. Environ. Sci. Comput. Sci. Eng. Technol.* **3**(1), 269–274 (2013)
13. Loni, J., Ayub, S., Singh, V.K.: Performance analysis of microstrip patch antenna by varying slot size for UMTS application. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 01–05 (2014)
14. Srivastava, R., Ayub, S., Singh, V.K.: Dual band rectangular and circular slot loaded microstrip antenna for WLAN/GPS/WiMax applications. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 45–48 (2014)
15. Dhupkariya, S., Singh, V.K.: Textile antenna for C-band satellite communication application. *J. Telecommun. Switch. Syst. Netw.* **2**(2), 20–25 (2015)
16. Loni, J., Singh, V.K.: Development of bandwidth enhanced microstrip patch antenna for UMTS application. *J. Microw. Eng. Technol.* **2**(1), 01–07 (2015)
17. Singh, Vinod Kumar, Naresh, B.: Multi resonant microstrip antenna with partial ground for radar application. *J. Telecommun. Switch. Syst. Netw.* **2**(1), 01–05 (2015)
18. Singh, M., Singh, V.K., Naresh, B.: Rectangular slot loaded circular patch antenna for WLAN application. *J. Telecommun. Switch. Syst. Netw.* **2**(1), 07–10 (2015)
19. Din, N.M., Chakrabarty, C.K., Bin Ismail, A., Devi, K.K.A., Chen, W.Y.: Design of RF energy harvesting system for energizing low power devices. *Prog. Electromagn. Res.* **132**, 49–69 (2012)
20. Osman, M.A.R., Rahim, M.K.A., Azfar, M., Samsuri, N.A., Zubir, F., Kamardin, K.: Design, implementation and performance of ultra-wideband textile antenna. *Prog. Electromagn. Res. B* **27**, 307–325 (2011)
21. Singh, V.K., Singh, N.K.: Compact circular slotted microstrip antenna for wireless communication systems. *J. Microw. Eng. Technol.* **1**(1), 07–14 (2015)

Weighted Wavelet Tree Sparsity Regularization for Compressed Sensing Magnetic Resonance Image Reconstruction

Bhabesh Deka and Sumit Datta

Abstract Compressed sensing in magnetic resonance imaging (CS-MRI) improves the MRI scan time by acquiring only a few k -space samples and then reconstructs the image using a nonlinear procedure from the highly undersampled measurements. Besides the standard wavelet sparsity, MR images are also found to exhibit tree sparsity across various scales of the wavelet decomposition which are generally modeled as overlapping group sparsity. In this chapter, we propose a novel iteratively weighted wavelet tree sparsity based CS-MRI reconstruction technique to estimate MR images from highly undersampled Fourier measurements. Simulations on various real MR images show that the proposed technique offers significant improvements compared to the state-of-the-art either in terms of visual quality or k -space measurements with the same reconstruction time.

Keywords Magnetic resonance imaging · Undersampled Fourier measurements
Compressed sensing · Wavelet tree sparsity · Overlapping group sparsity
Weighted CS-MRI reconstruction

1 Introduction

Compressed sensing or compressive sampling (CS) [1] is a modern signal processing approach for simultaneous acquisition and reconstruction of a signal from a few linear measurements by solving an under-determined system of equations. The number of such linear measurements is very less than that required by the traditional Shannon–Nyquist sampling rate provided the signal to be reconstructed is sparse or at least compressible in some known transform domain and the acquisition scheme

B. Deka (✉) · S. Datta

Computer Vision and Image Processing Laboratory, Department of Electronics
and Communication Engineering, Tezpur University, Tezpur 784028, Assam, India
e-mail: bdeka@tezu.ernet.in

S. Datta

e-mail: sumit89@tezu.ernet.in

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication
and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_48

or sensing matrix is incoherent with respect to the sparse representation basis. There are some commonly encountered occasions where data acquisition is very expensive, either for limitations of number of data acquisition sensors or for the sensing process being very slow, like, in the magnetic resonance imaging (MRI).

MRI is one of the most preferred medical imaging modalities for soft tissues like, the brain, and the heart. It provides good contrast and high-resolution images without using any ionizing radiation. But slow data acquisition speed is a major burden in conventional MRI. Moreover, some instrumental and physiological processes also limit the speed of MR image acquisition. For example, in cardiac imaging, one needs to hold the breath several times at 20–30 s intervals for the acquisition of images with diagnostic resolution. Many people, especially children and those having heart and lungs ailments, cannot hold the breath for a long period [2, Chap. 4]. Also, during this long scanning time, any movement of the body results in non-recoverable imaging artifacts. Thus, radiologists have to use sedation or general anesthesia for pediatric, cardiac patients. In the quest for a practical alternative, researchers have turned to the exciting and emerging field of compressed sensing MRI (CS-MRI) to reduce the scanning time without sacrificing the diagnostic imaging quality.

Lustig et al. [3] demonstrated that it is possible to reconstruct good quality MR images from a fewer random measurements because it satisfies the two key requirements of the CS theory, i.e., the *sparsity* and the *incoherence* [1]. It is because MR images are generally compressible in the transform domain, like the wavelet. Additionally, MRI data acquisition is performed in the Fourier domain which is incoherent with respect to the wavelet domain. Mathematically, consider $\mathbf{x} \in \mathcal{R}^n$ to be an array representing the MR image and $\mathbf{F}_u \in \mathcal{R}^{m \times n}$ the partial Fourier operator which maps \mathbf{x} to the undersampled k -space data, $\mathbf{y} \in \mathcal{R}^m$ i.e. $\mathbf{y} = \mathbf{F}_u \mathbf{x}$, where $m \ll n$. Now, to estimate \mathbf{x} from \mathbf{y} , they proposed the following TV – ℓ_1 – ℓ_2 model of CS-MRI reconstruction:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{F}_u \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1 \|\Psi \mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_{\text{TV}}, \quad (1)$$

where Ψ represents the standard wavelet basis set; λ_1 and λ_2 being regularization parameters that establish balance between the data consistency term and the sparsity terms. Here, the term $\|\mathbf{x}\|_{\text{TV}}$ represents the total variation of \mathbf{x} in isotropic sense, i.e.,

$\|\mathbf{x}\|_{\text{TV}} = \sum_{i,j} \sqrt{\{(\nabla_h \mathbf{x})_{ij}\}^2 + \{(\nabla_v \mathbf{x})_{ij}\}^2}$, where ∇_h and ∇_v are first-order horizontal and vertical finite difference operators, respectively. The above minimization problem is solved using the *nonlinear conjugate gradient* (NCG) algorithm.

Huang et al. [4] proposed to decompose the problem in Eq. (1) into two relatively simpler subproblems, one is the TV regularization and the other is the ℓ_1 -norm regularization. These subproblems are independently solved and individual solutions are then linearly combined to get the solution of the composite problem. They called it as the *Fast Composite Splitting Algorithm* (FCSA). This technique significantly improves both the reconstruction time as well as quality of the reconstructed image compared to previous techniques.

Recently, Chan and Huang [5] proposed the *wavelet tree sparsity MRI* (WaTMRI) CS reconstruction model where another sparsity promoting regularization term is added with the existing TV – ℓ_1 – ℓ_2 model, especially, to exploit the wavelet quadtree structure [5, Fig. 1]. It means that if a parent coefficient has a significant value then its children also tend to have significant values. To exploit this quadtree structure they assign pairs of parent–child wavelet coefficients into different overlapping groups. This particular arrangement not only improves the reconstruction quality but also reduces k -space measurements from $O(k + \log(n/k))$ to $O(k + k \log(n/k))$ for a k -sparse signal of length n for the same reconstruction quality. This reconstruction model can be represented as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{F}_u \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{x}\|_{\text{TV}} + \lambda_2 \|\Psi \mathbf{x}\|_1 + \lambda_2 \sum_{g \in \mathbf{G}} \|(\Psi \mathbf{x})_g\|_2, \quad (2)$$

where \mathbf{G} is a binary matrix containing overlapping indices of all parent–child groups. Equation (2) is decomposed into three relatively simpler subproblems dealing with (a) the tree sparsity norm, (b) the TV norm, and (c) the ℓ_1 -norm regularizations. Mathematically, the tree sparsity norm is defined by the $\ell_{2,1}$ -norm, i.e., first ℓ_2 -norm is obtained for each parent–child pair and then individual results of all such pairs are used to obtain the ℓ_1 -norm. This arrangement forces each individual member of the group either to zeros or nonzeros during the iterative thresholding process. Further, [6] experimentally demonstrates that among different possible overlapping group configurations the one having only single parent–child pair gives the best results in terms of both computational complexity and quality of reconstruction. However, in Eq. (2), as the tree sparsity already computes the ℓ_1 -norm of all individual parent–child groups from the wavelet decomposition, so having an additional standard wavelet sparsity regularization term seems to be an overhead although their setting is shown to outperform the existing TV – ℓ_1 – ℓ_2 model-based CS-MRI reconstruction. Therefore, in order to overcome the computational burden and also to mimic the joint effect of both the tree sparsity and the standard sparsity of wavelet coefficients, it seems to be somewhat logical to have only the tree sparsity term. It would definitely reduce the computational complexity of the reconstruction problem as it requires solving only two subproblems instead of three.

In this chapter, we propose a novel iteratively weighted tree sparsity based CS-MRI reconstruction technique having only two regularization terms, namely, the weighted tree sparsity and the weighted TV norm. Reconstruction results show significant improvements over state-of-the-art, either in terms of quality of reconstruction or number of k -space measurements. The remaining part of the chapter is divided into the following sections. Section 2 discusses on the related theory of the present work. Section 3 describes the proposed method. Section 4 gives details of various simulation results and their analysis. Section 5 gives a brief account of the conclusions drawn from the present research work.

2 Theory

Candes et al. [7] proposed the idea of reweighted ℓ_1 -minimization. For sparse representation using the ℓ_1 -norm minimization, larger coefficients are relatively heavily penalized as compared to the smaller coefficients. But ℓ_0 -norm minimization is independent of the magnitudes of sparse coefficients and computes only the total number of nonzero coefficients. To solve the nonuniform penalty issue in the ℓ_1 -minimization, they proposed the weighted ℓ_1 -norm minimization model. The weighted ℓ_1 -norm minimization problem for a sparse signal \mathbf{x} acquired by the measurement matrix \mathbf{A} may be defined as follows:

$$\begin{aligned} \min_{\mathbf{x}} \sum_{i=1}^n w_i |x_i| \\ \text{subject to } \mathbf{y} = \mathbf{A}\mathbf{x}, \end{aligned} \quad (3)$$

where w_1, w_2, \dots, w_n are positive weights and are obtained as follows:

$$w_i = \frac{1}{x_i + \varepsilon}. \quad (4)$$

Here, ε is a small positive constant which ensures nonzero estimates of weights when x_i is zero. It was demonstrated that weighted ℓ_1 -norm minimization gives significant improvements in terms of quality of reconstruction or number of CS measurements.

3 Proposed Work

In this chapter, we propose a novel iteratively weighted tree sparsity based MR image reconstruction technique. We call it as the *Iteratively Weighted Tree sparsity MRI* (IWTMRI) algorithm. It uses two regularization terms consisting of (a) the tree sparsity norm and (b) the TV norm. These regularization terms are appropriately weighted to solve the CS-MRI reconstruction problem. In our formulation, anisotropic TV norm is used, i.e., $\|\mathbf{x}\|_{\text{TV}} = \|\nabla\mathbf{x}\|_1 = \sum_{ij} |(\nabla_{\text{h}\mathbf{x}})_{ij}| + |(\nabla_{\text{v}\mathbf{x}})_{ij}|$, where $\nabla = (\nabla_{\text{h}}; \nabla_{\text{v}})$. The proposed weighted TV-tree sparsity based regularization problem can be written as follows:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{F}_u \mathbf{x} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{W}_{\text{TV}} \nabla \mathbf{x}\|_1 + \lambda_2 \sum_{g \in \mathbf{G}} \mathbf{W}_G \|(\Psi \mathbf{x})_g\|_2, \quad (5)$$

where \mathbf{W}_{TV} and \mathbf{W}_G are two weighting matrices corresponding to the TV and the wavelet tree sparsity terms, respectively. \mathbf{W}_{TV} contains two diagonal matrices, i.e., $\mathbf{W}_{\text{TV}} = [\mathbf{W}_{\text{TV}}^{\text{h}}; \mathbf{W}_{\text{TV}}^{\text{v}}]$ and nonzero elements of $\mathbf{W}_{\text{TV}}^{\text{h}}$ and $\mathbf{W}_{\text{TV}}^{\text{v}}$ at any location (i, i) ,

respectively, contain first-order horizontal and vertical finite differences of \mathbf{x} at that location. On the other hand, \mathbf{W}_G is a diagonal matrix where nonzero elements consist of ℓ_2 -norms of overlapping groups. We define them as follows:

$$(\mathbf{W}_{\text{TV}})_{i,j}^{(k+1)} = \frac{1}{|(\nabla \mathbf{x})_{i,j}^{(k)}| + \varepsilon_{\text{TV}}} \quad (6)$$

$$(\mathbf{W}_G)_{g \in \mathbf{G}}^{(k+1)} = \frac{1}{|(\Psi \mathbf{x})_g^{(k)}| + \varepsilon_G}, \quad (7)$$

where ε_{TV} and ε_G are small positive parameters.

Algorithm 1: Iteratively Weighted Tree – sparsity MRI (IWTMRI)

Input: $\Psi, \mathbf{F}_u, \mathbf{y}, \lambda_1, \lambda_2$

Initialization: $\{\mathbf{x}^{(0)}, \mathbf{r}^{(1)}\} \leftarrow \mathbf{F}_u^T \mathbf{y}, \{\mathbf{W}_{\text{TV}}^{h(1)}, \mathbf{W}_{\text{TV}}^{v(1)}, \mathbf{W}_G^{(1)}\} \leftarrow \mathbf{I}, \rho \leftarrow \frac{1}{L}, t^{(1)} \leftarrow 1, k \leftarrow 1$

1. **While** not converged **do**

2. $\mathbf{z}^{(k)} \leftarrow \text{shrinkgroup}\left(\mathbf{G}\Psi\mathbf{x}^{(k-1)}, \frac{\lambda_2 \mathbf{W}_G^{(k)}}{\beta}\right)$

3. $\mathbf{x}_g^{(k)} \leftarrow \mathbf{r}^{(k)} - \rho\left(\mathbf{F}_u^T(\mathbf{F}_u \mathbf{r}^{(k)} - \mathbf{y}) + \beta \Psi^T \mathbf{G}^T(\mathbf{G}\Psi\mathbf{x}^{(k-1)} - \mathbf{z}^{(k)})\right)$

4. $\mathbf{x}^{(k)} \leftarrow \text{prox}_\rho\left(2\lambda_1 \|\mathbf{W}_{\text{TV}}^{(k)} \nabla \mathbf{x}^{(k-1)}\|_1\right)(\mathbf{x}_g^{(k)})$

5. $\mathbf{x}^{(k)} \leftarrow \text{project}\left(\mathbf{x}^{(k)}, [l, u]\right)$

6. $(\mathbf{W}_{\text{TV}})_{i,j}^{(k+1)} \leftarrow \frac{1}{|(\nabla \mathbf{x})_{i,j}^{(k)}| + \varepsilon_{\text{TV}}}$

7. $(\mathbf{W}_G)_{g \in \mathbf{G}}^{(k+1)} \leftarrow \frac{1}{|(\Psi \mathbf{x})_g^{(k)}| + \varepsilon_G}$

8. $t^{(k+1)} \leftarrow \frac{\sqrt{1+4(t^{(k)})^2}}{2}$

9. $\mathbf{r}^{(k+1)} \leftarrow \mathbf{x}^{(k)} + \frac{t^{(k)} - 1}{t^{(k+1)}}(\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})$

10. $k \leftarrow k + 1$

11. **End While**

Output: $\mathbf{x}^* \leftarrow \mathbf{x}^{(k)}$

Detailed steps of the proposed algorithm are given in Algorithm 1. We first decompose the composite problem into two relatively simpler weighted subproblems. One is the weighted tree sparsity regularization subproblem and other is the weighted TV regularization subproblem. Both these problems are independently solved to get the solution of the composite problem. For solving the first subproblem, we apply the concept of variable splitting and introduce the variable \mathbf{z} which constrains \mathbf{x} . It yields a closed form solution given by

$$z_{g_i} = \max\left(\|(\mathbf{G}\Psi\mathbf{x})_{g_i}\|_2 - \frac{\lambda_2(\mathbf{W}_G)_{g_i}}{\beta}, 0\right) \frac{(\mathbf{G}\Psi\mathbf{x})_{g_i}}{\|(\mathbf{G}\Psi\mathbf{x})_{g_i}\|_2}, \quad i = 1, 2, \dots, s, \quad (8)$$

where g_i denotes the i th group and s counts the number of total groups. We denote Eq. (8) by $\text{shrinkgroup}(\cdot)$ as in [5] and include it in step 2 of the algorithm. Steps 3–5 of the algorithm show the process of solving the second subproblem consisting of the TV regularization term and two data fidelity terms in terms of \mathbf{x} and \mathbf{z} . These steps are similar to the iterative scheme of the FCSA [4] containing the proximal map function defined by

$$\text{prox}_\rho(\phi)(u) = \underset{\mathbf{x}}{\text{argmin}} \left\{ \phi(\mathbf{x}) + \frac{1}{2\rho} \|\mathbf{x} - \mathbf{u}\|_2^2 \right\}, \quad (9)$$

which is nothing but element-wise shrinkage operation carried out by the following function:

$$x_i = \max(|u_i| - \rho, 0) \frac{u_i}{|u_i|}, \quad (10)$$

and the project function:

$$\mathbf{x}_i = \text{project}(\mathbf{x}_i, [l, u]) = \begin{cases} \mathbf{x}_i, & \text{if } l \leq \mathbf{x}_i \leq u \\ l, & \text{if } \mathbf{x}_i < l \\ u, & \text{if } \mathbf{x}_i > u \end{cases}, \quad (11)$$

where l and u denote the range of pixel values in the MR image. Finally, steps 8 and 9 of the algorithm provide necessary acceleration for convergence of the algorithm similar to the FCSA.

4 Experimental Results

All simulations in this chapter are carried out in the MATLAB (R2012b) environment on a PC having 3.4 GHz Intel *i7* core CPU with 4 GB of RAM. For fair comparisons, we use a common stopping criterion for all algorithms. To evaluate performances of various algorithms, we carry out simulations on several real MR images. We collected a set of 3D BRAVO T1 h brain MR volume from a local Hospital repository. This data set is acquired using a GE 1.5 T signa HDxt scanner with the following parameters: TR/TE: 14.912/6.456 ms, slice thickness: 1.2 mm, spacing between scans: 1.2 mm, sampling (%): 100, and Flip angle: 15 deg. Another RF spoiled gradient echo (GE) abdomen MRI data set is collected from the publicly available database at <http://mridata.org>. We also perform simulations on a real brain MR image collected from [5].

To evaluate performance of the proposed IWTMRI algorithm, we perform three sets of experiments—(a) Evaluation of reconstruction quality in terms of SNR (dB) and MSSIM with varying sampling ratio 15–25%; (b) Test of convergence in terms CPU Time and number of iterations; and (c) Computation of number of measurements to achieve the same quality of reconstruction. We compare results of the IWTMRI with the FCSA and the WaTMRI as they are the most recently developed and computationally faster algorithms for CS-MRI reconstruction.

Experiment 1: Reconstruction performances of various algorithms in terms of SNR and MSSIM at different sampling ratios are shown in Fig. 1. We observe that on an average SNR improvement of approximately 1.5 dB is achieved by the IWTMRI compared to the WaTMRI at varying sampling ratios. This improvement would be still higher when compared to the FCSA. Similarly, for visual interpretation, cropped portions of reconstructed outputs using real brain and abdomen are shown in Fig. 2. They clearly indicate that the IWTMRI outperforms other algorithms in terms of preservation of edges and anatomical details, and less visible aliasing artifacts than others.

Experiment 2: To compute the computational complexity of the proposed algorithm for convergence, we measure the relative error of various algorithms with the progress of iterations and the CPU time taken for reconstruction. Results of various algorithms for brain and abdomen MR images at 20% sampling ratio are shown in Fig. 3. From the figure we observe that the IWTMRI requires less number of iterations and CPU time as compared to the FCSA and the WaTMRI. In case of the abdomen, IWTMRI requires 21 iterations whereas the WaTMRI and the FCSA require 67 and 102 iterations, respectively, for convergence. Similarly, IWTMRI requires 2.3 s, whereas the WaTMRI and the FCSA require 4.7 and 5.1 s, respectively, for convergence.

Experiment 3: Finally, we show reduction in the number of CS measurements for the same quality of reconstruction using different methods. For that, we select the SNR of the FCSA as the reference and then try to reconstruct the same quality of images with the WaTMRI and the IWTMRI. From Table 1 we observe that number of measurements required for IWTMRI is significantly less as compared to the FCSA and the WaTMRI.

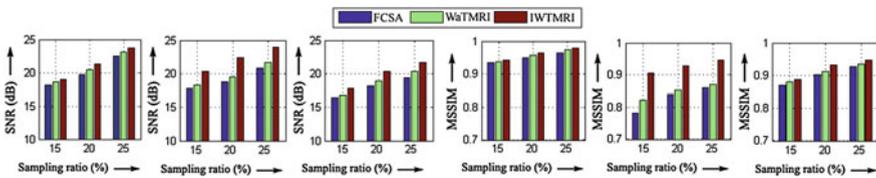


Fig. 1 SNR (in dB) and MSSIM values at different sampling ratios for brain and abdomen MR images, respectively

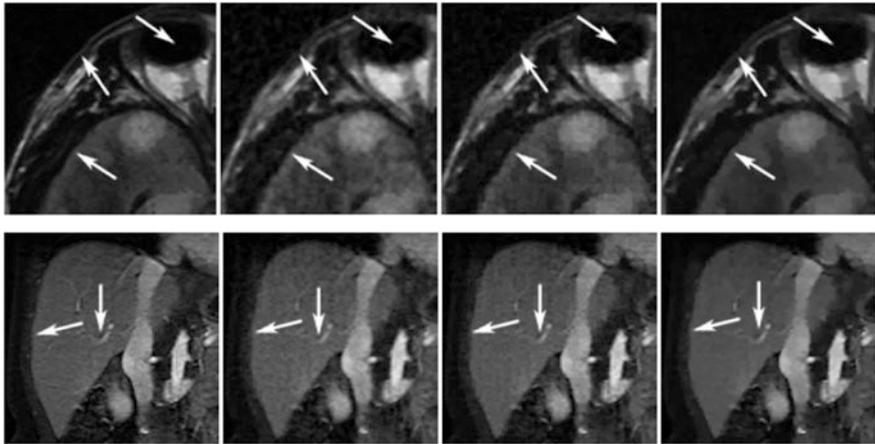


Fig. 2 First row left to right Cropped portion of a brain image and results using the FCSA, the WaTMRI, and IWTMRI at 20% sampling ratio. Second row left to right Cropped portion of an abdomen image and results using the FCSA, the WaTMRI, and the IWTMRI at 20% sampling ratio

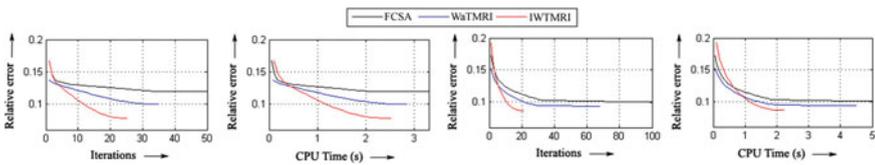


Fig. 3 Relative error versus iteration and relative error versus CPU time (s) for brain and abdomen MR images, respectively

Table 1 Comparison of percentage k -space under sampling ratios (% SR) of different methods for the abdomen MR image, Δm = Measurement Reduction

Target SNR (dB)	FCSA	WaTMRI		IWTMRI	
	SR (%)	SR (%)	Δm	SR (%)	Δm
16.46	15.0	14.5	328	14.0	655
18.11	20.0	18.5	983	17.5	1638
19.43	25.0	22.5	1638	19.0	3932

5 Conclusion

We have proposed a novel iteratively weighted MR image reconstruction technique to reconstruct MR images from highly undersampled measurements. It is observed that the proposed algorithm achieves significant improvements in terms of quality of reconstructions for the same number of measurements and vice versa compared to some of the most recent algorithms.

References

1. Candes, E.J., Romberg, J.K., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**, 489–509 (2006)
2. Lustig, M., Keutzer, K., Vasanawala, S.: Introduction to parallelizing compressed sensing magnetic resonance imaging. In: Patterson, D., Gannon, D., Wrinn, M. (eds) *The Berkeley Par Lab: progress in the parallel computing landscape*, Microsoft Corporation, pp. 105–139 (2013)
3. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007)
4. Huang, J., Zhang, S., Metaxas, D.N.: Efficient MR image reconstruction for compressed MR imaging. *Med. Image Anal.* **15**, 670–679 (2011)
5. Chen, C., Huang, J.: Exploiting the wavelet structure in compressed sensing MRI. *Magn. Reson. Imaging* **32**, 1377–1389 (2014)
6. Chen, C., Huang, J.: The benefit of tree sparsity in accelerated MRI. *Med. Image Anal.* **18**, 834–842 (2014)
7. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted L1 minimization. *J. Fourier Anal. Appl.* **14**, 877–905 (2008)

A Score-Level Solution to Speaker Verification Using UBM Pooling and Adaptive Cohort Selection

Pranab Das

Abstract In an environment that is highly unpredictable in nature, a speaker verification system needs a good background model to carry out the verification task reliably. In this paper, a 1024-component UBM is created by pooling a noisy speech UBM and clean speech UBM. This pooled UBM is used for speaker adaptation as well as for speaker testing. Experimental results have shown minor improvement with pooled UBM as compared to baseline UBM. In addition to this, a score-level solution is proposed by means of cohort model selection using HT-normalization to reduce undesirable variation arising from acoustically mismatched devices and environment. For cohort selection a simple distance metric based on similarity modeling of each client speaker is used. The normalization parameters computed over a group of speakers (cohort) having some common characteristics are used in the final score calculation. Experiments on a noisy corpus has shown reasonable improvements in performance, when normalization parameters were taken from a cohort than from a general group. Experiments have shown a recognition rate of 90.58 and 87.64% for matched handset type in office and roadside environment respectively.

Keywords Score normalization · Cohort · T-normalization · UBM pooling
GMM-UBM

1 Introduction

In general, a speaker verification system models the alternative hypothesis by a universe of probable background imposter speakers for optimal likelihood ratio test. Such a universe of background imposter speakers is called a Universal Background Model (UBM) [1] or a World Model. The UBM plays vital role in the verification task by minimizing the non-speaker-related variations and thus helps in a stable

P. Das (✉)

Assam Don Bosco University, Guwahati, India
e-mail: pranab.das@dbuniversity.ac.in

decision. One of the major approaches in modeling a UBM is to collect speech from a number of speaker's representative of the population of speakers. A single model is then trained from that population. The focus of this approach is mainly on the composition, selection of the speakers, and the speech used to train the UBM [2, 3].

To improve the effectiveness of detection threshold in a speaker verification system, a technique called score normalization is commonly used. In this technique the output scores are transformed by aligning with respect to the score distribution of individual speaker models. Score normalization is also minimizes the speaker-dependent and speaker-independent changes in the signal. For example, Z-norm minimizes speaker differences of imposter scores distributions, whereas H-norm minimizes bias effects resulting from different microphones and channels.

One of the most widely used methods for score normalization at the time of testing is Test-normalization or T-norm [4]. In this method, a test utterance is scored against a group of imposter speaker models also called T-norm speaker model to obtain a set of imposter score distribution. The test utterance is also scored against the target speaker model. This score distribution so obtained from the T-norm speaker models is used to compute the parameters mean and standard deviation. These normalization parameters are used during testing and help in reducing deviation in imposter scores and session difference between the test utterance and the imposter utterance. These normalization parameters are subsequently used to fine-tune the target speaker score. In a speaker verification system with adaptive UBM modeling, T-norm is found to be effective as the adapted UBM provides fast scoring [5] with little additional computation by scoring a large set of T-norm speaker models.

It has been proposed in [6–10] that score normalization using speaker-dependent cohort speakers set is much more effective in a speaker verification system. While using cohort a number of issues need to be considered like cohort selection technique, number of imposters in the cohort set, and the technique used for score normalization. Although both online and offline methods of cohort selection using T-norm have been used, in general T-norm-based cohort selection uses speaker's characteristic like gender and distance metric [11].

2 UBM Pooling

A 1024-component UBM is created by pooling two 512-component UBMs. The first 512-component UBM is built using only Office and Roadside Noise. The second 512-component UBM is built using clean data from NIST 2003 SRE only. These two 512-component UBMs are pooled to create a new UBM. Figure 1 shows the above approach. All the speaker models are created by adapting the training data with respect to 1024-component POOLED UBM. All the speakers are tested in noisy office environment and noisy roadside environment with respect to the clean UBM, Office noise UBM, Roadside noise UBM, and Pooled UBM. Table 1 shows performance of the system measured in equal error rate with different UBMs.

Fig. 1 A new UBM built using UBM pooling method

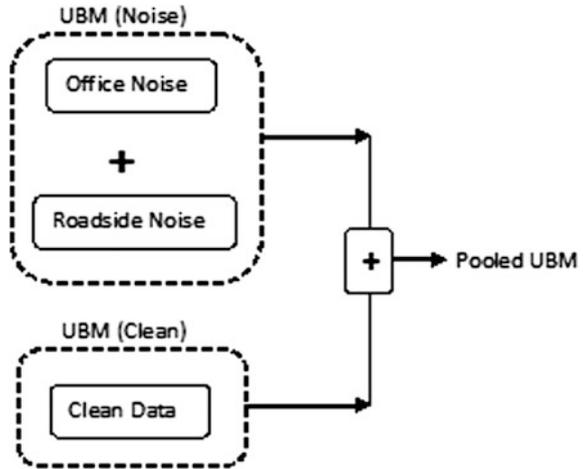


Table 1 Equal error rate for speaker verification system trained with different UBM

UBM model	EER (%)	Recognition rate (%)
Clean (UBM)	28.83	71.17
Office noise (UBM)	24.12	75.88
Roadside noise (UBM)	26.74	73.26
Pooled UBM (office noise and roadside noise + clean data)	18.06	81.94

3 Score Normalization

In a GMM-UBM based speaker verification system, the decision to accept or reject a claimed identity is achieved through a log likelihood ratio (LLR). This ratio is computed for an unknown test utterance X_{Test} between the claimed speaker S and the speaker-independent background model or universal background model. Score normalization is used to improve the detection threshold of the verification system by normalizing the log likelihood ratio score for a test utterance X and target model S as

$$LLR(X_{Test}, S) = \frac{LLR(X_{Test}, S) - \mu}{\sigma}$$

where the mean (μ) and the standard deviation (σ) are computed from imposter models.

Although there are many score normalization techniques like Z-Norm, H-Norm, T-Norm, etc. that can be applied with either speaker model or with test data. T-Norm is one of the most popular and efficient method, where parameters for normalizing the detection score are estimated from a group of imposter speaker

models during testing. Therefore, a method is proposed for cohort set selection for each speaker using T-Norm. The normalization parameters selected from each cohort set is used for final score normalization.

4 Cohort Model Selection Using HT-Normalization

As marginal improvement is observed in speaker verification system with UBM pooling in Sect. 2, so as a next step toward improving system performance is by cohort selection using HT-Norm. In this experiment, cohorts are selected for each speaker, the selection criteria are based on three factors.

- Speakers are from the same environment (e.g., road, office).
- Speakers are using same handset type.
- The distance score of (test data, imposter model) and (test data, target model) is very much small.

It has been observed that when normalization parameters, i.e., mean and standard deviation are taken from a group of speakers (cohort) having some common characteristics the performance of the system is improved compared to when normalization parameters are taken from a general group. So, an algorithm is proposed using HT-norm for cohort selection.

4.1 Proposed Methodology

- Step 1: All speakers are adapted with respect to the pooled UBM.
- Step 2: For each speaker, a set of imposter models having minimum distance score is selected which is called cohort for each speaker. Figure 2 shows the cohort selection process.
- Step 3: Cohort selection is done for each speaker and for each handset type.
- Step 4: During testing, the test data is used with the cohort (set of imposter models) to obtain a vector of scores. This is followed by computing normalization parameters. This step is done for each handset type.
- Step 5: The test data now is used with the target model to obtain a score. The normalization parameters obtained in step 4 is then used to calculate the final score.
- Step 6: The score obtained in step 5 is used against the decision threshold for speaker verification. Figure 3 elaborates the cohort selection and application of normalization parameters.

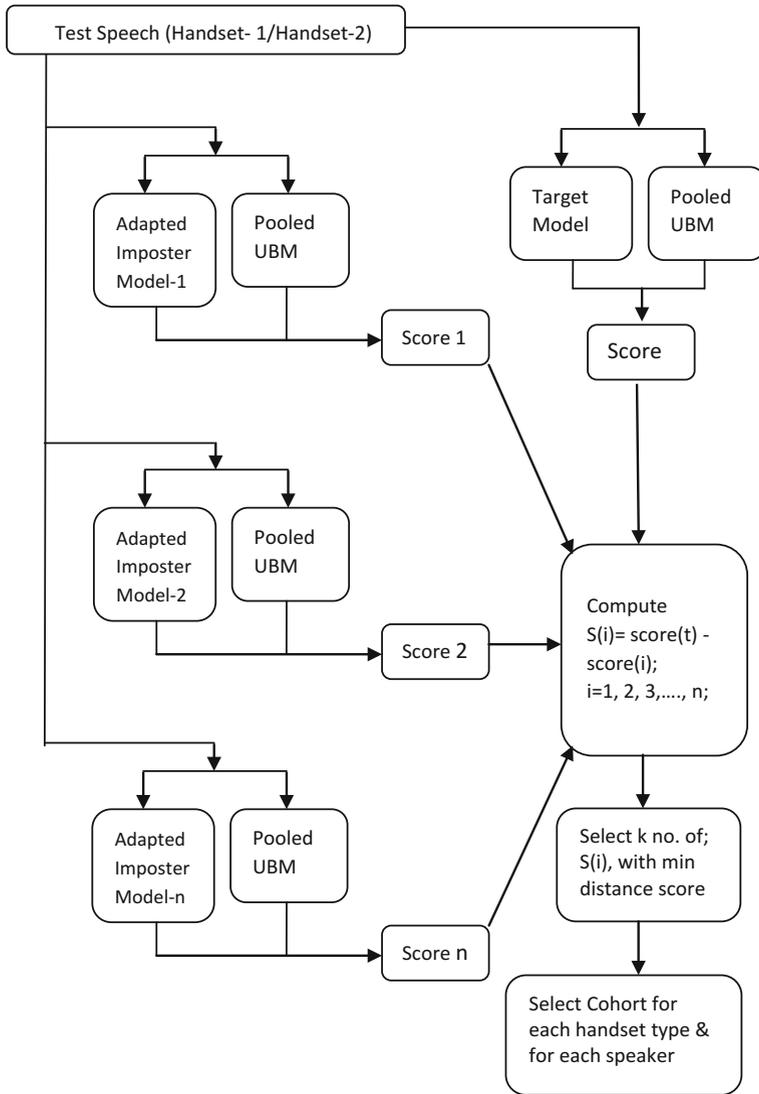


Fig. 2 Cohort selection process using T-norm for different handsets

5 Experimental Setup and Results

For the experiments, a noisy database (NS-DB) is created. The speaker verification system is trained and tested with respect to this database. NS-DB is a multisession, multichannel, and multi-environment database. It contains the speech samples from 50 speakers (25 male and 25 female) in two noisy environments namely office and

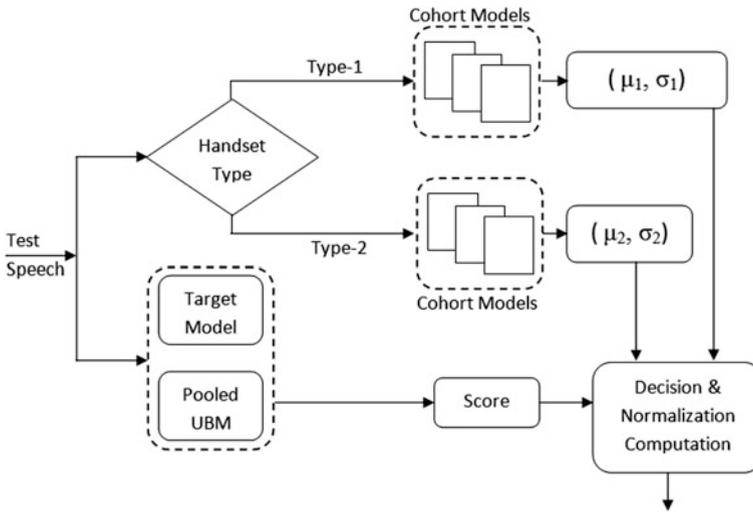


Fig. 3 Cohort selection and calculating normalization parameters

roadside. The speech samples have been collected from each of the environments in only English language. Speech data has been recorded in parallel across two recording devices namely Mobile Phone-1 and Mobile Phone-2. Recording has been done separately in office and roadside environment to obtain noisy speech. All the recordings have been done in 16 kHz sampling frequency in mono channel and 16-bit resolution. The speech materials have been down sampled to 8 kHz to keep the system compatible with telephone quality voice. To detect and eliminate silent frames within a speech utterance, an energy-based silence detector is used prior to feature extraction.

Also NIST SRE 2003 is a standard speech database of National Institute of Standard and Testing (NIST) for speaker recognition evaluation 2003 has been used in the present study for creating the clean UBM.

The verification system is built using efficient GMMs and universal background model (UBM) for imposter speaker representation. Bayesian adaptation is used to adapt speaker models with respect to the UBM. A 39-dimensional feature vector, made up of 13 mel-frequency cepstral coefficients along with its first-order and second-order derivatives is being used. To reduce the effect of channel mismatch Cepstral Mean Subtraction (CMS) is applied on feature vector prior to modeling.

In the first experiment, noisy office environment is considered and all the speaker models are adapted with respect to the pooled UBM. Similarly in the second experiment noisy roadside environment is considered and all the speaker models are adapted with respect to the pooled UBM. Cohorts are selected for each speaker using the method described in Sect. 4.1 and for each environment. Also it was ensured that the cohort's selection is done for each handset type. Tables 2 and 3 show equal error rate obtained for speaker verification system for both matched and

Table 2 EER for speaker verification system using HT-norm in office environment

UBM model	EER (%)	Recognition rate (%)
POOLED UBM (office noise and roadside noise + clean data)	18.06	81.94
Adaptive cohort selection using HT-norm (mismatched handset)	12.22	87.78
Adaptive cohort selection using HT-norm (matched handset)	9.42	90.58

Table 3 EER for speaker verification system using HT-norm in roadside environment

UBM model	EER (%)	Recognition rate (%)
POOLED UBM (office noise and roadside noise + clean data)	19.28	80.72
Adaptive cohort selection using HT-norm (mismatched handset)	14.67	85.33
Adaptive cohort selection using HT-norm (matched handset)	12.36	87.64

mismatched handset type in noisy office environment and noisy roadside environment.

It can be clearly observed from the results in Tables 2 and 3, recognition rate are 87.78 and 85.33% using adaptive cohort selection using HT-norm for mismatched handset type in office and roadside environment respectively. Further, it is observed that improvement of 90.58 and 87.64% in recognition rate for matched handset type in office and roadside environment respectively.

6 Conclusion

UBM Pooling and adaptive cohort selection using HT-Norm for score normalization has shown significant improvement in the performance of the speaker verification system. This improvement can be observed because Pooled UBM contains information for both clean and noisy data. Further when the normalization parameters mean and standard deviation are taken from a group of speakers (cohort) having some common characteristics recognition rate is higher than to normalization parameters taken from a general group.

References

1. Reynolds, D.A.: Comparison of background normalization methods for text-independent speaker verification. In: Proceedings of the European Conference on Speech Communication and Technology, pp. 963–967 (1997)
2. Matsui, T., Furui, S.: Likelihood normalization for speaker verification using a phoneme- and speaker-independent model. *Speech Commun.* **17**, 109–116 (1995)

3. Rosenberg, A.E., Parthasarathy, S.: Speaker background models for connected digit password speaker verification. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp. 81–84 (1996)
4. Auckenthaler, C., Thomas, L.: Score normalization for text-independent speaker verification systems. *Digit. Sig. Process.* **10**(1–3) (2000)
5. Reynolds, D.A.: Comparison of background normalization methods for text-independent speaker verification. *EUROSPEECH* (1997)
6. Higgins, A., Bahler, L., Porter, J.: Speaker verification using randomized phrase prompting. *Digit. Sig. Proc.* **1**, 89–106 (1991)
7. Rosenberg, A.E., DeLong, J., Lee, C.-H., Juang, B.-H., Soong, F.K.: The use of cohort normalized scores for speaker verification. *Proc. ICSLP* **92**, 599–602 (1992)
8. Tran, D., Wagner, M.: A proposed likelihood transformation for speaker verification. In: Proceedings of ICASSP 2000, Turkey (2000)
9. Reynolds, D.A.: Speaker identification and verification using Gaussian mixture speaker models. *Speech Commun.* **17**, 91–108 (1995)
10. Colombi, J.M., Reider, J.S., Campbell, J.P.: Allowing good impostors to test. In: Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers, 1997, pp. 296–300 (1998)
11. Reynolds, D.A., Rose, R.: Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Trans. Speech Audio Process.* **3**, 72–83 (1995)

A Novel Encryption Scheme for Secure SMS Communication

Aeneaus Malla, Pranab Sahu and Minati Mishra

Abstract Security is a challenging issue in today's digital world and for this, role of cryptography is significant. Cryptography secures data and information using an encryption algorithm and keys. There is always demand for a stronger encryption algorithm that would be very hard or impossible to crack without valid keys so that the privacy of intended sender and receiver information can be protected from hackers. This paper suggests a novel encryption scheme for SMS communication that uses multiple rounds of encryption to provide better privacy to the transmitted message against unauthorized access. Message integrity and authentication have been ensured by addition of authentication characters to the message. Keeping in mind the limited memory and processing capacity of mobile devices, the algorithm has been kept simpler and at the same time the strength of the scheme has been ensured with a stronger key and better mix of confusion and diffusion principles.

Keywords Playfair · Encryption · Cryptography · Key · Security

1 Introduction

In today's digital world security is a challenging issue. Cryptography plays a vital role in providing data security against attackers, so cryptographic algorithms are widely and frequently used as a means of ensuring that information is secure.

A. Malla · P. Sahu · M. Mishra (✉)
Department of Information & Communication Technology,
Fakir Mohan University, Balasore, Odisha, India
e-mail: minatiminu@gmail.com

A. Malla
e-mail: arian.aeneaus119@gmail.com

P. Sahu
e-mail: pranabsahu.bdps@gmail.com

The importance of cryptography is growing with the increasing worldwide utilization of the Internet. Some of the terminologies related to cryptography are mentioned below.

Cryptography has been derived from two Greek words and it means ‘the process of secret writing’. In this process, a ‘plain text’ message is converted or ‘encrypted’/ ‘enciphered’ to a coded message known as ‘cipher text’. There is always a demand for stronger encryption which is very hard to crack by an unauthorized user. The process of converting the readable ‘plain text’ to unintelligible ‘cipher text’ and the reverse is known as cryptography. ‘Cryptanalysis’, analyzes the strengths of a crypto scheme. Cryptography and cryptanalysis together constitute the domain ‘cryptology’ [1]. The overall process of encryption and decryption is given in Fig. 1.

1.1 Types of Cryptosystem

Depending on key used, cryptosystems are divided into three types; symmetric key cryptosystem, asymmetric key cryptosystem, and keyless cryptosystem. In Symmetric cryptosystem both encryption and decryption keys are same whereas in asymmetric cryptosystem a pair of keys, public and private are used. Public key is announced to all members while the private key is only known to the receiver. Sender uses the public key for encryption and the receiver decrypts the message using the private key. Symmetric key methods require smaller keys and hence are faster, stronger and easier for implementation but require a secure channel for key transmission. This makes these cryptosystems less practical in comparison to the public key systems. In keyless cryptosystem, there will be no key for encryption and decryption [2, 3].

This paper suggests a scheme for SMS encryption which can ensure that only the sender and receiver read the SMS but no third person; not even the service provider. Though the core of the algorithm, proposed in this paper, is a 9×10 playfair but, the original playfair technique has been modified with several shift and round functions to ensure higher degree of transposition. The paper also emphasizes on stronger keys. The rest of the paper has been organized as follows.

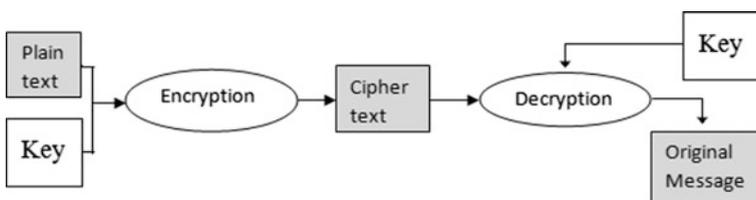


Fig. 1 Process of encryption and decryption

SMS working principle is discussed in Sect. 2, literature survey on SMS encryption schemes in Sect. 3, proposed method, results and discussion and conclusion in Sects. 4, 5 and 6 respectively.

2 Short Message Service (SMS)

Short message service (SMS) is the process of wireless communication which sends text from cell phone to cell phone. Short refers to the maximum size of text message, i.e., 160 characters which include uppercase alphabets, lowercase alphabets, numbers, and special characters.

2.1 SMS Architecture

To use SMS service, a SIM card (Subscriber identification module) is needed. SMS service needs SMS center client which is built in SIM card. A Base Transceiver Station (BTS) helps wireless communication between user equipment, i.e., cell phone and a network. Mobile Switching Center (MSC) is the primary service delivery node for GSM or CDMA and is responsible for routing SMSs and calls. It handles mobility, handover requirement during the call, takes care of charging and real-time prepaid account monitoring. Base station Controller (BSC), controls the GSM network, manages radio resources, and allocates channels. SMS Center (SMSC) receives a message when it is transmitted from a cell phone, finds the destination and sends it to destination cell phone. It is installed on mobile carrier core network and acts as a temporary storage for SMSs. In case the destination cell phone is switched off or inactive, SMSC stores the messages, delivers afterward when the destination cell becomes active and notifies the sender about the delivery status [4]. Figure 2 shows the architecture of SMS communication.

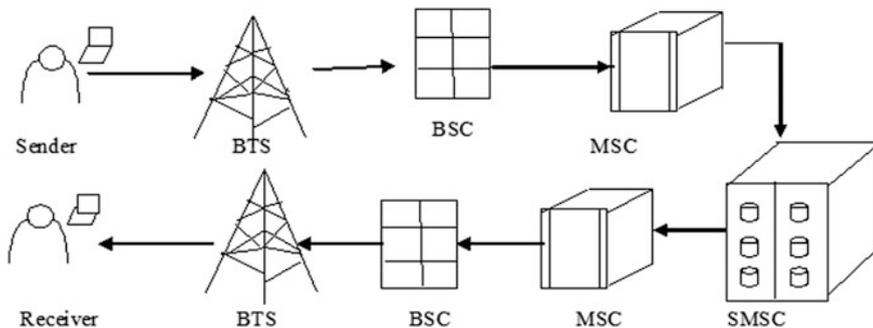


Fig. 2 Architecture of SMS

3 Text Encryption Schemes

Playfair cipher is one of the popular encryption methods suggested for SMS encryption from time to time. It is a substitution cipher which was first developed by Charles Wheatstone in 1854. Later it was popularized by Lord Playfair and hence known as playfair. It was used by the British, Australians, and Germans during World War II as it is reasonably faster to use and requires no special equipment.

The original playfair cipher uses a 5×5 matrix, a keyword and involves matrix creation and encryption. In the matrix creation step, the cells of the matrix are first populated with the unique characters of the keyword and then the remaining cells are filled with rest of the characters considering I and J as a single character. Before encryption, the message is divided into pairs. If the message length is odd then, letter X is appended to the end of message. In case of repetition of characters in a pair, one of the characters is replaced by an X. Figure 3 explains the entire procedure through an example.

Encryption and Decryption RULES [1]

The plain text, in playfair, is encrypted based on the following set of rules:

- (i) If the pair of characters belongs to the same row of matrix then replace each character by a character to the right considering the row as a circular row. For example, RA will become AB.
- (ii) If the pair of characters is in the same column of matrix then replace each character by a character below it considering the column as a circular column. For example, BX becomes LP.
- (iii) If the pair of characters does not fall into the above two categories then each character is replaced by a character that lies on its own row but in the column of the other character of the pair. For example, RL becomes BG and PM becomes HL.

The decryption rules are same as the encryption rules except that when the pair of characters fall into a single column, instead of replacing each character with the

Fig. 3 An example of encryption using 5×5 playfair

Example- keyword – CIPHER
 5x5 Plaifair Matrix after the population step :

C	I/J	P	H	E
R	A	B	D	F
G	K	L	M	N
O	Q	S	T	U
V	W	X	Y	Z

Let the Message is: LOVE YOU INDIA

Encryption Step:

Plain Text -	LO	VE	YO	UI	ND	IA
	↓	↓	↓	↓	↓	↓
Cipher Text -	GS	ZC	VT	QE	MF	AK

Encrypted Message -GSZCVTQEMFAK

character below it, the characters to be replaced by a characters above. Similarly, when the pair of characters falls into a single row, instead of replacing the characters with characters to the right, the characters are replaced by characters to the left. Rest of the algorithm remains same as the encryption algorithm.

Though playfair is one of the simplest substitution ciphers, some of the drawbacks of 5×5 traditional playfair cipher are, it takes only 25 uppercase alphabets leaving no room for the lowercase alphabets, numbers, special characters and forces two characters P, Q or I, J to be put together to a single cell. To overcome the above drawbacks, Chand, Bhattacharya [5] in their paper have suggested a text message encryption scheme using a 6×6 playfair. In their playfair, they had included numbers as well as I and J as two separate alphabets but skipped the special characters and lowercase alphabets. Aftab Alam, Shah Khalid and Muhammad Salam [6] modified the playfair to include two symbols * and # and I, J as separate alphabets. They had suggested a 7×4 playfair. Shrivastav, Chouhan, and Dhawan [7] have used an 8×8 playfair and included the numbers and 28 special characters. I and J were counted as two different characters. Dhenakaran and Llayaraja [8] in their paper have further modified the algorithm to 16×16 playfair including all the ASCII characters. Their playfair used 194 special characters. But, in practice, all these 194 special characters are not used for SMS rather a limited set of characters are used. Therefore, we have developed a 9×10 playfair that only encompasses the characters generally used for short message communications. Besides the playfair, the following are the other SMS encryption schemes suggested by researchers from time to time.

Goyal, Klinger [9] in their paper suggested a modified Caesar Cipher. Caesar Cipher is a mono alphabetic substitution cipher, where each letter in the plane text is replaced by another letter. Here they used an alphabet index for substitution. Depending upon whether the alphabet index is odd or even the value is either incremented or decremented by one. Almasri and Matjani [10] used double-secured IDEA (DS-IDEA) for text encryption. IDEA is a block cipher that operates with 64 bits plain text block and 128-bit key. It operates on three different algebraic operations such as XOR, addition modulo, and multiplication modulo. Though the major drawback of IDEA is its weak key, DS-IDEA claimed to give better security to the user's password in the online password management system. Mahajan, Sachdeva [11] in their paper implemented three encryption techniques such as AES, DES, and RSA. The author has compared their performance based on the analysis of simulation time of encryption and decryption. DES algorithm has been proposed to afford standard method for protecting sensitive, commercial, and unclassified data. Key is same for both encryption as well as decryption. AES algorithm is used not only for security but also for great speed. RSA is a public-key encryption algorithm. RSA involves three steps: key generation, encryption, and decryption. In this paper, by doing simulation the authors have shown that AES is better than DES and RSA.

Akash, Adarsh, Chandrasekhar [12] in their paper have discussed about advanced hill cipher which is distinct from conventional hill cipher. Advanced hill cipher uses involuntary matrix and permuted key. Here, the authors have

implemented a second-level encryption using permutation method that makes the cipher highly protected. Here, encryption technique is highly decisive because it uses tamper detection technique of cipher text assuring successful decryption of cipher. Second level of encryption technique is based on random value of public accomplice with modular arithmetic function. The authors have justified that this encryption technique is highly robust and secure. Schneier [13] in his paper proposed Blowfish, a new secret key block cipher technique. Blowfish cipher is a simple iterating encryption function which iterates 16 times. The block size of Blowfish is of 64 bits and it can be any length up to 448 bits. Encryption occurs by 16-round feistel network where each round consists of key. Key-dependent permutation and data-independent substitution, All operations are done using XOR and 32-bit word addition. Blowfish cipher uses large number of sub keys which are precomputed before encryption and decryption. The most efficient way to break blowfish is by exhaustive search of key space. Nanda, Awasti [14] in their paper discussed about SMS security, SMS encryption, programming platform for mobile phones. In this paper, the authors have analyzed different cryptosystems along with NTRU for implementation. NTRU is an open-source public-key cryptosystem which uses lattice-based cryptography to encrypt and decrypt data. NTRU consists of two algorithms, for encryption NTRUEncrypt and for digital signature NTRVsign. They have concluded that NTRU is faster and provides stronger security than other traditional cryptosystem.

Pavitra, Ramadevi [15] in their paper have compared various cryptography algorithms such as, AES, BLOWFISH, DES on the basis of time taken for encryption and decryption for different text files. They have also analyzed throughput and have concluded that AES algorithm is faster and produces more throughput. Nakashima and Ukamoto [16] in their paper estimated the exact key size needed by each scheme in order to theoretically guarantee security based on precise evaluation of reduction efficiency. They considered the RSA-based encryption scheme obtained by combining RSA function with five encryption schemes such as OAEP, OAEP+, SAEP, SAEP+, REACT which are proven under random oracle model. By comparing all, they found the result among schemes that RSA-OAEP has worst efficiency and far from its naming optimal. On the other hand, RSA-REACT has most efficient reduction and assures the same security with the minimum key length. Pointcheval and Okamoto [17] in their paper have proposed an enhanced asymmetric cryptosystem based on random oracle. Frozen [18] in their paper have given a brief review of different algorithms such as RSA, Diffie Hellman, DSS, DES, 3DES, IDEA. They have tried their strength in future applications. They have argued, in order to get both the security advantages of public-key systems and the speed advantages of secret key systems for encryption, the best solution is to combine public and secret key systems. Gunasundari and Elangovan [19] in their paper provided a candid resemblance between four symmetric key cryptography algorithms such as RC2, RC4, RC5, and RC6 (RC: "Ron's code" or "Rivest cipher"). Among all these, RC6 algorithm uses a variable number of bits ranging from 8 to 1024 bits and it encrypts the data 16 times, making it impossible for a hacker to tamper. Banerjee and Roy [4] in their paper presented a

example, if the sender’s and the receiver’s mobile numbers are 9823451765 and 9456893625 respectively and if both belong to India then

$IK = From:[+91-9438119366];(^&L)B@To:[+ 91-9437915662];(^&[y>)$

• *Round key Generation*

For each round, a 13-character long round key (RK_n) is generated from the 57 character long initial key (IK) following the logic given below:

Step 1: find the largest and smallest numbers possible from the sender’s and receiver’s mobile numbers.

E.g., $LS = 9986643311$, $SS = 1133466899$, $LR = 9976654321$, $SR = 1234566799$, in this case.

Step 2: Compute LS-SS

E.g., 8853176412, in this case

Step 3: Compute RK_1 taking first 13 characters after right shifting IK by number of characters same as the 1st character of LS-SS in a circular manner.

For example, to compute RK_1 , right circular shifting IK by 8 characters, the new IK becomes:

$IK = ;(^&[y>)From:[+91-9438119366];(^&L)B@To:[+91-9437915662]$

$RK_1 =$ first 13 characters of $IK = ;(^&[8 >)From:$

Step 4: Repeat step 3 with remaining values of LS-SS to obtain the other round keys RK_2 through RK_{10} .

For example, to compute RK_2 , right circular shift IK of the previous round by 8 characters and take the first 13 characters.

$IK = 7915662];(^&[y>)From:[+ 91-9438119366];(^&L)B@To:[+91-94379$

$RK_2 =$ first 13 characters of $IK = 7915662] ;(^&[$

Figure 5 shows the 10 round keys for LS-SS = 8853176412.

• *Message Authentication*

To verify whether the message has been undergone any change during transmission or not, the algorithm appends 20 characters from the character set given in Fig. 4 such as, 3 digits of message length count, 14 checksum characters and 3 digits of word count. The 14 checksum characters are calculated as follows:

Step 1: Arrange the characters of the plaintext message into a 14×10 matrix

Step 2: Calculate the column sums $CSum_1$ through $CSum_{14}$

Step 3: Calculate ASCII values $(CS_i) = CSum_i$ modulo $95 + 32$.

RK ₁	;	(^&[y>)	From:	RK ₃	1-9437915662]	RK ₅	:[+91-9437915	RK ₇	6];(^&L)B@To	RK ₉	119366];(^&L]
RK ₂	7915662];(^&[RK ₄	[+91-94379156	RK ₆	L)B@To:[+91-	RK ₈	19366];(^&L)B	RK ₁₀	38119366];(^&		

Fig. 5 Round keys for sender and receiver numbers +91-9438119366 and +91-9437915662

4.2 Plain Text Preparation and Encryption

Initial plain text (IPT) is the message created by the sender.

For example, let $IPT = \text{"For balance enquiry dial *123#."}$

This plain text is appended with the number of characters (including space) in the message, a set of 14 authentication characters and the total number of words in the message to create a new plain text $PT1$, for error checking and ensuring message integrity. The $PT1$, before every round, is right shifted (circular) by a number of characters equal to the digit of LR-SR to create the plain text (PT). The steps of the algorithm are as given below.

- Step 1: Read the plain text IPT
- Step 2: Append the total number of character to the IPT , the 14 authentication characters and total no of words present in the IPT to the end of the IPT to generate $PT1$. For example, the $PT1$ corresponding to the IPT given above is
- $$PT1 = \text{For balance enquiry dial*123#.029=G>T>C3-22QrYc4\#}$$
- Step 3: Compute LR-SR
e.g., 8741087522, in this case
- Step 4: Obtain PT_{R1} for round 1 by shifting $PT1$ by number of characters same as the first character of LR-SR in a circular manner.
Example: PT_{R1} for the above $PT1$ is
- $$PT_{R1} = \text{-22QrYc4For balance enquiry dial*123#.29 = G>T > C3}$$
- Step 5: Populate the 10×9 playfair matrix with unique characters from $RK1$. The remaining cells to be appended by taking characters from the character set in ascending order of their ASCII values. For odd and even rounds, the cell after the key is entered, to be populated differently, in ascending order during odd rounds and in descending order during the even round.
- Step 6: Follow the playfair encryption procedure to encrypt PT_{R1} to CT_{R1} .
- Step 7: Repeat steps 4 through 6 by replacing PT_R of n th round with CT_{Rn-1} and considering the n th digit of LR-SR for shifting.

5 Results and Discussions

The encryption/decryption modules have been developed using MATLAB and some plaintext and their corresponding cipher texts are given in Fig. 6. The cipher text generated at the end of the tenth round is transmitted and at the receiver's end the cipher text is decrypted with keys K_{10} through K_1 to get back the plain text. The scheme is found to be robust against various attack models.

Sl. No	Plain Text	Cipher Text
1	For balance enquiry dial*123#.	UC#*uc+nqkxi;z\$dlI;2B2i)1IK9&lr)MEE_97UrfmkX<vLj?s
2	Please send the book by Dan Brown.	[qou#c"ZyqkUc1;?BhtUs%jx#D'MKY MXdIl.S^A=Y%>?aM8A3P&wo
3	Please send the book of Dan Brown.	[qou#c"Zz5b6pt2Q4qtUs%jx#D'MKY M TdkM7.^>=Y%>?aM8A3P&wo
4	Please read the book by Dan Brown.	!Soz#c"ZyqkUc1;?BhtUs%jx#D'MKY MXdIl=Sl2=Y%>?aM8Dzs Bk
5	Please read the book of Dan Brown.	!Soz#c"Zz5b6pt2Q4qtUs%jx#D'MKY M Udh1-Zt?=Y%>?aM8Dzs Bk
6	the book of Dan Brown.	7/#+##"Zz5b6pt2Q4qtUs%jx DmwCAvJ,1Jl>9.YZ5
7	Dan Brown.	;R_FHmE,S-h)7+/HBItUs%jx N .x0
8	Please	^S/7*nvs#!!!!+O/i/>'BW&<q
9	Hi! This is Minati Mishra from Fakir Mohan University, Balasore, Odisha.	h<a=]4.23l_0"r6#\$SlaihT;.5-aanVf)W>#QpKV3)S%\$;+Hta('h6UaRhSy-;0tzkO"w-orPootl=jMObR,S;YSj
10	Hi! This is Pranab Sahu from Fakir Mohan University, Balasore, Odisha.	h<a=]4.23l_0"r6#\$SlaihT;;3.<<bess%95zepam>%fs+)Hta('nv)gvSwoaV_mjunh=jrP0otl=jMObR,S;YSj
11	Fakir Mohan University	vF#lkkk?0_fffSlAwsPitq[4s8C9-"6'UI&+OOPR^}
12	Please send the book of Dan Brown.(Validation Failed)	[qou#c"Zz5b6pt2Q4qtUs%jx#D'MKY MXdIl.S^A=Y%>?aM8A3P&wo

Fig. 6 Encryption rounds with an example at the transmitter’s end and the decrypted message

5.1 Message Authentication

At the receiving end, the system checks the integrity of the message by counting the number of characters, words, calculating the authentication characters and matching those values with the values appended to the message. If the values match, the plain text message is presented to the receiver as an authentic message. If the authentication fails then the decrypted message is presented with ‘validation failed’ comment, as shown in serial no. 12 of Fig. 6.

5.2 Cipher Text Only Attack (CTA)

It can be seen from Fig. 6 that, an adversary cannot derive the plaintext (PT) from the cipher texts as the cipher texts (CT) do not exhibit any direct relation with the PT.

5.3 Known Plaintext Attack (KPA)

Having access into a random set of plaintexts and cipher texts, e.g., knowing PT-CT pairs given in Sl. No. 1, 5, and 10, the adversary cannot say the alphabet n will be mapped to what.

5.4 Chosen Plaintext Attack (CPA)

It will not be possible for an adversary to guess the right PT even if he has allowed access to a chosen set of plaintexts and their corresponding cipher texts.

For example, knowing the PT-CT pairs given in Sl. No. 1 through 4, an adversary cannot say how ‘Dan Brown’ or ‘please’ will be encrypted.

5.5 *Replay Attack*

Though the method survives against various possible attacks during message transmission, it fails to detect the replay attack. An adversary, who has access to the medium, though, cannot partially change a message without getting noticed, can successfully replace an entire message with some previous message!

6 Summary and Conclusion

Here, we have suggested a novel encryption scheme for SMS communication which is simple as well as robust against various attacks. The encryption/decryption processes have been developed with a balanced mix of both the confusion and diffusion techniques to make it very hard for an adversary to crack. The round keys are auto generated from the sender’s and receiver’s numbers making those unique for each pair of communicators. Different round keys for each round as well as a different population technique for odd and even rounds make the algorithm stronger. The authentication characters ensure message integrity. To make the system robust against the replay attack, we are planning to use a suitable time-stamp in the future version.

References

1. Kahate, A.: *Cryptography and Network Security*, 2nd edn. Tata McGraw Hill Publishing Company Limited, New Delhi
2. Stallings, W.: *Cryptography and Network Security: Principles and Practice*, 4th edn. Prentice Hall (2006)
3. Gosh, K., Mishra, M.: Elliptic key cryptography and time bound key management. *Anvesa* **10** (1&2), 34–40 (2015)
4. Banerjee, S., Roy, U.: Secure SMS communication in android based system with two stage protection. *Int. J. Comput. Sci. Mob. Comput.* **4**(6), 1057–1064 (2015)
5. Chand, N., Bhattacharyya, S.: A novel approach for encryption of text message using palyafair cipher 6 by 6 matrix with four iteration steps. *Int. J. Eng. Sci. Innovative Technol. (IJESIT)* **3**(1), 478–484 (2014)
6. Aftab Alam, A., Shah Khalid, B., Muhammad Salam, C.: A modified version of playfair cipher using 7×4 matrix. *Int. J. Comput. Theor. Eng.* **5**(4), 626–628 (2013)
7. Shrivastava, G., Chouhan, M., Dhawan, M.: A modified version of Extended playfair cipher (8×8). *Int. J. Eng. Comput. Sci.* **2**(4), 956–961 (2013)

8. Dhenakaran, S.S., Llyaraja, M.: Extension of playfair cipher using 16×16 matrix. *Int. J. Comput. Appl.* **48**(7), 37–41 (2012)
9. Goyal, K., Klinger, S.: Modified caesar cipher for better security enhancement. *Proc. Int. J. Comput. Appl.* **73**(3), 26–31 (2013)
10. Almasari, O., Rmatjani, H.: Introducing encryption algorithm based on IDEA. *Int. J. Sci. Res. (IJSR)* **2**(9), 334–339 (2013)
11. Mahajan, P., Sachdeva, A.: A study of encryption algorithms AES, DES and RSA for security. *Glob. J. Comput. Sci. Technol. Netw.* **13**(15) (2013)
12. Akash, H.P., Adarsh, K., Chandrasekhar, S.: A secure encryption technology based on advanced hill cipher for public key encryption. *IOSR J. Comput. Eng.* **11**(2), 10–14 (2013)
13. Schmeier, B.: Description of a New Variable-Length Key, 64-Bit Block Cipher (Blowfish), Counterpane Systems, 730 Fair Oaks Ave, Oak Park, IL 60302 (2005)
14. Nanda, A.K., Awasti, L.K.: A proposal for SMS security using NTRU cryptosystem. In: *Proceeding to 9th International Conference*, pp. 706–718. Springer (2013)
15. Pavitra, S., Ramadevi, E.: Performance evolution of symmetric algorithm. *J. Glob. Res. Comput. Sci.* **3**(8), 43–45 (2012)
16. Nakashima, T., Ukamoto, T.: Natural science report. *Ochanomizu Univ.* **57**(1), 37–55 (2006)
17. Pointcheval, D., Okamoto, T.: REACT: rapid enhanced asymmetric cryptosystem transform. In: *Proceeding of Cryptographer's Track at RSA Conference, San Francisco, California*, Springer, pp. 159–175 (2001)
18. Frosen, J.: *Practical Cryptosystem and Their Strength*. Department of Computer Science, Helsinki University of Technology
19. Gunasundari, T., Elangovan, K.: Symmetric Key Cryptography: Current Trends. *Int. J. Comput. Sci. Mob. Appl.* **2**(2), 78–83 (2014)

Reviewing the Performance of an Improved Seven-Level Multilevel Inverter for Various Pulse Width Modulation Techniques

K. Aroul, S. Umashankar, K.R. Prabhu and P. Sanjeevikumar

Abstract Multilevel inverter has become a unique choice in industries at high-power applications. It is mainly due to the high-quality output waveform. Conventional MLIs use many power switches and the design of such a topology is not economical. An improved single-phase MLI is suggested in this article. The circuit consists of reduced number of power switches and hence it possesses a simple control circuitry. The appealing feature of the designed circuit is only a single power switch is switched at a high frequency at anytime during the operation. It minimizes the electromagnetic interference and switching losses in the circuit. The performance of the said topology is investigated using sinusoidal pulse width modulation (SPWM) techniques in Simulink software. The performance of the proposed converter is compared with different PWM techniques.

Keywords Seven-level inverter • SPWM techniques • Total harmonic distortion Modulation frequency ratio • Side band harmonics

K. Aroul · S. Umashankar · K.R. Prabhu (✉)
School of Electrical Engineering, Vellore Institute of Technology (VIT) University,
Vellore, Tamil Nadu, India
e-mail: prabhu.kr@vit.ac.in

K. Aroul
e-mail: aroul.k@vit.ac.in

S. Umashankar
e-mail: umashankar.s@vit.ac.in

P. Sanjeevikumar
Department of Electrical and Electronics Engineering, University of Johannesburg,
Auckland Park, Johannesburg, South Africa
e-mail: sanjeevi_12@yahoo.co.in

P. Sanjeevikumar
Research and Development, Power Electronics Division, Ohm Technologies,
Chennai, India

1 Introduction

Multilevel inverter has been receiving a wide attention due to its lesser electromagnetic interference and reduced harmonic distortion. The most frequent MLIs discussed in the literature are diode-clamped, flying capacitor and cascaded H-Bridge MLI. Diode-clamped MLIs have an uneven transitional DC voltage levels. This leads to a distortion in the output waveform producing more harmonics in the system. The number of diodes used for such an inverter at higher level is significantly very high than the other techniques [1, 2]. Flying capacitor MLI is similar to its predecessor, with its diode being replaced by capacitor. This topology also faces a similar disadvantage requiring additional capacitors at higher levels increasing the complexity of the circuit [3, 4]. Cascaded H-Bridge topology has been gaining popularity in recent times since a minimum number of components are required for a given level [5–8]. Many control strategies are used to improve the performance of the converters [9–11].

Higher levels in MLI demand more number of switches operating at high frequency. It leads to heavy switching losses in the topology. The prime concern in MLI design is to construct a topology that uses reduced number of switches during each level of its operation [12]. Conduction and switching losses are high in a series-connected MLI presented in [13]. Switches involved in the operation at the higher levels for the circuit presented in [14] are very high. To bypass a DC-source and power switch diodes are used in anti-parallel increasing the complexity of the topology presented in [15].

This research article proposes a seven-level inverter using a full-bridge power converter. At any time in its working only a single power switch is in operation at a high frequency. This ensures negligible switching loss in the device and improved power efficiency. The supremacy of the said converter is related with equal pulse width technique, selective sinusoidal pulse width modulation (S-SPWM) and multicarrier sinusoidal pulse width modulation (MC-SPWM).

This article is segmented as follows: Sect. 2 discusses about the various modes of operation of the proposed converter. Section 3 briefs about the various PWM techniques. Section 4 focuses on the simulation results of the seven-level inverter. Sections 5 and 6 discusses the result obtained from simulation.

2 The Proposed Topology

The topology consists of three voltage sources V_1 , V_2 , V_3 for obtaining a seven-level inverter. It can be generalized by the expression,

$$2V + 1 = N \quad (1)$$

where,

V Required no. of Voltage Sources

N Inverter level

For instance, to achieve a 13-level inverter, six voltage sources will be required for the topology shown in Fig. 1. Figures 2 and 3 are presented here to establish the novelty of the recommended topology. The switching order for N -level topology shown in Fig. 2 is complex as it involves a lot of power switches; hence the control circuit is also intricate. The switching order shown in Fig. 3 involves many

Fig. 1 The proposed topology (switches S1, S2 and S3 are unidirectional)

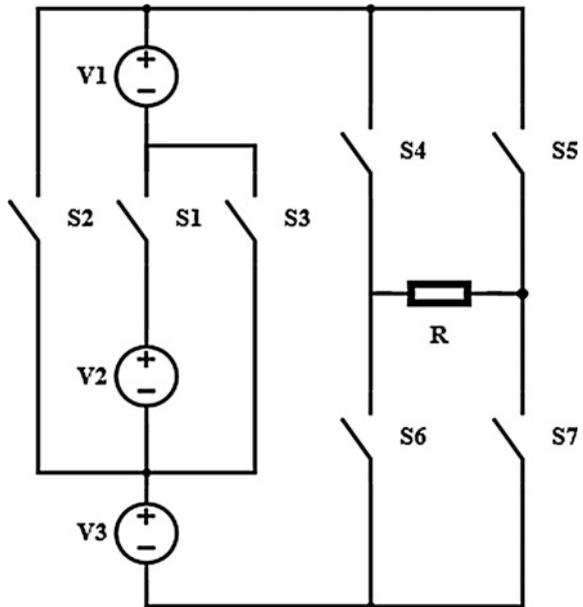


Fig. 2 A multilevel inverter topology presented in [13]

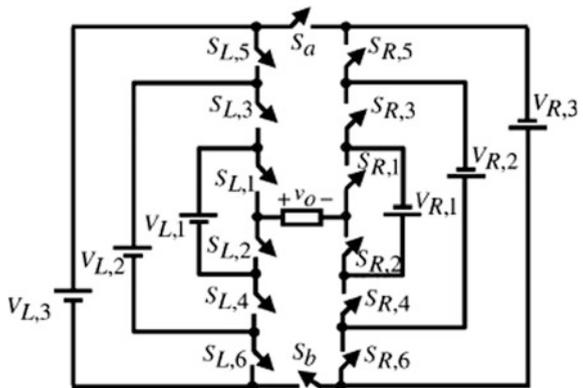


Fig. 3 A 7-level inverter topology presented in [14]

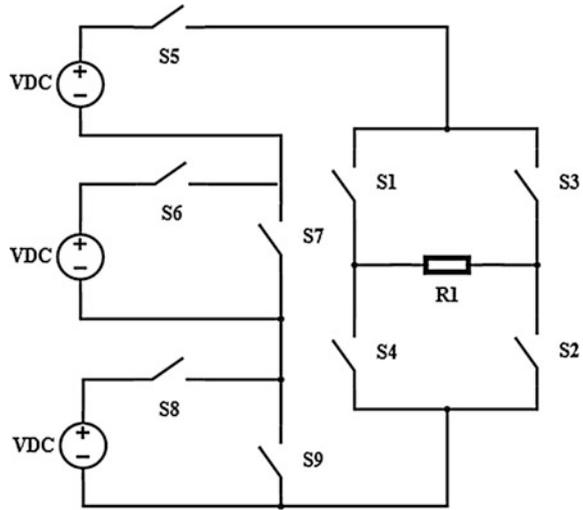


Table 1 Switching scheme of the proposed converter

No.	S1	S2	S3	S4, S7	S5, S6	V0
1	–	–	–	ON	–	0
2	–	ON	–	ON	–	V3
3	–	–	ON	ON	–	V3 + V1
4	ON	–	–	ON	–	V3 + V2 + V1

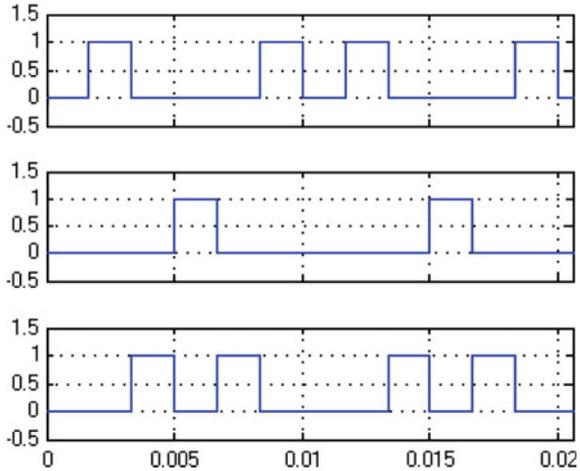
switches operating at high frequency. This leads to an increase in conduction and switching loss. Hence, the proposed topology has a simple control circuitry and less switching losses (Table 1).

3 Equal Pulse Width Technique

An N -level inverter has $2N - 2$ equal pulse widths. Hence, a seven-level inverter has 12 pulse widths of equal proportion. The conduction period is 30° for each state. The conduction states are 0, $V3$, $V1 + V3$, $V1 + V2 + V3$, $V1 + V3$, $V3$, 0, $-V3$, $-(V1 + V3)$, $-(V1 + V2 + V3)$, $-(V1 + V3)$, $-V3$.

There are two ‘zero’ states, hence only ‘ten’ states requires pulse signals. $V1 + V3$ state is obtained by triggering switch S3. There are four $V1 + V3$ states. Hence, the switch S3 needs to be turned on 4 times each cycle. The four pulse signal shown in Fig. 4 is fed into ‘OR’ gate. Similarly switch S2 needs to be turned on 4 times for each cycle to obtain the state $V3$. Switch S1 needs to be turned on only twice to obtain the state $V1 + V2 + V3$.

Fig. 4 Generated pulse width signals for switches S2, S1, S3 respectively



4 SPWM Technique

Sinusoidal Pulse Width Modulation technique has pulses with different duty cycles but constant amplitude for each period. To generate the SPWM signal, conventionally, carrier triangle wave is compared with the sine wave of operating frequency. Repeating sequence block from Simulink is used to generate the triangular waveform. Relational operator is used as a comparator to generate pulse signals. The frequency of carrier wave is chosen as 10 kHz and the frequency of sine wave is 50 Hz (fundamental frequency).

4.1 Selective-SPWM Technique (S-SPWM)

In this technique, the generated SPWM obtained by relating a sine and triangular wave is logically AND with the PWM pulses of equal pulse width. The obtained high-frequency pulse signal is fed to Switches S1, S2, and S3. The generated pulse signals using this technique is shown in Fig. 5.

4.2 Multi Carrier-SPWM Technique (MC-SPWM)

In this technique, four triangular carrier waves are compared with a sinusoidal wave of fundamental frequency to obtain the gate pulses. Based on the alignment of four triangular carrier waves vertically MC-SPWM can be further classified into (i) Phase Opposition Disposition PWM (POD-PWM) (ii) Alternative Phase

Fig. 5 Generated pulse width signals using S-SPWM technique

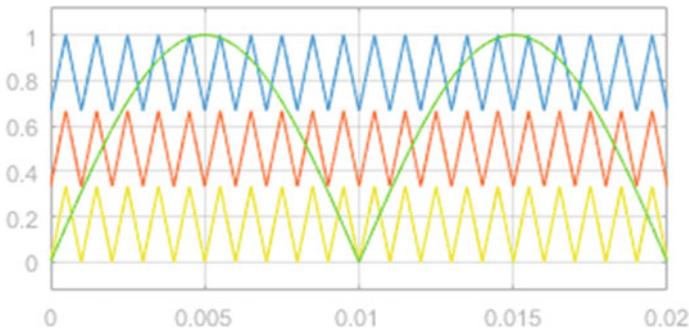
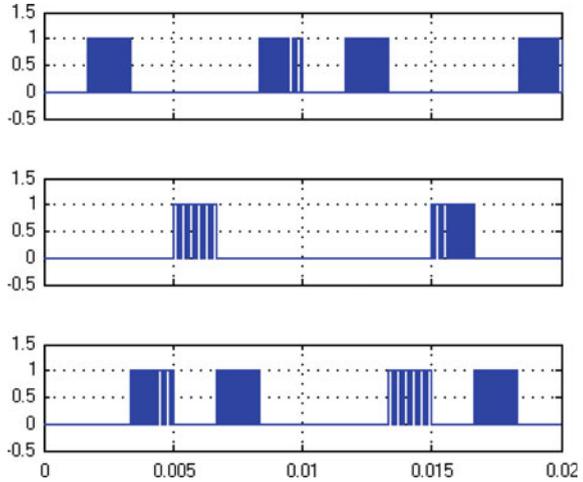


Fig. 6 POD-PWM technique at 1 kHz carrier frequency

Opposition Disposition PWM (APOD-PWM) (iii) Phase Disposition PWM (PD-PWM).

4.2.1 Phase Opposition Disposition PWM (POD-PWM)

Conventionally for an N -level inverter ($N - 1$) numbers of triangular waveforms are generated. For a 7-level inverter, the three triangular waves meant for generating pulse signals during positive half of sine wave are out of phase with the three triangular waveforms that generate signals during negative half. Due to the symmetrical nature of the said carrier waveforms, only three triangular waveforms are generated to compare with the sine wave as shown in Fig. 6. This reduces the complexity in pulse generation circuit. To enable comparison between the carrier and fundamental waves, sine is fed to absolute function to make it oscillate only in the positive half.

Each triangular wave has amplitude of $1/3$. The three triangular waves make an amplitude one. Sine wave also is generated for amplitude of one. So, the amplitude modulation index is unity. The carrier wave is chosen as 10 kHz. The generated pulse width signal is presented in Fig. 7.

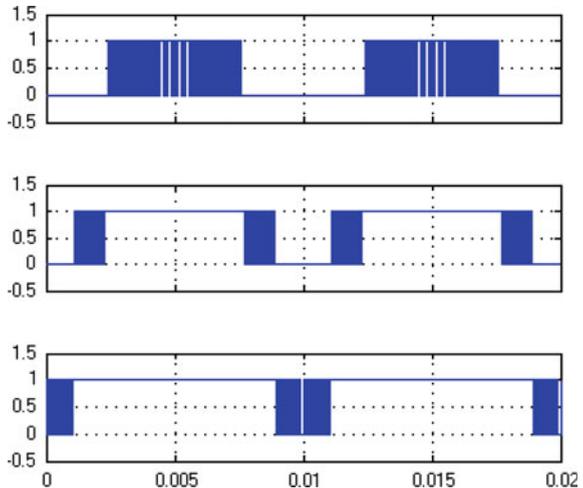
4.2.2 Alternated Phase Opposition Disposition PWM (APOD-PWM)

The alternate triangular waves are in phase opposition. Hence, six triangular waves are used in this technique, which ranges from -1 to $+1$ as shown in Fig. 8. The sine wave also varies from -1 to $+1$. The amplitude modulation index is chosen as unity in this case too for a better THD. One modulated waveform from the positive half-cycle and a corresponding modulated waveform from the negative half-cycle is logically ‘OR’ ed. The resulting high-frequency pulse width modulated signals are fed to the switches S1, S2, and S3. The generated pulse signal is similar to POD-PWM technique.

4.2.3 Phase Disposition PWM (PD-PWM)

In this technique all the six waveforms, which ranges from -1 to $+1$ are in-phase as shown in Fig. 9. The procedure carried out in creating Simulink model is similar to APOD-PWM technique.

Fig. 7 Generated pulse width signals using POD-PWM technique



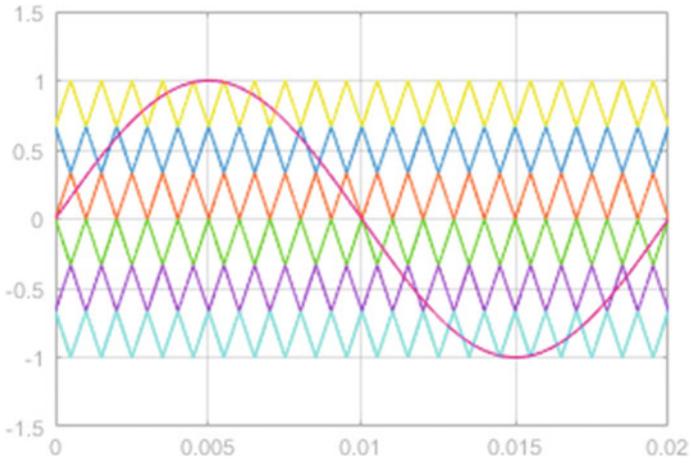


Fig. 8 APOD-PWM technique at 1 kHz carrier frequency

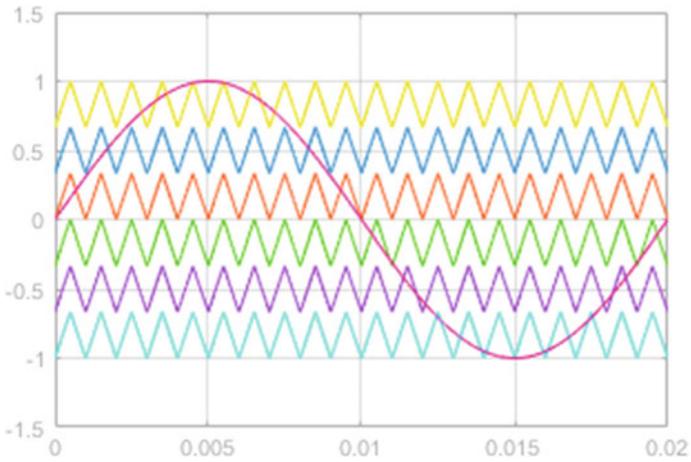


Fig. 9 PD-PWM technique at 1 kHz carrier frequency

5 Simulation Results

Three Voltage source each of 80 V is used in the proposed topology. The high-frequency switches are unidirectional. The topology is designed to deliver a power of 500 W. The carrier frequency is chosen as 10 kHz and the amplitude modulation index is 1.

In equal pulse width technique, the harmonics decay exponentially; hence it has a considerable amount of lower order harmonics as shown in Fig. 11a and the output voltage is shown in Fig. 10b. In S-SPWM technique considerable amount of lower order harmonics and sideband harmonics are present in the system. The magnitude of the side band harmonics at carrier frequency is 10% of the fundamental. So the S-SPWM technique has the poorest THD as shown in Fig. 11b. MC-SPWM provides the best THD and fundamental component among the discussed techniques. The only drawback of the said technique is the sideband harmonics, which occurs at multiples of carrier frequency. The magnitude of side band harmonics at carrier frequency is 6% of fundamental in all the three MC-SPWM techniques (Figs. 10 and 11).

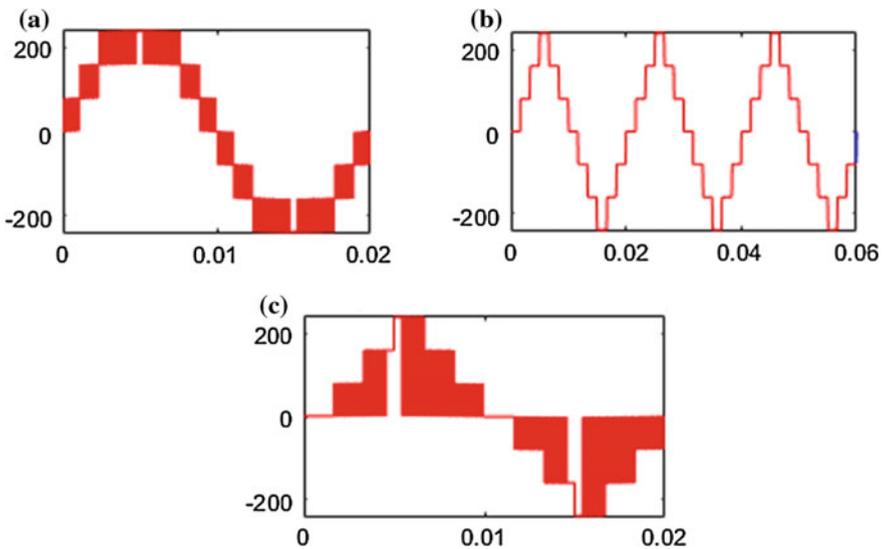


Fig. 10 Output voltage for **a** MC-SPWM **b** equal pulse width **c** S-SPWM

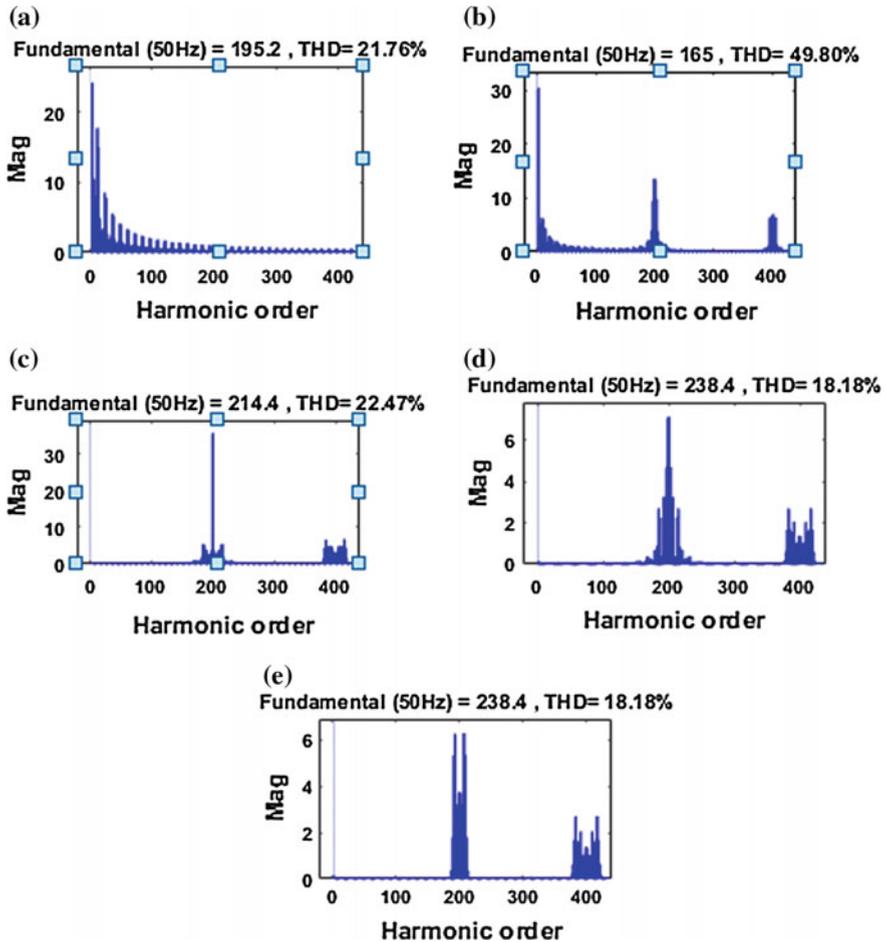


Fig. 11 THD for a equal pulse width b S-SPWM c PD-PWM d POD-PWM e APOD-PWM

6 Conclusion

The percentage harmonics for the carrier-based PWM techniques is presented in Table 2. The percentage harmonics as a function of modulation frequency ration is presented in Table 3 for MC-SPWM techniques alone.

Table 3 clearly indicates that POD-PWM gives a better THD at lower switching frequencies and it is found to vary as a function of carrier wave frequency. The other two MC-SPWM techniques are resistant to the carrier wave frequency variations.

In this research article, an effective seven-level inverter is presented, where only one high-frequency switch is in operation at any instant. The performance of the

Table 2 Harmonics percentage at 10 kHz carrier frequency

PWM technique	Fundamental (V) (240 V, 50 Hz)	THD (%)
Equal pulse width	195.2	21.76
S-SPWM	165	49.80
PD-PWM	238.4	18.18
APOD-PWM	238.4	18.18
POD-PWM	238.4	18.18

Table 3 Harmonics percentage at various carrier frequencies

Carrier wave frequency (kHz)	APOD-PWM (%)	PD-PWM (%)	POD-PWM (%)
1	18.49	22.51	16.03
2	18.26	22.44	17.99
4	18.13	22.48	18.14
6	18.12	22.49	18.20
8	18.21	22.49	18.20
10	18.18	22.47	18.18

proposed inverter is analyzed with several PWM techniques. Unlike the other MC-SPWM, POD-PWM requires only three triangular waveforms to generate PWM pulses. Hence, POD-PWM stands tall in terms of control technique as well as provides better THD at lower frequencies.

References

- Hasegawa, K., Akagi, H.: Low-modulation-index operation of a five level diode-clamped pwm inverter with a dc-voltage-balancing circuit for a motor drive. *IEEE Trans. Power Electron.* **27**, 3495–3505 (2012)
- Chaves, M., Margato, E., Silva, J.F., Pinto, S.F.: New approach in back-to-back m-level diode clamped multilevel converter modeling and direct current bus voltages balancing. *IET Power Electron.* **3**, 578–589 (2010)
- McGrath, B.P., Holmes, D.G.: Analytical determination of the capacitor voltage balancing dynamics for three-phase flying-capacitor converters. *IEEE Trans. Ind. Appl.* **45**, 1425–1433 (2009)
- Zhang, L., Waite, M.L., Chong, B.: Three-phase four-leg flying-capacitor multi-level inverter-based active power filters for unbalanced current operation. *IET Power Electron.* **6**, 153–163 (2013)
- Ebrahimi, J., Babaei, E., Gharephetian, G.B.: A new topology of cascaded multilevel converters with reduced number of components for high-voltage applications. *IEEE Trans. Power Electron.* **26**, 3109–3118 (2011)
- Ebrahimi, J., Babaei, E., Gharephetian, G.B.: A new multilevel converter topology with reduced number of power electronic components. *IEEE Trans. Ind. Electron.* **59**, 655–667 (2012)

7. Babaei, E., Farhadi Kangarlu, M., Sabahi, M.: Extended multilevel converters: an attempt to reduce the number of independent dc voltage sources in cascaded multilevel converters. *IET Power Electron.* **7**, 157–166 (2014)
8. Napoles, J., Watson, A.J., Padilla, J.J.: Selective harmonic mitigation technique for cascaded H-bridge converter with nonequal dc link voltages. *IEEE Trans. Ind. Electron.* **60**, 1963–1971 (2013)
9. Sanjeevikumar, P., Blaabjerg, F., Wheeler, P., Khanna, R., Mahajan, S.B., Dwivedi, S.: Optimized carrier based five-level generated modified dual three-phase open-winding inverter for medium power application. *IEEE Int. Transport. Electrification Conf. Expo.* 40–45 (2016)
10. Sanjeevikumar, P., Blaabjerg, F., Wheeler, P., Ojo, J.O.: Three-phase multilevel inverter configuration for open-winding high power application. In: *The 6th IEEE International Symposium on Power Electronics for Distributed Generation Systems* (2015)
11. Sanjeevikumar, P., Rajambal, K., Chinnaponnu, A.: Comparison of MLI and Z-source inverter for transformer-less operation of single-phase photovoltaic system. *Thammasat Int. J. Sci. Technol.* **15**, 37–47 (2010)
12. Albert Alexander, S., Sanjeevikumar, P.: Reduction of Semiconductor Devices in Multilevel Inverter with Digital Switching Technique. *Lecture Notes in Electrical Engineering*. Springer Journal Publications (2016) (Accepted for Publication)
13. Babaei, E., Laali, S.: Optimum structures of proposed new cascaded multilevel inverter with reduced number of components. *IEEE Trans. Ind. Electron.* **62**, 6887–6895 (2015)
14. Kirubakaran, A., Vijayakumar, D.: Development of Lab VIEW-based multilevel inverter with reduced number of switches. *Int. J. Power Electron.* **6**, 88–102 (2014)
15. Arif, A.-J., Edwin, N.: Cascading of diode bypassed transistor-voltage source units in multilevel inverters. *IET Power Electron.* **6**, 554–560 (2013)

Robust Multi-thresholding in Noisy Grayscale Images Using Otsu's Function and Harmony Search Optimization Algorithm

K. Suresh and U. Sakthi

Abstract Multilevel segmentation in images clusters pixels depends on the total thresholds and intensity values. To find optimal thresholds and to maximize the objective function, entails a lot of computational power and memory. In this work gray-level segmentation is proposed by Otsu-based Harmonic Search Optimization Algorithm (HSOA) algorithm to resolve such drawbacks. The HS algorithm is employed to explore the optimum values of threshold by Otsu's maximization objective function. Its effectiveness based on HS technique has been applied on 5 standard images with a size of 512×512 . The images are associated with Gaussian (GN) and Salt-and-Pepper (SAP) noise. The measurable examination is performed with the parameters of between-class variance (Objective Function) value and quality measures, such as Root Mean Square Error (RMSE) and Peak Signal-to-Noise Ratio (PSNR). The experimental procedure is employed with MATLAB software. Experimental outcomes of Otsu-based harmony search offers an optimal solution to multilevel thresholding problem for the GN and SAP noise applied images with improved objective function and faster convergence.

Keywords Otsu between-class variance • Multi-thresholding • Harmony search optimization • Noise • Image quality measures

K. Suresh (✉)

Department of Computer Science and Engineering,
Sathyabama University, Chennai, India
e-mail: sureshk@stjosephs.ac.in

U. Sakthi

Department of Computer Science and Engineering,
St. Joseph's Institute of Technology, Chennai 600119, India
e-mail: sakthi.ulaganathan@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_52

1 Introduction

In recent years, image processing demonstrates a significant part in image investigation to have better understanding of an image in several domains such as pattern analysis, destruction detection, geoscience, and medical field. The progress of computing technologies and digital imaging trials enriched the growth of imaging research [1–3].

Several procedures have been discussed and implemented in the image-processing literature [4, 5]. Among them, image segmentation with heuristic algorithm procedures is widely adopted by many researchers because of its simplicity, flexibility, and ease of implementation [6]. In order to solve the multilevel thresholding problem, between-class variance and the entropy-based analysis are widely adopted for the grayscale and RGB image. Among these two, it is proved that, the Otsu's between-class variance function will offer better image quality measures compared with entropy function. Hence, in this proposed work, Otsu's function of between-class variance-guided heuristic search has been considered to segment noise stained test images.

In image thresholding literature, various heuristic algorithms, namely Particle swarm optimization [7, 8], Bat optimization [9], Bacterial foraging optimization [10], Cuckoo search [11], optimization based on Artificial bee colony [1] and Firefly optimization [12, 13] are implemented by most of the researchers to solve the multi-thresholding issue. In this projected work, Harmony Search Optimization Algorithm (HSOA) discussed by Oliva et al., is considered [14]. During the optimization problem, HSOA explores the ' T -dimensional search space, until the Cost Function (CF) is maximized. The suggested method is tested with 5 typical images of size 512×512 , stained with Gaussian and also with salt and pepper noise. Further, the performance of this method is validated by computing the image parameters, such as Otsu's function, RMSE and PSNR.

2 Overview of Harmony Search Optimization Algorithm

The Harmony Search Optimization Algorithm (HSOA) was initially developed by Geem et al. in 2001 to solve water distribution network [15] based on the music regulation process. During the music composition process, the musician will search for the better harmony by trying different combinations of pitches. The basic HSOA depends on these two essential parameters: (i) Harmony Memory Considering Rate (HMCR) and (ii) Pitch Adjusting Rate (PAR).

A detailed description of the HSOA is available in [16, 17]. In this work, the HSOA discussed by Oliva et al., is adopted [14].

In this work, the pitch adjustment is calculated as follows:

$$x_{\text{new}}(j) = \begin{cases} x_{\text{new}}(j) = x_{\text{new}}(j) \pm \text{rand}(0,1) \cdot \text{BW}; \text{with probability PAR,} \\ x_{\text{new}}(j), \text{ with probability } (1 - \text{PAR}) \end{cases} \tag{1}$$

Pitch altering is accountable to produce new promising harmonies by vaguely adjusting the original positions. This process reflects the mutation process in existing evolutionary algorithms. This process continues till the assigned iteration value and the final updated position is considered as the optimal solution of HSOA.

3 Otsu’s Function

Otsu’s proposed an image thresholding method in 1979 which is widely used for various applications in a modified form till recent years [18]. This scheme offers the optimum thresholds of an input image by maximizing the sum of weighted variance value. A complete depiction of Otsu’s weighted sum of between-class variance technique is found in [18].

The Otsu-assisted segmentation technique is given as:

For any image, say the intensity levels be L which range from $[0, 1, 2, 3, \dots, L - 1]$. Forbi-level thresholding, given input image is distributed into C_0 and C_1 (background and object of interest) Classes, through a threshold value ‘ t ’. The cluster C_0 covers gray values from 0 to $t-1$ levels and the cluster C_1 covers the gray values from ‘ t ’ to $L-1$ levels.

The class probabilities for the two gray levels C_0 and C_1 are expressed as [18]:

$$C_0 = \frac{\rho_0}{\omega_0(t)} \dots \frac{\rho_{t-1}}{\omega_0(t)} \text{ and } C_1 = \frac{\rho_t}{\omega_1(t)} \dots \frac{\rho_{L-1}}{\omega_1(t)} \tag{2}$$

where $\omega_0(t) = \sum_{i=0}^{t-1} \rho_i, \omega_1(t) = \sum_{i=t}^{L-1} \rho_i \quad 0 \leq t \leq L(255)$

Note: ρ_i signifies the probability distribution of gray-level intensities ‘ i ’ of the input image. The mean levels μ_0 and μ_1 for C_0 and C_1 is expressed as

$$\mu_0 = \sum_{i=0}^{t-1} \frac{i\rho_i}{\omega_0(t)} \text{ and } \mu_1 = \sum_{i=t}^{L-1} \frac{i\rho_i}{\omega_1(t)} \tag{3}$$

The mean intensity value (μ_t) of the complete image is characterized as $\mu_t = \omega_0\mu_0 + \omega_1\mu_1$ and $\omega_0 + \omega_1 = 1$.

Finally the Objective (Cost) function of bi-level thresholding is given as

$$\text{Maximize } J(t) = \sigma_0 + \sigma_1 \tag{4}$$

where $\sigma_0 = \omega_0(\mu_0 - \mu_t)^2 \quad \sigma_1 = \omega_1(\mu_1 - \mu_t)^2$

The between-class variance technique is able to prolong for multilevel thresholding process with a variety of threshold values is represented as:

Let us take the total number of thresholds be ' n ' (t_1, t_2, \dots, t_n), which divide the input sample image to ' $n + 1$ ' classes: C_0 with gray levels ranging from 0 to $t_1 - 1$, C_1 with bounded gray levels ranging from t_1 to $t_2 - 1, \dots$, and C_n includes gray levels from t_n to $L-1$.

For multilevel thresholding problem, the objective function is represented as

$$J_{\max}(n) = \sigma_0 + \sigma_1 + \dots + \sigma_m \quad (5)$$

where $\sigma_0 = \omega_0(\mu_0 - \mu_n)^2$,

$$\sigma_1 = \omega_1(\mu_1 - \mu_n)^2, \dots,$$

$\sigma_m = \omega_m(\mu_m - \mu_n)^2$ and $J_{\max}(n)$ signifies Otsu's function of between-class variance.

The segmented image quality depends on the Cost Function (CF) chosen to guide the search.

In this work, J_{\max} , RMSE and PSNR are considered to get segmentation accuracy is given below

$$J_{\max} = w_1 \times J_{\max}(n) + w_2 \times \text{PSNR} \quad (6)$$

the weight functions are assigned as $w_1 = 2$, $w_2 = 1$.

4 Results and Discussions

This section provides the outcome obtained with multi-thresholding on the considered images. The proposed work is simulated using the MATLAB software. Well-known gray-scale image dataset of size 512×512 , such as Barbara, Gold hill, House, Mandrill, and Butterfly is reflected in this work. The proposed work is made with primary algorithm limits, search dimension selected as ' n ', stopping criteria is made static as J_{\max} and overall iteration is made 1000 for every threshold. In this work, robustness of Otsu-guided HSOA algorithm is experienced with standard gray-scale data tainted with Gaussian and salt-and-pepper noise values.

Table 1 the first column shows the representative images, column 2 presents Gaussian noise stained and column 3 shows Salt-and-pepper noise-included investigation images which are included for this study. In Table 1 column 4 and column 5 present the PSNR values of the noise-stained images. After noise addition, the histogram altitudes of the images transformed vigorously, which increases the complexity of the segmentation process also the histogram distribution of noise tainted images is maintained between 0 and $L-1$. Compared with the smooth

Table 1 512 × 512-sized grayscale dataset and PSNR (dB) value of noise-stained images

	Original image	Noise stained image		PSNR (dB) value	
		Gaussian	Salt and Pepper	Gaussian	Salt and Pepper
Barbara				19.6045	18.1456
Goldhill				19.5870	18.2714
House				20.2186	17.5030
Mandrill				19.4670	18.4341
Butterfly				19.4787	18.4407

image, finding the optimal thresholds for the noise-stained images are quite difficult and a challenging work in image segmentation field.

In the proposed work, initially, Otsu and HSOA-based segmentation practice is executed for Gaussian noise-stained Barbara. The thresholding process is repeated for 15 times and the mean values are recorded as the optimized value. Table 2 shows segmented output of the Gaussian noise-stained images for thresholds $t = \{2, 3, 4, 5\}$ and its equivalent cost function and quality measure values are presented in Table 3. Similar procedure is executed with the salt-and-pepper noise-tainted test images and the outcomes are exhibited in Tables 4 and 5. From the proposed result, the average CPU time required for the Gaussian noise-tainted image is around 128.16 s and the salt-and-pepper noise tainted is around 117.94 s. From Tables 3 and 5 values, it is perceived that, Otsu-enabled HSOA-based segmentation provides satisfactory results with acceptable values of RMSE and PSNR for the considered image dataset.

Table 2 Segmented Gaussian noise-stained image

	$T = 2$	$T = 3$	$T = 4$	$T = 5$
Barbara				
Goldhill				
House				
Mandrill				
Butterfly				

Table 3 Image quality measures for the segmented Gaussian noise images

	T	OF	Optimal threshold	RMSE	PSNR
Barbara	2	2513.65	115, 186	66.1561	11.7194
	3	2663.01	92,167, 214	47.2121	14.6497
	4	3069.09	64,115, 163, 230	34.3875	17.4028
	5	3110.49	55, 89, 131,191, 241	28.7095	18.9703
Gold hill	2	2002.19	122, 179	71.4662	11.0488
	3	2493.79	93,175,202	50.1274	14.1293
	4	2493.79	84,140,178,217	40.8911	15.8982
	5	2661.36	47,97,139,202,243	28.5555	19.0170
House	2	3042.21	98,172	55.7183	13.2108
	3	3571.05	56,141,185	39.8144	16.1300
	4	3664.35	41,103,153,216	32.8358	17.8039
	5	3671.59	46,101,139,171,239	26.3654	19.7101
Butterfly	2	1679.89	123, 186	69.2775	11.3190
	3	1912.98	83,139,202	43.6502	15.3311
	4	1848.24	77,125,181,215	35.6424	17.0915
	5	2264.32	56,101,14,175,240	26.5074	19.6635
Man drill	2	1104.49	126,201	69.4477	11.2976
	3	1694.39	83,144,206	43.1620	15.4288
	4	2121.18	88,136,179,224	36.6459	16.8503
	5	2072.82	85,121,139,200,242	34.7405	17.3141

Table 4 Segmented salt-and-pepper noise-stained image

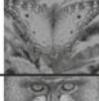
	$T = 2$	$T = 3$	$T = 4$	$T = 5$
Barbara				
Goldhill				
House				
Butterfly				
Mandrill				

Table 5 Image quality measures for the segmented salt-and-pepper noise images

	T	OF	Optimal threshold	RMSE	PSNR
Barbara	2	1857.56	120, 194	67.5302	11.5408
	3	1908.98	89,164, 205	46.2712	14.8246
	4	3208.04	73, 131,184,216	36.6446	16.8506
	5	3395.33	62, 93,138,181,236	29.7721	18.6546
Gold hill	2	1996.17	136,186	81.4050	9.9178
	3	2458.75	93,165,222	50.3458	14.0915
	4	2685.68	71,136,189,228	38.8902	16.3340
	5	2882.76	65,112,150,185,240	31.6276	18.1295
House	2	2107.81	100,166	57.5671	12.9273
	3	3247.69	80,172,202	43.1943	15.4223
	4	3969.19	56,103,176,228	32.6700	17.8478
	5	3839.55	30,86,132,201,238	27.0598	19.4843
Butterfly	2	1664.98	124,174	72.2234	10.9572
	3	1706.37	90,154, 192	47.5288	14.5917
	4	2058.19	58,118,163, 204	34.4348	17.3908
	5	2213.87	57,96,139, 231, 242	29.7852	18.6508
Mandrill	2	2411.71	125,206	69.4437	11.2982
	3	1972.66	93,153,232	44.0350	15.2548
	4	2036.63	88,131,168,24	36.1079	16.9787
	5	1530.91	48,94,144,179,250	28.0658	19.1673

5 Conclusion

This work represents the problem in determining multilevel thresholds ‘ T ’ for the noise-tainted test dataset of size 512×512 is employed based on Otsu’s guided HSOA. Here, number of thresholds are chosen as $T = \{2-5\}$. To validate the robustness of proposed procedure, the image dataset is tainted using Gaussian and salt-and-pepper noise values. J_{\max} , RMSE, and PSNR are considered to evaluate the prominence of segmentation procedure. The simulated results of the proposed work ratifies that, HSOA-based segmentation offers promising results on the Gaussian noise and salt-and-pepper-tainted image datasets. In future, the superiority of the proposed approach can be verified with other heuristic algorithms existing in the literature.

References

1. Akay, B.: A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding. *Appl. Soft Comput.* **13**(6), 3066–3091 (2013)
2. Ghamisi, P., Couceiro, M.S., Benediktsson, J.N.A., Ferreira, N.M.: An efficient method for segmentation of images based on fractional calculus and natural selection. *Expert Syst. Appl.* **39**(16), 12407–12417 (2012)
3. Ghamisi, P., Couceiro, M.S., Martins, F.M.L., Benediktsson, J.A.: Multilevel image segmentation based on fractional-order Darwinian particle swarm optimization. *IEEE Trans. Geosci. Remote Sens.* **52**(5), 2382–2394 (2014)
4. Pal, N.R., Pal, S.K.: A review on image segmentation techniques. *Pattern Recogn.* **26**(9), 1277–1294 (1993)
5. Sezgin, M.: Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **13**(1), 146–168 (2004)
6. Tuba, M.: Multilevel image thresholding by nature-inspired algorithms-A short review. *Comput. Sci. J. Moldova* **22**(3), 318–338 (2014)
7. Raja, N.S.M., Sukanya, S.A., Nikita, Y.: Improved PSO based multi-level thresholding for cancer infected breast thermal images using Otsu. *Procedia Comput. Sci.* **48**, 524–529 (2015)
8. Maitra, M., Chatterjee, A.: A hybrid cooperative-comprehensive learning based PSO algorithm for image segmentation using multilevel thresholding. *Expert Syst. Appl.* **34**(2), 1341–1350 (2008)
9. Rajinikanth, V., Aashiha, J.P., Atchaya, A.: Gray-level histogram based multilevel threshold selection with bat algorithm. *Int. J. Comput. Appl.* **93**(16) (2014)
10. Sathya, P.D., Kayalvizhi, R.: Modified bacterial foraging algorithm based multilevel thresholding for image segmentation. *Eng. Appl. Artif. Intell.* **24**(4), 595–615 (2011)
11. Abhinaya, B., Raja, N.S.M.: Solving multi-level image thresholding problem—an analysis with cuckoo search algorithm. *Inform. Syst. Design Intell. Appl.* pp. 177–186, Springer, India (2015)
12. Horng, M.-H., Liou, R.-J.: Multilevel minimum cross entropy threshold selection based on the firefly algorithm. *Expert Syst. Appl.* **38**(12), 14805–14811 (2011)
13. Rajinikanth, V., Couceiro, M.S.: RGB histogram based color image segmentation using firefly algorithm. *Procedia Comput. Sci.* **46**, 1449–1457 (2015)
14. Oliva, D., Cuevas, E., Pajares, G., Zaldivar, D., Perez-Cisneros, M.: Multilevel thresholding segmentation based on harmony search optimization. *J. Appl. Math.* **2013** (2013)

15. Geem, Z.W., Kim, J.H., Loganathan, G.V.: A new heuristic optimization algorithm: harmony search. *Simulation* **76**(2), 60–68 (2001)
16. Geem, Z.W.: Optimal cost design of water distribution networks using harmony search. *Eng. Optim.* **38**(03), 259–277 (2006)
17. Geem, Z.W., Lee, K.S., Park, Y.: Application of harmony search to vehicle routing, *Am. J. Appl. Sci.* **2**(12), 1552–1557 (2005)
18. Otsu, N.: A threshold selection method from gray-level histograms. *Automatica* **11**(285–296), 23–27 (1975)

An Approach for Visual Cryptography Scheme on Color Images

Chinmoy Kar, Suman Kalyan Kar and Sreeparna Banerjee

Abstract Visual cryptography, a popular and simple technique uses the characteristics of human vision to decrypt encrypted images. It requires no computation for decryption from the user end. This paper emphasizes on basic cryptographic scheme proposed by Naor and Shamir with changed shares. Here, we focus on different shares with different combination of CMY color scheme to check whether proper encryption has taken place. Different combinations of shares were employed and the contrast reduction for decrypted images was 50% for some of the combinations which were also achieved earlier. Most of these studies, however, concentrated on binary images but very few on grayscale and color images. In this paper, a technique is presented to apply the visual cryptography scheme on colored images.

Keywords Visual cryptography · RGB color model · CMY color model
Halftone

1 Introduction

In the modern twenty-first century, multimedia data are transmitted via Internet, thereby evoking an urgent need of data security simultaneously. In order to secure data from the unintended recipients, many fields of science have evolved like cryptography, steganography, visual cryptography. Cryptography is a science of encrypting data, where the data get disordered after encryption and can only be

C. Kar · S.K. Kar (✉)

Sikkim Manipal Institute of Technology, Majitar, India
e-mail: sumansit@gmail.com

C. Kar

e-mail: info.chinmoy@gmail.com

S. Banerjee

Maulana Abul Kalam Azad University of Technology, Kolkata, India
e-mail: sreeparnab@hotmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_53

501

recovered by a particular key [1]. In order to retrieve the data, it is very important for the retriever to know the appropriate key. Without knowing the appropriate key, the encrypted data can hardly be detected even if an unauthorized person has stolen the data [2]. Steganography is a science of hiding information such that the existence of message itself is hidden. Visual cryptography was mainly introduced for solving the problem of secret sharing. One of the main issues which arise in cryptographic techniques to be considered in cryptography is secret sharing. Secret sharing is one type of key establishment protocols [3]. Visual cryptography was introduced by Moni Naor and Adi Shamir in 1994. They proposed visual secret sharing scheme, in which an image is divided into ' n ' shares so that only someone with all ' n ' shares could decrypt the image, but the same is not possible with any ' $n-1$ ' shares of the original image [3]. Each share is imprinted upon a different transparent paper and decryption is performed by overlaying the shares when all ' n ' shares are overlaid, the original image gets visible. The main objective of data hiding is to communicate securely in such a way that the secret message is not visible to the attacker. That is unauthorized person should not be able to identify in any sense the target image.

2 Preliminaries

A basic idea of visual cryptography is presented below:

A secret image is divided into two component shares and stack of shares returns the secret image (Fig. 1).

The method presented here exploits two important modules in order to build shares; these are color decomposition principle and halftone technique.

Color decomposition generally deals with decomposing any color image into their respective color channels, based on the color models. This method incorporated CMY color model, where the three primary colors are cyan, magenta, yellow such that cyan, magenta, yellow are complementary colors to red, green, blue respectively. The printers use CMY color model.

In order to understand halftone technique, it is important to know that the printers such as dot matrix printers, laser printers, jet printers can either print a single pixel or dot of a particular color channel (particular color pixel) or not print (white pixel). There is no color shade in between that these can print [4]. However,

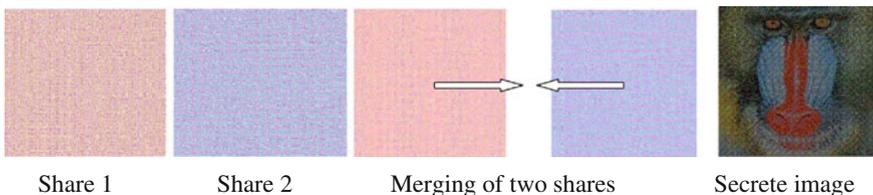


Fig. 1 Basic idea of visual cryptography scheme

these shades are attained by the printers, using a technique that exploits the density of net dots in order to produce gray levels in an image. Bright regions of image have sparse density of dots (as per the intensity of brightness). The dark region in an image is given by high density of these dots. Halftone is a method that exploits the technique of using density of dots to simulate gray level and transforms a gray level image into a binary image before processing [3]. This method maintains the advantage of traditional visual cryptography, which is the use of Human Visual System (HVS) for decryption. The shares and the decoded image obtained, using this algorithm, are larger in size than the original secret image and decoded image has a contrast reduction of 50%.

3 Related Works

Naor and Shamir [3] proposed new cryptography scheme where secret image is divided in shares and decryption can be done by human visual system. Hou [5] proposed two algorithms for visual cryptography of color images. Bhagate and Kulkarni [4] and Jena and Das [6] provided an overview of different cryptographic schemes while Hou et al. [7] provided new design for friendly visual cryptography which was different from conventional visual cryptography scheme.

4 Proposed Method

The proposed method can be decomposed further in three modules.

Module 1: Color Decomposition Technique:

The original secret image is read and it is then broken down into its constituent primary color channel components. This algorithm used CMY color model, thus the obtained component images are cyan, magenta, and yellow component images.

Module 2: Halftone Technique:

Floyd's halftone algorithm is used to convert the 3 grayscale component images, obtained in module 1, into their respective binary images. Diffuse the calculated error in neighboring pixels of $\text{pix}[m, n]$. The following formula can be used for calculating pixel value for four neighbor pixels.

$$\begin{aligned} \text{pix}[m + 1, n] &= \text{pix}[m + 1, n] + \text{error}^*7/16 \\ \text{pix}[m - 1, n + 1] &= \text{pix}[m - 1, n + 1] + \text{error}^*3/16 \\ \text{pix}[m, n + 1] &= \text{pix}[m, n + 1] + \text{error}^*5/16 \\ \text{pix}[m + 1, n + 1] &= \text{pix}[m + 1, n + 1] + \text{error}^*1/16 \end{aligned}$$

Module 3: Building Share Images:

Based on the pixel values of 3 halftone images (CH_{ij} , MH_{ij} , YH_{ij}), there are 8 possible combinations of the triplet (CH_{ij} , MH_{ij} , YH_{ij}). For each of these combinations, shares are built differently as shown in Table 1. The algorithm for building share image is given below:

1. Start.
2. Run a loop for number of rows.
 - 2.1. Run a loop for number of columns.
 - 2.1.1. Read image CH and store each pixel values in array cp (similarly read MH, YH and store their pixel values in arrays mp, yp respectively).

Table 1 Scheme of color cryptography [2, 3]

Mask	(C,M,Y)	Share 1	Share 2	Share 3	Stacked	Revealed
	(0,0,0)					1/2, 1/2, 1/2
	(0,0,1)					1/2, 1/2, 1
	(0,1,0)					1/2, 1, 1/2
	(0,1,1)					1/2, 1, 1
	(1,0,0)					1, 1/2, 1/2
	(1,0,1)					1, 1/2, 1
	(1,1,0)					1, 1, 1/2
	(1,1,1)					1, 1, 1

- 2.1.2. Build a 2×2 matrix called mask, which is 4 blocks, such that any two of these four blocks are black pixels and other two are white.
- 2.1.3 Build three such more 2×2 matrices, to create rest of three shares for each pixel P_{ij} of the composed image I (I is composed of CH , MH , YH). This is done based on the pixel values of CH_{ij} , MH_{ij} , YH_{ij} and the position of black pixels in the mask.
 - 2.1.3.1 If $CH_{ij} = 1$ (Cyan component is revealed), then fill one of the blocks in share image 1 and one block in share image 3 (both corresponding to the two different white blocks in the mask) with the cyan pixel.
 - 2.1.3.2. If $CH_{ij} = 0$ (cyan component is hidden), then fill one of the block in share image 1 and one block in share image 3 (both corresponding to the two different black blocks in the mask) with the cyan pixel.
 - 2.1.3.3. If $MH_{ij} = 1$ (magenta component is revealed), then fill one of the blocks in share image 1 and one block in share image 2 (both corresponding to the two different white blocks in the mask) with the magenta pixel.
 - 2.1.3.4. If $MH_{ij} = 0$ (magenta component is hidden), then fill one of the blocks in share image 1 and one block in share image 2 (both corresponding to the two different black blocks in the mask) with the magenta pixel.
 - 2.1.3.5. If $YH_{ij} = 1$ (yellow component is revealed), then fill one of the block in share image 2 and one block in share image 3 (both corresponding to the two different white blocks in the mask) with the yellow pixel.
 - 2.1.3.6. If $YH_{ij} = 0$ (yellow component is hidden), then fill one of the blocks in share image 2 and one block in share image 3 (both corresponding to the two different black blocks in the mask) with the Yellow pixel.
 - 2.1.4. Repeat step 2.1.3. unless all the pixels P_{ij} (CH_{ij} , MH_{ij} , YH_{ij}) of the composed image I are decomposed similarly.
 - 2.2. end.
 - 2.3. end.
3. Obtain and display all three shares along with mask.
4. These shares when overlapped produce the decoded image.
5. Stop.

5 Results and Discussions

This simulation of the algorithm has been done in MATLAB 7.10.0.499 and tested on a 64-bit 2.30 GHz Intel CORE i7 processor computer. Standard test images were taken as inputs (Figs. 2 and 3).

This algorithm works for color as well as grayscale images. It was applied on 20 images. Ten color images and ten grayscale images were taken as input secret images and displayed in two different tables, corresponding to which their



Fig. 2 Input image



Fig. 3 *Cyan* component, *magenta* component, *yellow* component



Fig. 4 *Cyan* halftone, *magenta* halftone, *yellow* halftone

respective shares are also built and was observed that the algorithm works satisfactorily. This algorithm works for images of all sizes (Fig. 4 and 5).

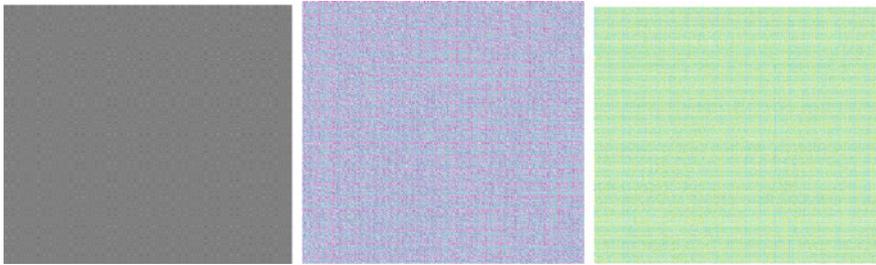


Fig. 5 Mask, Share 1, Share 2, Share 3

Fig. 6 Share 3, final image



6 Conclusion

In the implemented algorithm, a secret color image or grayscale image can be taken as input, followed by color decomposition and halftone techniques, and then shares are built such that on overlapping all these shares only, that particular secret image could be revealed (Fig. 6).

This algorithm gives a pixel expansion equal to 4, due to which the decoded image becomes four times the size of the original image. Such noisy shares when transmitted could draw the attention of hackers.

7 Future Scope of Work

This algorithm gives a pixel expansion equal to 4, due to which the decoded image becomes twice the size of the original image. There is no simple pixel expansion techniques present that could be used for improving upon the size factor of decoded image. Contrast reduction can be tried to increase. Such noisy shares when transmitted could draw the attention of hackers. This can be worked upon by using friendly visual cryptography.

References

1. Adhikari, A., Sikdar, S.: A new $(2, n)$ —visual threshold scheme for color images. In: Proceedings of the Indocrypt 2003. LNCS, vol. 2904, pp. 148–161. Springer (2005)
2. Mursi, M.F.M., Salama, M., Mansour, M.: Visual cryptography schemes: a comprehensive survey. *Int. J. Emerg. Res. Manage. Technol.* **3**(11) (2014). ISSN: 2278-9359
3. Naor, M., Shamir, A.: A visual cryptography. In proceedings of Advances in cryptology, EUROCRYPT 94. Lecture Notes in Computer Science (1994)
4. Bhagate, S.B., Kulkarni, P.J.: An overview of various visual cryptography schemes. *Int. J. Adv. Res. Comput. Commun. Eng.* **2**(9) (2012)
5. Hou, Y.C.: Visual cryptography for color images, Department of information Management. National Central University, Jung Li, Taiwan 320, ROC (Received 6 June 2002, accepted 26 Aug 2002)
6. Jena, P.C., Das, N.K.: A survey on visual cryptography using image encryption and decryption. *Int. J. Emerg. Technol. Adv. Eng. 2008 Certified J*, **3**(8) (2013)
7. Hou, Y.C., Quan, Z.Y., Liao, H.Y.: New design for friendly visual cryptography. *Int. J. Inform. Electron. Eng.* **5**(1) (2015)

A Novel Hybrid Approach for Influence Maximization in Online Social Networks Based on Node Neighborhoods

Gypsy Nandi, Uzzal Sharma and Anjan Das

Abstract Online social networks have nowadays become a buzzword for millions of users, who spend a lot of time online to remain in touch with other users by interacting online with them or to know about such other users' likings and views about a movie, product, place, and so on. Thus, there is a considerable amount of information being spread among such online users which help in maximizing influence for a particular product, movie, holiday destination, etc. But, the main question remains as to how to identify the top few best influential users so as to help in promotion of any such a product or movie. This paper discusses about influence maximization in online social networks and also studies efficient techniques for the same. Considering time complexity as the prime factor for influence maximization techniques, this paper also aims to propose a new algorithm *DegGreedy* which yields a much faster output than the two basic standard influence maximization algorithms.

Keywords Influence maximization · Online social networks · High-degree centrality · Greedy algorithm · *DegGreedy* algorithm

1 Introduction

Nowadays, accessing the Internet has become a part and parcel of everyone's life, mainly due to the interest gained by online users in Online Social Networks (OSNs). These OSNs play a vital role in influencing the online users for making any

G. Nandi (✉) · U. Sharma
Assam Don Bosco University, Guwahati, Assam, India
e-mail: gypsy.nandi@gmail.com

U. Sharma
e-mail: druzzalsharma@gmail.com

A. Das
St. Anthony's College, Shillong, Shillong, Meghalaya, India
e-mail: anjan_sh@rediffmail.com

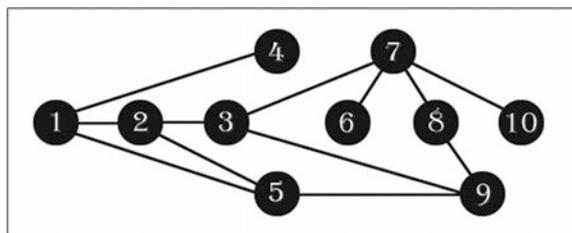
kind of choices or decisions [2]. As for example, one online user can influence his/her online friends to make a decision on which holiday destination to visit, which songs to listen, which item to purchase, and so on. As a result the marketing agencies or companies are keen on finding the top few influential online customers who can be targeted for promotion of a particular product, movie, place, and so on.

Any OSN can be graphically represented as shown in Fig. 1, in which the OSN consists of ten nodes or online users and connected by links or edges indicating the friendship or ties between them. Several researchers in the field of computer science have studied about influential maximization in OSNs to find few top best influential users in an OSN which are often called as *seed sets*. These seed sets play a major role in spreading information or influence with relate to a particular topic or product. Hence, influence maximization in OSNs is currently one of the most discussed topics in the field of Online Social Network (OSN) mining, in which the basic computational task is to select a set of initial seed set that can best influence more users in the network.

In general, the influence maximization problem can be considered as at first choosing a preliminary set of target nodes to initiate the process of diffusion. Let $I(S)$ denote the influence of a group of nodes Z . Hence, $I(Z)$ indicates the probable number of active nodes that is expected at the completion of the diffusion process. One of the important input parameters in this problem is k , where k refers to the total number of seed sets (most active nodes) that is required to be found. However, for most of the existing models that are considered, it is NP-hard to create the finest set for influence maximization.

Domingos and Richardson [1] are considered to be the pioneers of the influence maximization problem and had suggested an algorithm for the same. Later, Kempe, Kleinberg, and Tardos [3] have proposed the independent cascade (IC) model and the Linear Threshold (LT) model as the two basic influence maximization models, based on which the greedy algorithm was implemented to find the resultant seed set. However, as far as efficiency is concerned, the time required to generate the result using greedy algorithm was extremely high. To solve this problem, another algorithm, namely the cost-effective lazy forward selection (CELFF) algorithm was found [4] and the maximum influence arborescence (MIA) model was developed by Chen et al. [5] which generated a much faster result than the general greedy algorithm. In [3], the various diffusion models used for influence propagation have been extensively studied. Also, a method for influence maximization has been proposed using

Fig. 1 Example of an online social network for the illustration of nodes and edges



automata learning with the primary consideration to efficiently deal with large online social networks.

Chen et al. [7] proposed the local directed acyclic graph (LDAG) algorithm for the linear threshold model but their algorithm is not suitable for all kind of social media sites. Recently, a study on topic-based influence maximization problem in OSNs has been into limelight [8, 9] where emphasis is given on analyzing the spread of influence from a user to another neighboring user based on a specific topic. Recently, a study on time-sensitive influence maximization (TSIM) problem has been carried out by mainly considering the dependence of time on the information value [12]. Also, research on location-based influence maximization [14] is being carried out for efficient results. Thus, several researches on the topic of influence maximization are being carried out mainly to get better efficiency and generate efficient number of spread of influence among users in OSNs.

The later sections of the paper are structured as follows. Section 2 focuses on the two state-of-the-art existing influence maximization algorithms, namely, the degree centrality algorithm and the general greedy algorithm, and also a new proposed influence maximization algorithm, namely the hybrid *DegGreedy* algorithm. Section 3 shows the experimental results by comparing the spread of influence among these three algorithms. Section 4 emphasizes on a brief conclusion of the paper and the scope of future work.

2 Discussion on Standard Influence Maximization Algorithms

This section, at first, discusses the high-degree centrality algorithm used for the influence maximization problem which mainly emphasizes on the structural properties of an OSN. Next, the famous general greedy algorithm is discussed which relies mainly on the computation of influence spread and uses a standard propagation model for maximizing influence in the networks. Few such propagation models include the linear threshold (LT) model, the independent cascade (IC) model, and the weighted cascade (WC) model. The IC model works on the probability of a node n to succeed in activating a neighbor p which is based on the set of p 's neighbors that have already attempted to do so whereas in the LT model, a weight $w_{n,p} \in [0:1]$ is assigned to a node n based on the influences of its each incoming neighbor p [3]. A special case of IC model is the WC model in which each edge or link from a node m to n has a success probability $p = 1/d_n$ for triggering n , where d_n is the degree of the node n [13]. There are also several other proposed diffusion models for influence maximization such as decreasing cascade model, submodular threshold model, and so on.

All these propagation techniques have their own advantages and disadvantages. Keeping in mind the limitations of these two standard influence maximization algorithms that are often used to find the most influential users in an OSN, a new

hybrid *DegGreedy* algorithm has been proposed that aims to provide a larger spread of influence in an OSN with comparatively less running time.

In all the three algorithms, an OSN is considered which is represented as an undirected or directed graph $G = (V, E)$, where V represents a set of vertices or nodes in the OSN and E represents a set of edges between two vertices that indicates a connection or relationship between the two vertices.

2.1 The High-Degree Centrality Algorithm

Since many years, the interconnectedness of nodes in OSNs has been a fundamental issue for OSN mining. There are several ways of finding the strength of interconnection among nodes. One of the simplest approaches is to find for each node to how many nodes, it is directly connected in a network. This is known as the measure of degree centrality (DC), which is a useful indicator to analyze the strength of influence in an OSN. In [6], the DC measure is proposed and explained by Nieminen.

The basic concept of DC is that nodes having higher direct connections with other nodes are considered to have higher degree. Hence, nodes having maximum degree should result in larger influence spread in an OSN. Considering that there are N nodes in a network, for a node n , its degree centrality can be defined as

$$DC(n) = \sum_{i=1}^N \sigma(u_i, n) \quad (1)$$

where $\sigma(u_i, n)$ is defined as follows:

$$\begin{aligned} \sigma(u_i, n) &= 1, \text{ if } u_i \text{ and } n \text{ are connected;} \\ &= 0, \text{ otherwise} \end{aligned}$$

As a result, if the centrality score $DC(n)$ for a node n is higher than node x , it indicates node n has more direct contacts than node x . If we consider Fig. 1, the DC of each node in the network will be as displayed in Table 1. From Table 1, we can find that node 7 has the highest DC score of 4 and is thus considered to be the best networked member for spread of influence in an OSN, followed by nodes 1, 2, 3, 5, and 9.

The biggest advantage of using DC-based influence maximization is that the time complexity of the DC heuristic algorithm is considerably much better than any

Table 1 Measurement of degree centrality

Node n	1	2	3	4	5	6	7	8	9	10
DC(n)	3	3	3	1	3	1	4	2	3	1
Rank	2	2	2	8	2	8	1	7	2	8

other centrality algorithms such as betweenness centrality, closeness centrality, and so on. It is also significantly much better than the general greedy algorithm which is discussed in the next subsection. Hence, for large-scale OSNs, DC-based influence maximization can be chosen as an efficient technique for influence maximization in order to get faster results. Algorithm 1 illustrates the step-wise process to select the k best seeds in an OSN graph G . Ultimately, the top k seeds are stored in a set S which is considered as the output for the algorithm.

Algorithm 1: DegreeCentrality(G, k)

```

1: Initialize  $S = \phi$ 
2: for each vertex  $v$  do
3:   compute its degree  $d_v$ 
4: end for
5: for  $i = 1$  to  $k$  do
6:    $S = S \cup \{v_{d_i}\}$ 
7: end for
8: output  $S$ 

```

Here, v_{d_i} (in line 6) indicates the vertex having top i th highest degree centrality value. Once the set S having top k influential seeds is determined (as explained in Algorithm 1), a measure for estimating the total spread of influence in the OSN is carried out using some standard diffusion model.

Considering the enormous amount of data that needs to be analyzed for an OSN, DC-based influence maximization is an efficient and easy technique for determining the best seed set in the network. However, the output generated by implementing the DC heuristic algorithm results in a considerably lesser influence spread as compared to the output produced by the general greedy algorithm as found in the experimental result section. Also, another major disadvantage of DC-based technique for influence maximization is that indirect contacts are not considered at all. By taking into consideration only the direct links or edges, this technique fails to crawl the entire set of valuable data available in the entire graph.

2.2 The General Greedy Algorithm

The general greedy algorithm is also based on some influence propagation models, such as the linear threshold (LT) model, the independent cascade (IC) model, or the weighted cascade (WC) model and it computes the spread of influence by carrying out the simulation process for these models [3]. There are several other important influence maximization algorithms that are based on these models to find the seed set that consists of the top best influential users in an OSN.

 Algorithm 2: GeneralGreedy (G, k) [7]

```

1: Initialize  $S = \emptyset$ ,  $S_{im} = 20000$ 
2: for  $i = 1$  to  $k$  do
3:   for each vertex  $v \in V \setminus S$  do
4:      $s_v = 0$ 
5:     for  $i = 1$  to  $S_{im}$  do
6:        $s_v += |\text{Random}(S \cup \{v\})|$ 
7:     end for
8:      $s_v = s_v / S_{im}$ 
9:   end for
10:   $S = S \cup \{\arg \max_{v \in V \setminus S} \{s_v\}\}$ 
11: end for
12: output  $S$ 

```

The two input required for the general greedy algorithm is the OSN graph G and the number of expected influential users k . In Algorithm 2, initially an empty set S is used. $\text{Random}(S)$ denotes the random simulation process carried out using one of the influence propagation models to find the seed set from the origin set S . The final output includes the set S having top k influential seeds. In each iteration, the algorithm adds exactly one node v into the set S in such a way that this node v together with the current set S maximizes the spread of influence in G (Line 10). Thus, for each vertex $v \in S$, the simulation process is carried out S_{im} times repeatedly to estimate the influence spread of $S \cup \{v\}$ (Lines 3–9).

Experiments using the greedy algorithm show that there is a significant performance gain over several other widely used influence maximization algorithms that only use the structural properties of a network. However, when enormous amount of data in an OSN is needed to be analyzed, this algorithm will not prove to be efficient mainly because the result generated is very slow and time consuming.

2.3 The DegGreedy Centrality Algorithm

It has been experimentally found that the high-degree centrality algorithm yields considerably faster output but is not very efficient as far as spreading of influence over the nodes is concerned. The general Greedy algorithm, on the other hand, gives a better result than the high-degree centrality algorithm in terms of the spread of influence of nodes over the network but to generate the result, it takes a significantly much longer time. This creates a problem when dealing with considerably large-scale OSNs. The proposed novel *DegGreedy* algorithm is a compromise between the high-degree centrality algorithm and the general Greedy algorithm, and the output is generated in a much lesser time with high spread of influence. Algorithm 3 describes the new hybrid *DegGreedy* centrality-based algorithm to select the top k influential nodes in the network G .

Algorithm 3: *DegGreedy*(G, k, m)

```

1: Initialize  $T = \emptyset$ 
2: for each vertex  $v$  do
3:     compute its degree  $d_v$ 
4: end for
5: for  $i = 1$  to  $m$  do
6:      $T = T \cup \{v_{di}\}$ 
7: end for
8: Initialize  $S = \emptyset, S_{im} = 20000$ 
9:   for  $i = 1$  to  $k$  do
10:    for each vertex  $v \in T \setminus S$  do
11:         $s_v = 0$ 
12:        for  $i = 1$  to  $S_{im}$  do
13:             $s_v += |\text{Random}(S \cup \{v\})|$ 
14:        end for
15:         $s_v = s_v / S_{im}$ 
16:    end for
17:     $S = S \cup \{\arg \max_{v \in T \setminus S} \{s_v\}\}$ 
18:  end for
19:  output  $S$ 

```

Initially, in the algorithm discussed above, the degree of each vertex in the graph is computed as shown in steps 2–4. The value of m , as mentioned in step 5, should be user-defined and should be considerably larger than the value of k , where k refers to the number of top influential users that needs be obtained. The optimum size of m as has been experimentally found out can be twice or thrice the value of k for a decent value of k . For instance, if k is a one digit number, correspondingly the value of m can be considered as $3k$. And, if k is a two digit number, correspondingly the value of m can be considered as $2k$. Sequentially, the set T will consist of a set of top m vertices having highest degree values as shown in steps 5–7. Now, considering T as the set of nodes for the social network graph (which will be definitely much smaller than the original network graph), we need to find seed set S containing top- k nodes (Steps 9–19). This will result in a much faster execution time compared to the general greedy algorithm and will yield comparatively much better result than the high-degree centrality algorithm. Hence, this novel technique of influence propagation considers the scalability issue which is of a much concern when dealing with large-sized social network graphs.

3 Experimental Evaluation

In our experiment, we used two publicly available real-world datasets—the dolphin dataset and the Santa Fe institute collaboration dataset. The dolphin dataset includes an undirected dolphin social network having numerous connections among various dolphins [10]. This dataset contains 62 distinct nodes and 159 edges which indicates the ties among all dolphins. The entire dataset has been represented in a graph as shown in the Fig. 2. The unique nodes have been labeled from 1 to 62. The size and color of nodes have been given based on the values of degree of nodes. The various related structural details of the dolphin dataset are mentioned in Table 2.

The Santa Fe Institute Collaboration dataset contains a network of 118 nodes and 200 edges which have been used by many authors in the field of OSN analysis [11]. The entire dataset has been represented in a graph as shown in the Fig. 3. The unique nodes have been labeled from 1 to 118. The other related information of the dolphin dataset is mentioned in Table 3.

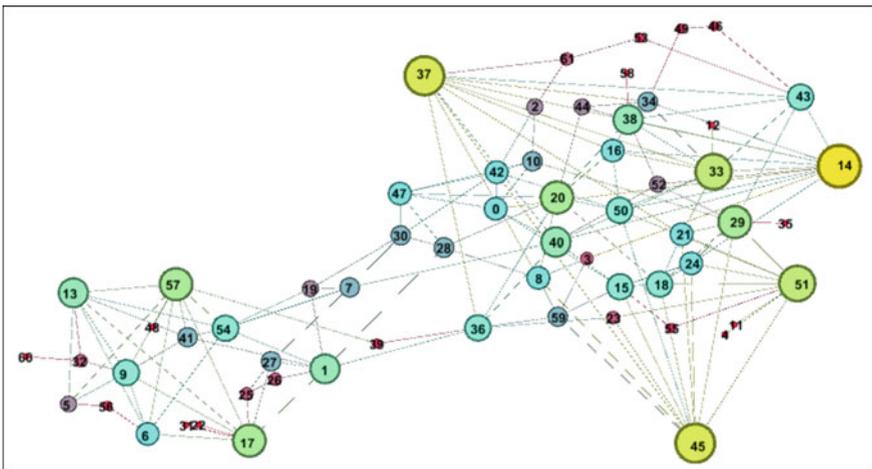


Fig. 2 The *dolphin* dataset represented as a graph

Table 2 Structural features of the *dolphin* dataset

Dataset	<i>dolphin</i>
Graph type	Undirected
#Nodes	62
#Edges	159
Average degree	5.129
Maximal degree	12
Average path length	3.453

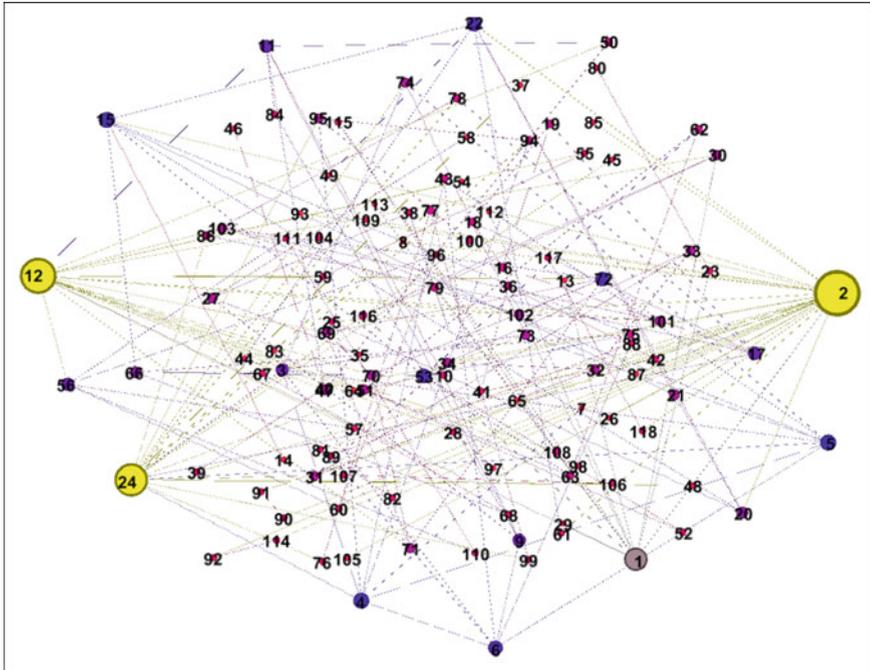


Fig. 3 The *Santa Fe Collaboration* dataset represented as a graph

Table 3 Structural Features of the *Santa Fe Collaboration* dataset

Dataset	<i>Santa Fe Collaboration</i>
Graph Type	Undirected
#Nodes	118
#Edges	120
Average Degree	3.39
Maximal Degree	29
Average Path Length	5.021

Experiments were conducted for the four different influence maximization algorithms namely, the high-degree heuristic algorithm, the general greedy algorithm, the *DegGreedy* centrality algorithm and the random generation algorithm on both the datasets. The random generation algorithm is considered as a baseline method for comparing with others mentioned influence maximization algorithms. In this method, nodes are chosen uniformly at random. These experiments mainly aim at illustrating the performance comparison of these algorithms with respect to the amount of spread of influence using the Weighted Cascade model. The value of m considered for *DegGreedy* technique differed from $2k$ to $3k$ depending on whether the value of k is two digit or single digit respectively. Higher values of

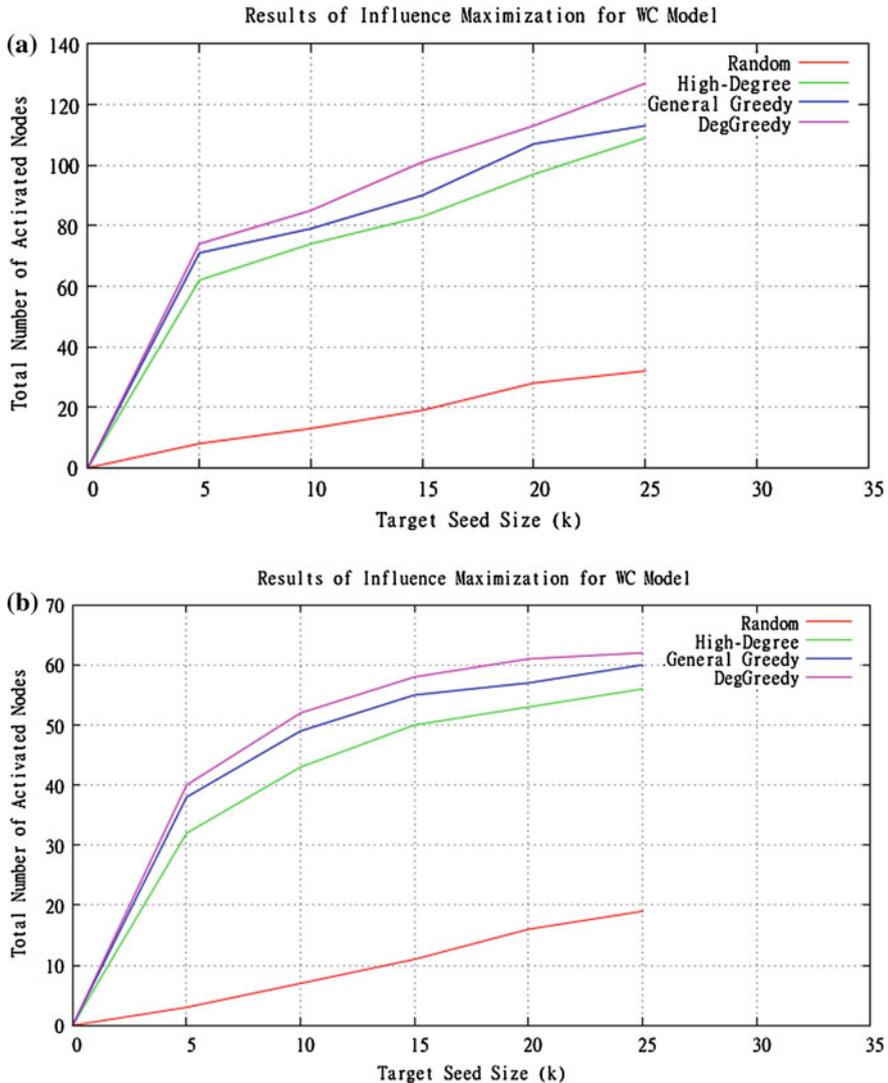


Fig. 4 Comparing spread of influence using random, high-degree, general greedy, and *degGreedy* algorithms for the **a** dolphin dataset **b** Santa Fe Collaboration dataset

m have also been considered during the experimental evaluation but it did not give any significantly better results for both the datasets.

In Fig. 4a, b, the target seed size is represented in *x*-axis which denotes the number of influential nodes being considered and the total number of activated nodes is represented in the *y*-axis which denotes the total number of nodes that got influenced for a given seed set size. As can be analyzed from Fig. 4a, b, the

proposed *DegGreedy* algorithm outperforms the general greedy heuristic algorithm as well as the high-degree centrality algorithm which indicates that relying exclusively on the structural properties of a graph is not a better marketing strategy but we need to also explicitly consider the dynamics of information in an OSN. The execution time for the *DegGreedy* algorithm is also considerably much better than the general Greedy algorithm and that is a major solution to the scalability issue involved in dealing with social network graphs.

4 Conclusions

This paper gives a comparative analysis of a novel influence maximization technique and the two basic standard techniques used for effective influence maximization in OSNs. The paper also provides algorithms and experimental results to analyze the same. The novel influence maximization technique proposed is a simple technique based on node neighborhoods that proves efficient in terms of scalability and higher spread of influence. We believe that the algorithms discussed in this paper will remain the simplest and most basic algorithms for studying and analyzing the concept of influence maximization in OSNs and will hence help a researcher in this field to get a preliminary idea about the same. As a future work, some comparative study of the latest algorithms with these basic algorithms will be studied for gain in the highest influence spread in OSNs.

References

1. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge discovery and data mining, pp. 57–66 (2001)
2. Nandi, G., Das, A.: Efficient spread of influence in online social networks. *Smart Innov Syst. Technol.* **32**(6), 293–300 (2015)
3. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: *Theory of Computing*, vol. 11, no. 4, pp. 105–147 (2015)
4. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., Briesen, J., Glance, N.S.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 420–429 (2007)
5. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM), pp. 88–97 (2010)
6. Nieminen, J.: On the centrality in a graph. *Scand. J. Psychol.* **15**, 332–336 (1974)
7. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 199–208 (2009)
8. Liu, L., Tang, J., Han, J., Jiang, M., Yang, S.: Mining topic-level influence in heterogeneous networks. In: Proceedings of the 19th ACM International Conference on Information and knowledge management (CIKM), pp. 199–208 (2010)

9. Barbieri, N., Bonchi, F., Manco, G.: Topic-aware social influence propagation models. In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining (ICDM), pp. 81–90 (2012)
10. Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., Dawson, S.M.: The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations, *Behav. Ecol. Sociobiol.* 396–405 (2003)
11. The Santa Fe Institute collaboration dataset, Retrieved from <http://dsec.pku.edu.cn/~jliu/> (2016). Accessed on Jan 2016
12. Mohammadi, A., Saraei, M., Mirzaei, A.: Time-sensitive influence maximization in social networks. *J. Inf. Sci.* **41**(6), 765–778 (2015)
13. Akrouf, S., Meriem, L., Yahia, B., Eddine, M.N.: Social network analysis and information propagation: a case study using flickr and youtube networks. *Int. J. Future Comput. Commun.* **2**(3), 246–252 (2013)
14. Zhou, T., Cao, J., Liu, B., Xu, S., Zhu, Z., Luo, J.: Location-based influence maximization in social networks. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1211–1220 (2015)

Diagnosis of Ground Glass Opacity in HRCT Lung Images Using High-Intensity Pixels

Punal M. Arabi, Nanditha Krishna, T.P. Prathibha
and N. Vamsha Deepa

Abstract According to the World Health Organization survey done in 2008, approximately 9.5 million people were affected by lung diseases worldwide. Some of the major lung diseases are asthma, chronic cough, asbestosis, chronic obstructive pulmonary disease (COPD), bronchiectasis, cystic fibrosis, influenza, lung cancer, pneumonia, pulmonary hypertension, sarcoidosis, sleep apnea. It is observed that approximately 40,000 people die from lung diseases every year. Diagnosis and treatment of the disease at right time play a crucial role in survival rate and ensure the quality of life to the affected person. Ground glass opacity refers to the blurred area of increased attenuation in the lung with visible vascular and bronchial markings. The lungs affected by diseases like pulmonary edema, pneumocystis, hypersensitive pneumonitis, carinii pneumonia, desquamative pneumonitis have the visibility of ground glass structures. This paper proposes a method to identify ground glass opacity in HRCT lung images by calculating the number of high-intensity pixels present in the lung image of interest. A computer-aided method of identifying the ground glass opacification is proposed based on the number of high-intensity pixels present. This method is tested on a set of 10 normal and ground glass opacity affected lung images. The results obtained show that the proposed method has an accuracy of 90%.

Keywords Diagnosis of ground glass opacity · HRCT lung · Mean pixel intensity level · Number of high-intensity pixels

P.M. Arabi (✉) · Nanditha Krishna · T.P. Prathibha · N.V. Deepa
Department of Biomedical Engineering, ACSCE, Bengaluru, India
e-mail: arabi.punal@gmail.com

Nanditha Krishna
e-mail: nanditha13@gmail.com

T.P. Prathibha
e-mail: prathi.tp@gmail.com

N.V. Deepa
e-mail: vamshi.deepa@gmail.com

1 Introduction

Chronic obstructive pulmonary disease is the fourth leading cause of death in the world [1]. Respiratory disorders like pneumoconiosis, pulmonary fibrosis, hypersensitivity pneumonitis, farmer's lung disease, occupational asthma, and occupational lung cancer are caused due to inhaling dust, silica, asbestos, hay, mold, spores, toluen diisocyanate, phthalic anhydride, sawdust, and hexavalent chromium [2].

Ground glass pulmonary opacity has a higher lung density and also has a persistent visibility of bronchial walls and vascular structures. Ground glass structures are termed as chronic interstitial disease and acute alveolar disease [3].

High-resolution computed tomography (HRCT) images contain lots of information about lung textures including bronchus, pulmonary veins, arteries and is the most preferred method for diagnosing a number of lung diseases [4].

Computer-Aided Diagnosis (CAD) system acts as an aid in the medical field as it helps the doctors for a better diagnosis of diseases with higher degree of accuracy in a shorter period of time. CAD works as a bridge between the radiologist and the input images under test [5].

Hosseini et al. [6] proposed an approach for automatic generation and tuning of parameters of type-2 fuzzy logic system which was applied for lung nodule classification of the CAD system. Al-Fahoum et al. [7] developed an automated intelligent system for nodule detection and displayed the size of the detected cancer.

Gomathi and Thangaraj [8] proposed a CAD system for automated detection of pulmonary parenchyma and nodule detection of lung cancer as applied to computed tomography examinations for early detection of lung cancer.

Bennet et al. [9] proposed a method for ground glass detection using banks of Gabor filter and corresponding masks for pattern identification with frequency spectrum analysis. Abbas et al. [10] proposed a segmentation algorithm for lungs on HRCT scan images to segment the whole CT slice into different regions based on dynamic and optimal thresholding techniques. Silva et al. [11] proposed a method for automatic segmentation of pulmonary regions in X-ray CT images.

This paper proposes an image processing technique to identify the ground glass opacity in CT images of lung using the mean pixel intensity values and high-intensity pixel pattern.

2 Methodology

The proposed method is tested on normal and ground glass HRCT lung images. The acquired images undergo a preprocessing stage, where the lung portion is segmented and is enhanced by contrast enhancement technique; noise is removed by low-pass filter. The method proposed here for computer-aided diagnosis of ground glass opacity in lung images consists of 3 steps.

Step 1: Fixing up the threshold value

A set of 10 HRCT lung images are taken, among which 5 images are of normal lung and 5 images are of ground glass structure lung.

Since the acquired image is a RGB image, it is converted into a gray image and then the lung portion is extracted from the thoracic region using binary thresholding technique. The extracted left and right portions are divided into 4 quadrants each and labeled as 1, 2, 3...8 in total. For the segmented right and left lung, the mean pixel intensity values are obtained separately. The average values of the of mean pixel intensity values of the left and right portions are found out for normal lung and as well as ground glass lung. Using these average values, the threshold value is fixed.

Step 2: Finding high-intensity pixels

The pixels having intensity values greater than the fixed threshold value are termed as high-intensity pixels. The left and right lung portions are divided into 4 quadrants each, giving 8 quadrants in total. The number of high-intensity pixels present in all the quadrants of left and right lung of all the images under analysis is found out.

Step 3: Fixing the reference value

Using the number of higher intensity pixels present in the normal lung images and ground glass lung images, the reference value is fixed. The highest value among the number of higher intensity pixels of left and right portions of normal lung quadrants is found out; similarly the lowest value among the number of higher intensity pixels of left and right portions of ground glass lung quadrants is found. Then the average of these two values is found out and is fixed as reference value.

2.1 Decision Rule

Check the number of higher intensity pixels present in the image quadrants.

If 6 quadrants of the total of 8 quadrants have the number of higher intensity pixels more than the reference value, the image under test is termed as a suspicious image for ground glass opacity lung image.

If 6 quadrants of the total of 8 quadrants have the number of higher intensity pixels lesser than the reference value, the image under test is labeled as a normal lung image.

3 Results

Figure 1 shows the block diagram of the proposed method. The set of five normal lung CT images taken for experimentation is shown in Fig. 2. The set of five ground glass lung CT images taken for experimentation is shown in Fig. 3. Table 1 shows the mean pixel intensity values of left and right lung. Table 2 shows the number of pixels having intensity >27 for normal lung quadrants. Table 3 shows number of pixels having intensity >27 for ground glass opacity CT lung images.

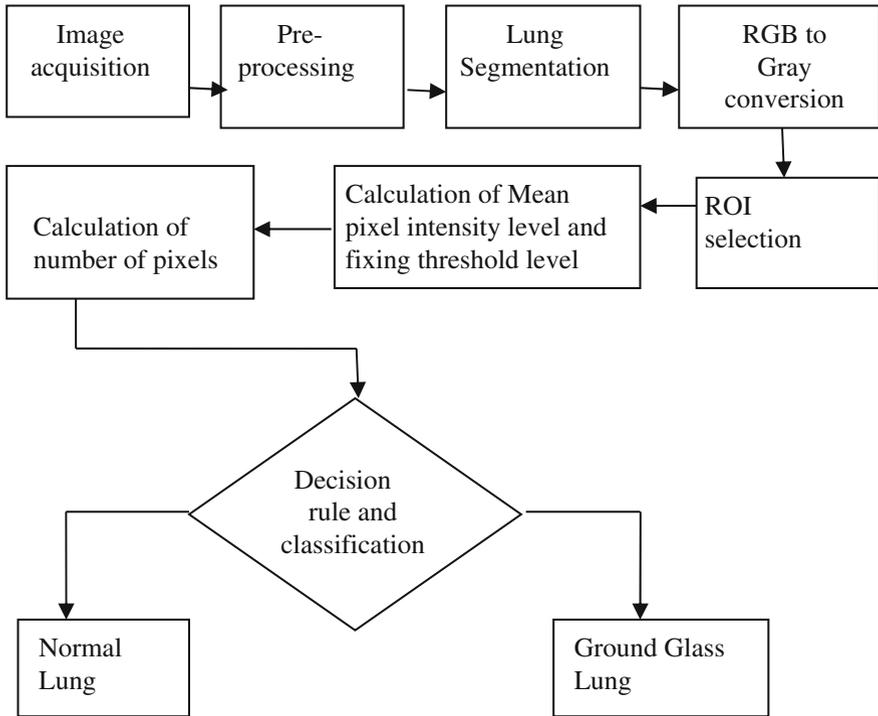


Fig. 1 Block diagram



Fig. 2 Set of normal lung CT images



Fig. 3 Set of ground glass structure lung CT images

Table 1 Mean pixel intensity values of left and right lung

Type of lung	Left	Right
Normal lung 1	15.62	14.49
Normal lung 2	18.17	16.57
Normal lung 3	12.09	11.96
Normal lung 4	24.16	18.67
Normal lung 5	11.57	11.76
Average value of normal lung	16.32	14.69
Ground glass lung 1	38.09	33.08
Ground glass lung 2	40.53	34.46
Ground glass lung 3	63.33	56.07
Ground glass lung 4	35.84	35.92
Ground glass lung 5	28.63	25.76
Average value of ground glass lung	41.28	37.05

Table 2 Number of pixels having intensity >27 for normal lung quadrants

Normal lung images	LL1	LL2	LL3	LL4	RL1	RL2	RL3	RL4
Normal lung 1	213	2361	264	3173	2310	292	2681	399
Normal lung 2	1011	2630	1117	2956	2409	1203	2286	1533
Normal lung 3	556	4815	452	4539	3743	633	4243	964
Normal lung 4	2425	6691	2418	4908	4356	2736	3978	2105
Normal lung 5	1819	1785	2766	2516	263	2465	1877	2580

Table 3 Number of pixels having intensity >27 for ground glass opacity CT lung

Ground glass lung images	LL1	LL2	LL3	LL4	RL1	RL2	RL3	RL4
Ground glass lung 1	14,208	8213	17,128	13,000	6112	15,293	6719	14,855
Ground glass lung 2	43,584	24,830	44,657	26,341	38,727	28,450	38,645	21,840
Ground glass lung 3	3476	3397	3185	3806	3014	3871	3744	3142
Ground glass lung 4	6723	6457	6522	3270	4665	5838	5777	7624
Ground glass lung 5	9734	8680	9394	9217	7985	8550	7338	9308

3.1 Calculations

Threshold value

$$Z = \frac{\text{Average Value of Normal lung(left + right)} + \text{Average Value of Ground glass lung(left + right)}}{4} \quad (1)$$

$$Z = \frac{16.32 + 14.69 + 41.28 + 37.05}{4}$$

$$Z = 27.33 \sim 27$$

where Z represents the threshold value. For calculating the threshold value, average values of left and right lung of both normal and ground glass lung are taken from Table 1, Eq. (2) is used to calculate reference value.

Reference Value

$$Y = \frac{\text{Maximum value of number of higher intensity pixels present in the normal lung quadrants} + \text{Minimum value of number of higher intensity pixels present in the ground glass lung quadrant}}{2} \quad (2)$$

$$Y = \frac{6691 + 3041}{2}$$

$$Y = 4852 \sim 5000$$

where Y is the reference value. For calculating the reference value, we take the respective minimum and maximum higher pixel intensity values from Table 2.

4 Discussion

A set of 10 images of normal lung and affected lung with ground glass opacity is taken for analysis. The acquired images undergo preprocessing stage in which the lung portion is segmented, filtered and enhanced. From the average mean pixel intensity values which are calculated for the left and right lungs of both normal and ground glass lung images, the threshold value is obtained and is found to be equal to pixel intensity level 27 (Eq. 1). The left and right lungs are divided into 4 quadrants each and labeled as 1, 2, 3... 8 in total. The number of higher intensity pixels which have intensity value > pixel intensity level 27, are then calculated for every quadrant and tabulated (Tables 2 and 3). From the tabulated values the

maximum value of number of higher intensity pixels present in the normal lung quadrants and minimum value of number of higher intensity pixels present in the ground glass lung quadrants are identified; using these values the reference value is calculated (Eq. 2). Based on this reference value, a decision about whether the lung is normal or suspicious one for ground glass opacity is made; if 6 or more quadrants of the total of 8 quadrants have higher intensity pixel values more than the reference value, the image under test is termed as a suspicious image for ground glass opacity lung image.

On the other hand if 6 or more quadrants of the total of 8 quadrants have higher intensity pixel values lesser than the reference value, the image under test is labeled as a normal lung image. The results obtained show that all the normal lung images are rightly identified by the proposed method while one ground glass lung image (Ground glass lung no. 3 of Table 3.) is wrongly identified as normal lung. Out of 10 images 9 images are rightly identified. The accuracy of the method is found to be 90% here.

5 Conclusion

The method proposed here identifies the number of higher intensity pixels present in the lung regions as the pixels having intensity value more than 27 where as 27 is the average mean pixel intensity value of all the lung images taken for experimentation. The left and right lung portions of every lung image are divided into four quadrants each giving eight quadrants in total. The number of higher intensity pixels present in every lung region is found out and tabulated. Using these values, a reference value is calculated and the decision rule to identify ground glass opacity lung is framed. The experimental results obtained show that the proposed method has an accuracy of 90%. However, more number of images are to be tested to confirm the accuracy. A deeper statistical analysis in fixing the threshold value and reference value may improve the accuracy of the proposed method.

Acknowledgements The authors thank the Management and Principal of ACS College of engineering, Mysore road, Bangalore for permitting and supporting to carry out the research work.

References

1. Nabil Al Lawati, M.D., FitzGerald, J.M.: Acute exacerbation of chronic obstructive pulmonary disease. *BC Med. J.* **50**, 138–142, (2008)
2. Uchiyama, I.: Chronic health effects of inhalation of dust or sludge. *J. Jpn. Med. Assoc.* **56**, 91–95 (2013)
3. Yamaguchi, T., Ashizawa, K., Nagaoki, K., Hayashi, K.: High-resolution CT evaluation of ground-glass opacity in diffuse lung disease. *A Rev. Art.* **43**, 1–11 (1998)

4. Tong, T., Huang, Y., Wang, X., Feng, H., Li, C.: Automatic extraction of three dimensional lung texture tree from hrct images. In: International Conference on Bioinformatics and Biomedical Technology, pp. 198–202 (2010)
5. Anshad, P.M., Kumar, S.S.: Recent methods for the detection of tumor using computer aided diagnosis—a review. In: International Conference on Control Instrumentation Communication and Computational Technologies, pp. 1014–1019 (2014)
6. Hosseini, R., Dehmeshki, J., Barman, S., Mazinani, M., Qanadli, S.: A genetic type-2 fuzzy logic system for pattern recognition in computer aided detection systems (2010)
7. Al-Fahoum, A.S., Jaber, E.B., Al-Jarrah, M.A.: Automated detection of lung cancer using statistical and morphological image processing techniques. *Sciedu Press. J. Biomed. Graph. Comput.* **4**, 33–42 (2014)
8. Gomathi, M., Thangaraj, P.: Automated CAD for lung nodule detection using CT scans. In: International Conference on Data Storage and Data Engineering, pp. 150–153 (2010)
9. Bennet, M.A., Nair, R.R., Mahalakshmi, V., Janakiraman, G.: Performance and analysis of ground-glass pattern detection in lung disease based on high-resolution computed tomography. *Indian J. Sci. Technol.* **9**(2) (2016)
10. Abbas, Q., Khan, M.T.A., Farooq, A., Celebi, M.E.: Segmentation of lungs in HRCT scan images using particle swarm optimization. *Int. J. Innovative Comput. Inf. Control.* **9**, 2155–2165 (2013)
11. Silva, J.S., Silva, A., Santos, B.S.: Lung segmentation methods in X-ray CT images. In: Proceedings of the SIARP, pp. 583–598 (2000)

Coefficient Random Permutation Based Compressed Sensing for Medical Image Compression

R. Monika, S. Dhanalakshmi and S. Sreejith

Abstract Compression of medical data remains challenging because of the loss in clarity of compressed images. In medical field, it is necessary to have high image quality in region of interest. This paper presents a Compressed Sensing (CS) method for the compression of medical images. Coefficient random permutation (CRP) based CS is used in this paper for compression of medical images. The different image block has different sparsity. If the nearby pixel values in a block have stronger correlation, then they are strongly sparse, otherwise they are said to be weakly sparse. The main objective of using this method is to provide high quality compressed images thereby maintaining a balanced sparsity of the reconstructed images. As a result performance gain would be high. Experimental results show that CRP based CS helps achieving better PSNR values even with lesser number of measurement values.

Keywords CRP · Compressed sensing · Gaussian matrix · Orthogonal matching pursuit · DWT

1 Introduction

Medical images are captured and transferred between hospitals for review by physicians. Also these images are required to be stored for the reference of physicians and patients in future. An efficient compression algorithm is required for effective storage of data for future use. Conventional compression (CC) helps in

R. Monika (✉) · S. Dhanalakshmi
ECE Department, SRM University, Kattankulathur, Chennai, India
e-mail: moni.rajendran@gmail.com

S. Dhanalakshmi
e-mail: dhanalakshmi.s@ktr.srmuniv.ac.in

S. Sreejith
School of Electrical Engineering, VIT, Vellore, India
e-mail: sreejith.s@vit.ac.in

reducing the amount of data. However they do not provide significant reduction in data rate. In order to achieve a good compression ratio, compressed sensing techniques can be used.

CS achieves good compression. A signal can be acquired at a rate which is much lesser than nyquist sampling rate by using CS. However in Conventional CS Sparsity of all blocks of the image is not even. Hence images are reconstructed with poor quality. Generally smoother region have stronger sparsity while edges have weaker sparsity. Due to imbalance in sparsity few regions (i.e., highly sparse) will be reconstructed well while other regions (i.e., poorly sparse) are poorly reconstructed. This in whole degrades the quality of the reconstructed images. Also conventional CS cannot be applied for medical image compression as it requires good quality reconstruction with balanced sparsity.

In order to achieve balanced sparsity in all blocks of the image, CRP is used. Balanced sparsity is achieved by means of random permutation of the coefficients. CRP method in DWT domain helps in improving CS sampling efficiency and performance.

2 Related Works

Candes and Wakin [1] explained that certain signals can be recovered successfully from fewer samples or measurements than traditional methods by using CS. Sampling and compressing data are performed simultaneously in CS.

Donoho [2] explains about the basic concepts of Compressed sensing (CS) and also representing an image with reduced no of samples still maintaining the quality of the recovered image. CS can be performed only if the signal is highly sparse or compressible. Most of the data acquired can be discarded without much loss in data.

Gao et al. [3] explains that different image block has different sparsity levels. Due to imbalance in the sparsity the reconstruction process of the blocks results in degradation of the image quality. Smoother regions are reconstructed well while the texture regions are poorly reconstructed. In order to balance the sparsity random permutations of the coefficients are performed. As a result all the blocks have even sparsity resulting in higher PSNR values. Reconstructed images have better visual quality that of the conventional CS process.

Gao et al. [4] proposed two simple techniques for the block CS (BCS) with discrete cosine transform (DCT). The techniques are effective in improving the CS sampling efficiency and CS sampling performance. This process of spreading the sparsity throughout the block helps in better reconstruction of all portions of the image. As the result the visual quality of the reconstructed image is greatly enhanced. Also High PSNR values are obtained.

In this paper CRP based CS in DWT domain is investigated for medical image compression application. The rest of the paper is organized as follows: overview of

compressed sensing is provided in Sects. 3 and 4 discusses about the CRP based CS in DWT domain, CRP for medical image compression is discussed in Sect. 5, experimental results are provided in Sects. 6 and 7 concludes the paper.

3 Overview of CS

CS is a mathematical framework that describes about accurate recovery of a signal from reduced measurements [6]. In order to perform CS, the signal should be sparse and compressible. This consists of projection of signal vectors into set of chosen projection vectors. Consider a signal of dimension N . Transform is applied to the signal and consider only k non-zero elements are present in the transformed matrix. In such case M measurements can be chosen from a total of N measurements.

Consider a N -dimensional signal, $x \in R^N$. Let Ψ be certain domain in which the signal is sparsely represented, Also consider that there are K non-zero coefficients in the Ψ domain ($K \ll N$).

In order to take CS measurements, we consider a M by N measurement matrix Φ which obeys the restricted isometric property.

The compressed sensing is given by

$$y = \Phi x \tag{1}$$

In this paper Gaussian random matrix is used as measurement matrix. The Gaussian matrix has random values and it has zero expectation and variance ($1/m$), where m is the mean. Also it has high RIP. Applying Gaussian matrix to entire image does not suit the need. Hence the image is divided to blocks and Gaussian matrix is applied to each block. To improve the efficiency block based sampling strategy is followed [7]. Orthogonal matching pursuit (OMP) [5] is used as the reconstruction algorithm. OMP is the simplest and easiest of the reconstruction algorithms.

4 Coefficient Random Permutation

The different image blocks have different sparsity. Sparsity level is not even. As a result only fewer regions are reconstructed well while other regions are poorly reconstructed. To balance the sparsity level random permutation of the coefficients are performed.

4.1 CRP Sampling Scheme [3]

The Sparsity balancing scheme is described as follows

1. Transform is applied to image block: $t_i = dct2(x_i)$, $i = 1, 2, \dots, B$, where x_i represents blocks and B represents number of blocks.
2. Coefficients corresponding to same frequency are then regrouped $r_j = \{t_i^j, i = 1, 2, 3, \dots, B\}$, $j = 1, 2, \dots, nB2$, where t_i^j is the j th coefficient of the i th block.
3. Regrouped coefficients belonging to same frequency are randomly permuted $p_j = p(r_j)$.
4. Permuted coefficients are rebuild $b_i = \{p_j^i, j = 1, 2, 3, \dots\}$, $i = 1, 2, 3, \dots, B$ where p_j^i is the i th permuted coefficient block.
5. CS is performed for rebuild vector.
6. OMP reconstruction algorithm is applied for each sampled vector.
7. The recovered coefficients are regrouped $r_j = \{b_i^j, i = 1, 2, \dots, B\}$, $j = 1, 2, \dots$, where b_i^j is j th coefficient in i th recovered vector.
8. Inverse permutation is performed for recovered frequency components $p_j = p(r_j)$, $j = 1, 2, \dots$.
9. DCT coefficients for each recovered block is rebuild and IDCT is applied.

5 CRP for Medical Image Compression

CRP can reconstruct image with good quality. CRP based medical image compression scheme is depicted below.

Figure 1 depicts the block diagram for medical image compression. CRP is applied to medical image followed by CS, finally image is reconstructed with better quality. CRP could retain the minute details of the image.

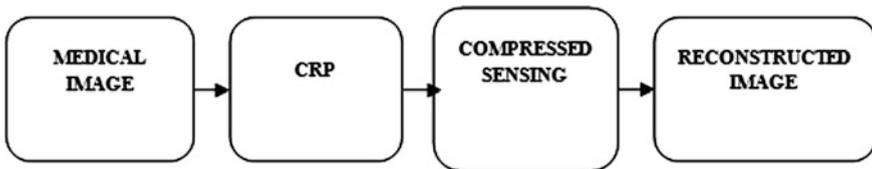


Fig. 1 Block diagram

6 Experimental Results

Atherosclerosis is solidifying of veins because of hypertension, progressive intimal accumulation of lipid, protein, and elevated cholesterol. It causes stroke, heart attack, stenosis and many other conditions which may lead to death. Compression of these images is essential in order to reduce the storage space and in order to store them effectively for future reference [8]. Therefore the images chosen for testing CRP algorithm is normal and abnormal plaque images.

Matlab R2012a is used for software simulation. The test images are medical images. Block size of 8×8 is tested in the simulation. Sparse binary random matrix is chosen as the measurement matrix. OMP algorithm is used for reconstruction. Transform used is DWT. The results are tabulated and compared with CCS.

The PSNR performance comparisons for various sparsity values of 8, 6, 4 and 2 for abnormal image are tabulated in Table 1.

Better is the PSNR value, better is the quality of reconstructed image. For a block size of 8×8 , the CRP has achieved an increase in PSNR value when compared to that of CCS. Also number of sparse elements chosen from each block varies from only 2 to 8. For such smaller number of sparse elements good PSNR value is achieved for CRP.

It can also be seen that as the sparsity is reduced there is decrease in PSNR values. Even with sparsity as 2 CRP has achieved good PSNR values and there is almost 3db difference in PSNR values for CRP when compared to CCS.

The graphical representation of PSNR values for various Sparsity values and techniques are shown in Fig. 2. Sparsity is chosen in the range of 2–8 in steps of 2. For smaller sparsity good PSNR difference is achieved.

It could be clearly seen from the graph that CRP is more efficient than CCS in achieving better PSNR values (Table 1).

Better SSIM values are obtained by using CRP (Table 2). SSIM value of 1 is reachable only if the original image and the reconstructed image look identical. SSIM values are close to 1 which indicates better reconstruction of the images.

To check the consistency of the techniques normal and abnormal medical images were chosen and PSNR and MSE of CCS and CRP techniques were found and are summarized in Tables 3 and 4 respectively.

CRP has reduced MSE values than CCS. MSE is inversely proportional to PSNR. Hence reduced MSE values indicate better reconstruction of images.

Table 1 PSNR comparison for various sparsity

Atherosclerosis plaque (480×320), Block size = 8×8				
Technique	PSNR (dB)			
	8	6	4	2
CCS	27.8409	26.3005	24.2554	22.6078
CS with CRP	29.0032	28.9460	27.3342	25.9908

Fig. 2 Techniques versus PSNR

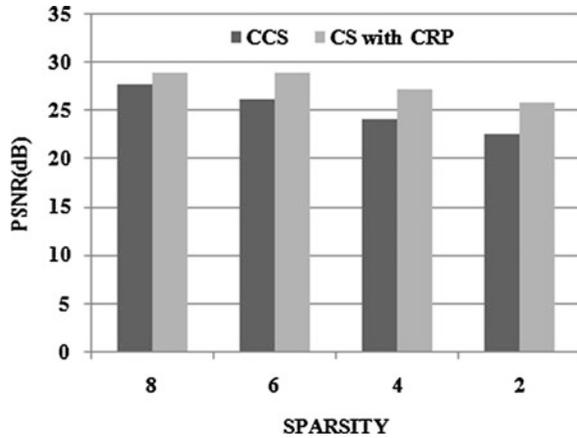


Table 2 SSIM comparison for various sparsity

Aretheroscleorsis plaque (480×320), Block size = 8×8

Technique	SSIM			
	8	6	4	2
CCS	0.7127	0.7130	0.6687	0.5823
CS with CRP	0.8602	0.8100	0.7912	0.6883

Table 3 PSNR comparison for various estimated images

Sparsity = 8, Block size = 8×8

Technique	PSNR (dB) Normal image	PSNR (dB) Abnormal image (fibroclacic plaque)
CCS	23.2202	29.7465
CS with CRP	26.6431	32.0605

Table 4 MSE comparison for various estimated images

Sparsity = 8, Block size = 8×8

Technique	PSNR (dB) Normal image	PSNR(dB) Abnormal image (fibroclacic plaque)
CCS	68.5910	135.4164
CS with CRP	40.4607	115.3669

The visual quality comparison of various medical images for CCS and CRP techniques considering only 8 sparse values from each of the blocks of size 8×8 are shown in Fig. 2.

Figure 3a–c shows the normal image, Aretheroscleorsis plaque, fibroclacic plaque images. Figure 3d–f from left to right shows the reconstructed images using CRP.

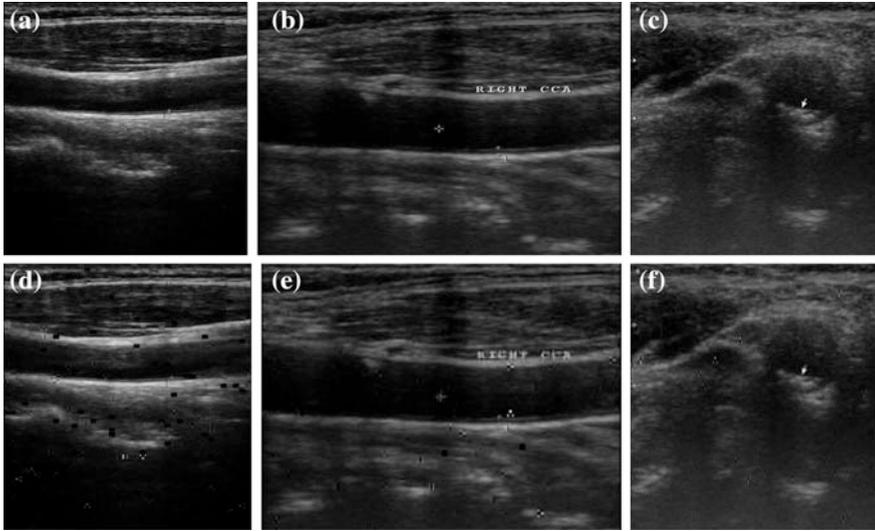


Fig. 3 Visual quality comparison. Original images. **a** Normal image. **b** Atherosclerosis plaque. **c** Fibroclastic plaque. **d–f** From *left to right* recovered images by CRP

7 Conclusion and Future Works

Image reconstruction using CRP in DWT domain for medical images is investigated in this paper. They are efficient in better reconstruction of all blocks of the image irrespective of their region i.e. smooth region or the edges. Therefore images can be reconstructed with very good PSNR even with only few Sparse values. Hence by using the CRP, enhanced performance can be achieved with reduced sparsity and complexity. The future work is to explore techniques which can further improve PSNR values.

References

1. Candes, E.J., Wakin, M.B.: An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008)
2. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006)
3. Gao, Z., Xiong, C., Zhou, C., Wang, H.: Compressive sampling with coefficients random permutations for image compression. In: *Proceedings of International Conference on Multimedia and Signal Processing*, pp. 321–324 (2011)
4. Gao, Z., Xiong, C., Ding, L., Zhou, C.: Image representation using block compressive sensing for compression applications. *J. Vis. Commun. Image Represent.* **24**(7), 885–894 (2013)
5. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* **53**(12), 4655–4666 (2007)

6. Deng, C., Lin, W., Lee, B.S., Lau, C.T.: Robust image compression based on compressive sensing. In: Proceedings of International Conference on Multimedia and Expo (ICME), pp. 462–467 (2010)
7. Gan, L.: Block compressed sensing of natural images. In: Proceedings of International Conference on Digital Signal Processing, pp. 403–406 (2007)
8. Sanei, S., Phan, A.H., Lo, J.L., Abolghasemi, V., Cichocki, A.: Compressive sensing approach for progressive transmission of images. In: Proceedings of International Conference on DSP, pp. 1–5 (2009)

A Flexible Scheme to Fault Detection for Electrical Assets Using Infrared Thermography

Deepak Kumar, Amit Kumar and M.A. Ansari

Abstract This paper approaches an infrared thermography methodology that can help to find fault and diagnosis for the electrical equipment. It uses noncontact and nondestructive technology. This technique is fast and reliable for inspecting the system without any interruption. In the field of electrical area maintenance, reliability of transmission and distribution system is one of the most critical issues and it suffers from some problems like loose connection, corrosion, and unbalanced loads. In this paper noninvasive methods are used to monitor the temperature of electrical assets and analysis of the hot region, and it used the watershed transform for the image segmentation and color-based segmentation which separates the red, green, blue area of the image. Dark red hot region areas are detected and the maximum temperature is 194 °C and the reference temperature is 40.8 °C of fuse cabinet. The proposed method is to detect the hotness and hot region of the electrical assets for fault detection.

Keywords Infrared thermogram · Fuse cabinet · Temperature measurement
Segmentation

1 Introduction

In the field of electrical area maintenance, reliability of transmission and distribution system is one of the most critical issues and it suffers from some problems like loose connection, corrosion, and unbalanced loads. These types of problems cause increase in impedance to current and increase resistive heating. If unchecked, heat

D. Kumar (✉) · A. Kumar · M.A. Ansari
School of Engineering, Gautam Buddha University, Greater Noida, India
e-mail: Deepu1796@gmail.com

A. Kumar
e-mail: abes.amit@gmail.com

M.A. Ansari
e-mail: mahmaditr@gmail.com

increases and a point builds at which the connection can melt and break the circuit, and in some cases can create fire. Thus, the thermal disorder of electrical equipment increases the working time, loading inequality, and increasing resistance or over current out from the bad contact. Buy this the whole transmission and distribution system will be under disorder. To protect and detection of these type of case Infrared thermography (IRT) or digital thermal image processing can be used that locks the heating point. It quickly locates the hot point and determines condition, by this system should be repaired [1].

The infrared thermography is used in many fields such as health indicator in medical field and malfunction sign in mechanical and electrical field. Infrared thermography is a non-contact technology means it does not use contact with the source of heat and is based on infrared radiation (electromagnetic radiation which have higher wavelengths than the visible light). Any object emits infrared radiation whose temperature above the absolute zero (i.e., $T > 0$ K) such radiation cannot be seen by the human eyes. The infrared device transfers the emitted radiation into electronic signal. The visible image and infrared image are different to each other, visible image represents the reflected light of the scene, and in the infrared image source is scene which can be observed by the infrared camera without light. Infrared camera acquired image and converted into visible image that is called as thermogram [2].

It is used noncontact and nondestructive technology. IRT can be used in power system because this technique is fast and reliable for inspecting the system without any interruption. It is most valuable diagnostic tool used for maintenance and performs important task to prevent electricity blackout and reduces the electrical equipment damage of both distribution and transmission system. So in power system thermography camera provides fast and most correct, accurate method to survey. It quickly locates the hot region and determines to check the problem and give result to operator that the equipment should be repair or not. By these advantages it increases system reliability and reduces the maintenance cost. So IRT can be used in power system to monitor the thermal behavior of equipment [3].

Many methods related to fault detection for electrical power system by the researchers Kumar and Saini [4] proposed a signal estimation technique used in wavelet transform for signal de-noising. It used new thresholding method which is based on generalized Gaussian distribution and analyzes the statistical parameters of wavelet sub-band which gives best performance better than the other methods of de-noising. Panigrahi and Verma [5] uses the methodology of the electrical equipment which faces the problems such as loose connection, unbalance load, and corrosion. The heat can be built on loose connection and melt and breaking the connection of circuit and equipment, so infrared inspection are designed to locate the problems of electrical system and reduces the failure in transmission and distribution system. Ahmad et al. [6] proposed a passive monitoring network using fuzzy rule-based neural network. Where the fuzzy system containing membership function used for fault diagnosis with the help of neural network and the combination of both systems called hybrid system. Shitole et al. [7] presented a Time of flight diffraction technique for nondestructive testing which is used the advance

image processing for detection the weld flaws. This technique reduces time, cost, and error in weld flaws in steel structure.

Varma et al. [8] developed design which is microcontroller-based thermo scanning and have real time monitoring system. It uses the computer with wireless connection, receive the real time video from source, transfer to the system where it applies some function, and store the thermal data in MYSQL database. Ibarahim and Amani [9] presented an approach for network monitoring by using neuro fuzzy approach, the hybride system has the dual knowledge database, one has the membership function value, and the other is rule-based formulation by the network admistration. Novizon et al. [10] proposed method to access the artificial neural network to classify ZnO Arrester condition with the temperature profile, ambient temperature and humidity, and third harmonic leakage current. Thermal data using thermal imaging as well as the leakage current third harmonic component were used as tools to investigate the surge arrester aging condition. Laurentys Almeida et al. [11] use the strategy for surge arrester image segmentation using image processing to classify the operative condition of arrester as faulty, normal, light, and suspicious. That aims to separate the information from image and detect and diagnose surge arrester fault using thermo-vision technique.

2 Infrared Thermogram Noninvasive Technique

The block diagram shows the Infrared thermogram noninvasive (IRTNI) system. This includes the IR thermogram, watershed transformation, color-based segmentation, and the other characteristics of an image (Fig. 1).

2.1 Process of Fault Diagnosis

There are the following steps for fault diagnosis as:

- Data acquisition
- Digital image segmentation process
- Watershed transformation
- Calculate standard deviation and thresholding of image
- Result.

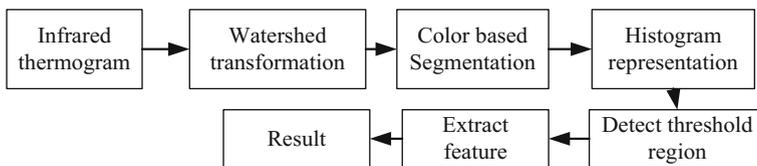


Fig. 1 Flow diagram of IRTNI

2.1.1 Data Acquisition

Data acquisition can be done through the infrared thermo-viewer camera. The camera has the spectral range from 7.5 to 13 μm and focal plane array detector type.

2.1.2 Digital Image Segmentation

Image segmentation is typically used to locate objects and borderline in images. Dividing an image into multiple parts, which is the process of image segmentation. This is typically used to determine objects or other important information in digital images. The segmentation objective is the image distribution, its set of parts, and objects. The location of object and image borderline are to be determine by the image segmentation process [4]. For a successful segmentation, whole information related to the problem is needed. The most useable function utilized for the division are: Point or Line discovery, Edge identification, Gradient administrators, Laplacian, Hough light Simple or Adaptive Threshold, district developing, and watershed transformation.

2.1.3 Watershed Transformation

The watershed transformation is the best technique for image segmentation; it is also called as flooding process. For the investigation of image, preparing a watershed of a gray-scale image is comparable to the idea of a catchment basin of a height map. The gray level represents the height of topography surface of an image [4, 5]. Basically a watershed is the line to determine the particular region of image where the water drop will fall. It provides closed contours for image segmentation and also takes less time for segmentation as compared to other method.

For image segmentation, first take the image edge detection, here the gradient topology is used, for edge detection use sobel operator. So, the original image gets converted into gray image and then apply sobel operator for edge detection or gradient topology [12]. Gradient image represents the edge strength. So, use the multi-scale edge detector to find morphological gradient of an image, expressed mathematically as following:

$$\text{GT}(f) = \frac{1}{n} \sum_{i=1}^n [(f \oplus a_i) - (f \ominus a_i) \oplus a_{i-1}] \quad (1)$$

where a_i denotes the group of square structure elements and n is scale and sizes are $(2i + 1) * (2i - 1)$ pixels. $(f \oplus a)$ represents the gray-scale dilation and $(f \ominus a)$ represent the gray-scale erosion [13].

The gray-scale closing and opening for a gray image f by structuring element a are to be expressed mathematically denoted as:

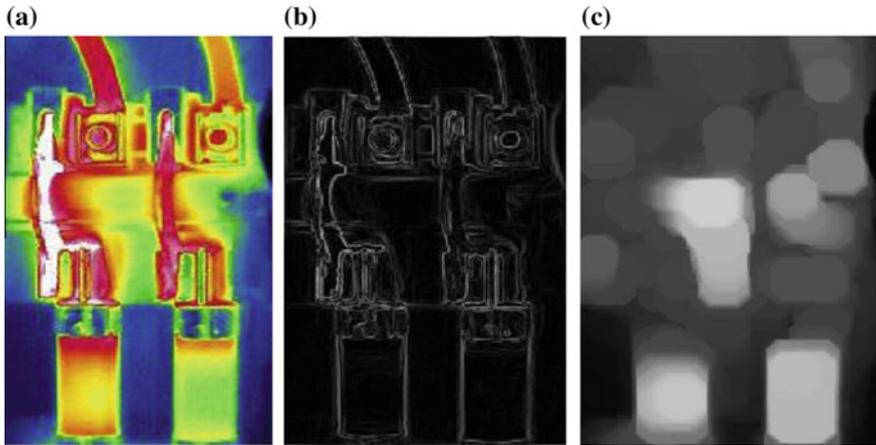


Fig. 2 a IRT image of fuse cabinet. b Gradient magnitude. c Image opening

For opening

$$f \cdot a = (f \ominus a) \oplus a \tag{2}$$

For closing

$$f \cdot a = (f \oplus a) \ominus a \tag{3}$$

and then used the opening- and closing-based image reconstruction which is effective than the other function and it does not affecting the shape of the object.

For fault detection and diagnosis, take an image of fuse cabinet by the IR thermogram camera which is shown below.

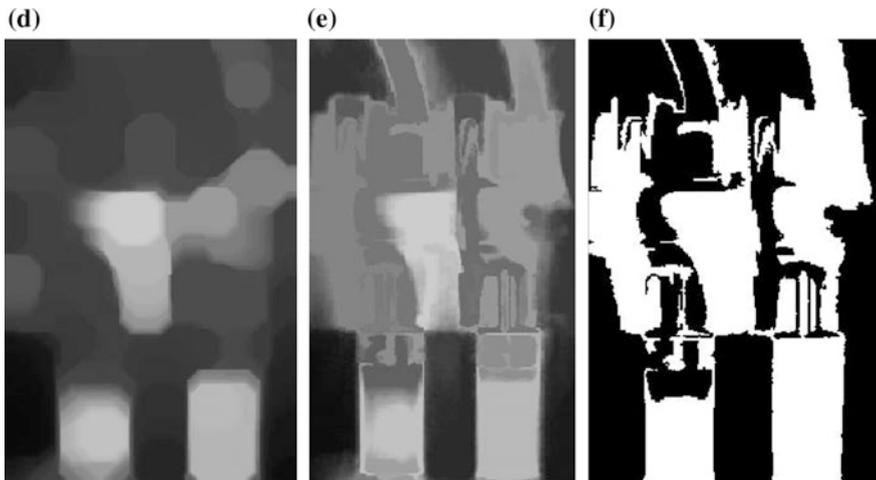


Fig. 3 a Opening-closing of image. b Opening-closing reconstruction image. c Threshold image

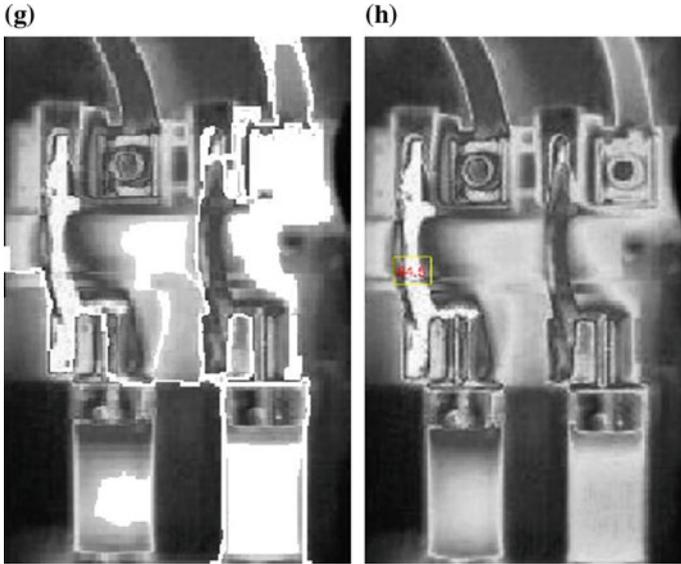


Fig. 4 a Segmented image. b Standard deviation of region

Figure 2a shows the original image of fuse cabinet and the RGB color region is clear seen. Now apply the watershed transform on to the image. Here Fig. 2b which is shows the gradient magnitude of image, means it shows the outline between objects. Figures 2c and 3a shows the opening and opening-closing image of gray level of the image which is shows the eater drops as blur and sharp lights and the solution of this show in Fig. 3b as opening-closing by reconstruction of image.

Figure 3c shows the thresholded image and Fig. 4a shows the outline of segmented region. The segmented region shows that here can be the possibility to come fault because this region has come into the red color.

2.1.4 Calculate Standard Deviation and Thresholding of Image

The thresholding of image which can be very helpful to find the largest value of standard deviation image. The thresholding can be representing mathematically as [6, 7]:

$$g(i,j) = \begin{cases} 1, & f(i,j) > T \\ 0, & f(i,j) \leq T \end{cases} \quad (4)$$

where T is threshold, the image background and object pixels replace by 0 and 1 and it can also say as 1 represent the color desire region and 0 undesired region of image. $f(i,j)$ and $g(i,j)$ shows the original image and thresholding image.

The standard deviation (σ) of image can be mathematically expressed as [9, 14]:

$$(\sigma) = \sqrt{\sum \frac{P_{x,y} \times P_{x,y}}{x \times y} - (\mu^2)} \tag{5}$$

μ is the mean of pixel values which can be mathematically expressed as:

$$\mu = \sum \frac{P_{x,y}}{x \times y} \tag{6}$$

where $P_{x,y}$ is pixel value of segmented image at x, y location. The standard deviation value of segmented image is shown in Fig. 4b the value of standard deviation is 64.5 by this region has high noise level. And the intensity of this region should be high.

The histogram performances of the IRT image are shown in Figs. 5, 6 and 7

Fig. 5 Histogram of *red plane* of IRT image

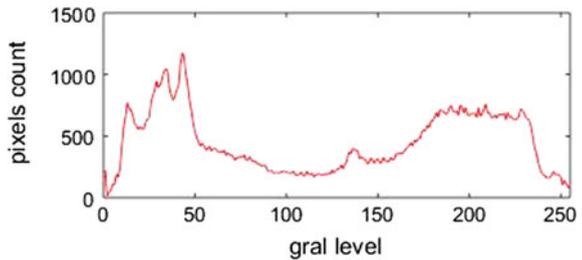


Fig. 6 Histogram of *green plane*

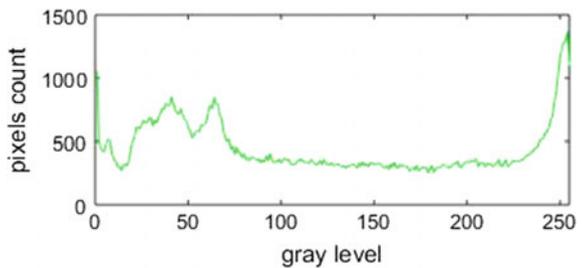
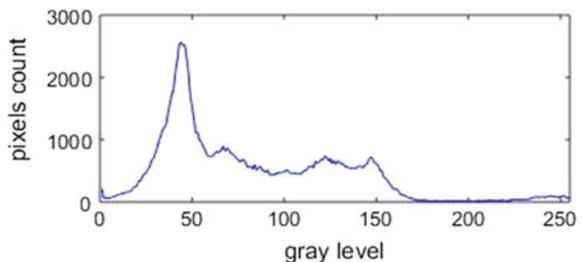


Fig. 7 Histogram of *blue plane*



3 Color-Based Segmentation

The color-based segmentation is the process where the image changed into hot color map then apply the smooth filter for noise removal from the image, the figures are shown below of color-based segmentation [8, 15] (Figs. 8 and 9).

It is filtered mask of the neighborhood pixels which is slid across the image for smoothing. The average of neighborhood pixels is used to replace the smoothing pixels, the average of the neighborhood pixels are expressed as [11]:

$$\bar{a}(i, j) = \frac{1}{k} \sum_{(l, m) \in P_{ij}} a(l, m) \quad (7)$$

where P_{ij} defines the neighborhood pixels coordinates in an RGB color image which is shown in Fig. 10.

Figure 10 shows the three level RGB color map image where the dark red region are shown. This part of fuse cabinet which have very high temperature and the other region are consider to be reference region which have temperature less then to the dark hot red region.

Fig. 8 Hot color map of image

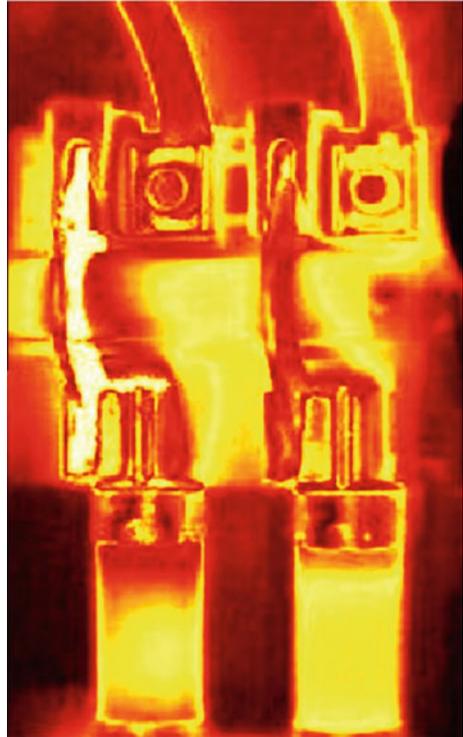
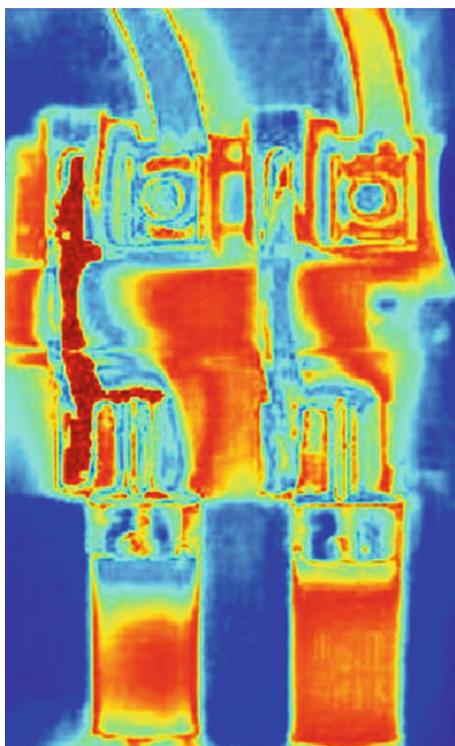


Fig. 9 Filtered image



Fig. 10 Three levels RGB image



4 Result and Discussion

The image features are to be extract in from of area and the intensity which shown in Table 1.

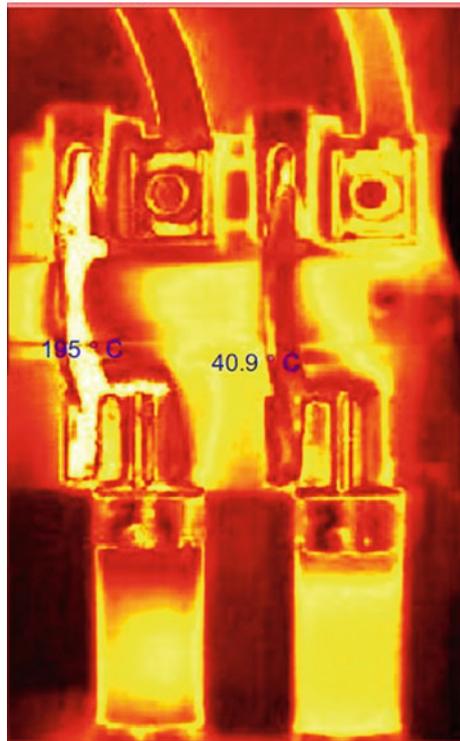
The table shows the area of image and the mean intensity range where the low intensity is 40.8661 and the high intensity is 195.2291.

In Fig. 11, it shows the dark red spot which have the maximum temperature 194 °C and reference temperature of fuse cabinet is 40.8 °C that is indicated in Fig. 11. So, it is clear seen that the temperature of fuse cabinet wire is very high, it causes fault can be occur. And the threshold values of these temperatures are 68 and 178.

Table 1 The image area and the mean intensity

Area	Mean intensity
23,600	40.8661
53,680	99.8926
37,995	195.2291

Fig. 11 Segmented region min and max temperature



The threshold value of area of redness for controlling and the alarming may be taken as following:

Upper Threshold Area = 80% of A.

Medium Threshold Area = 70% of A.

Lower Threshold Area = 60% of A.

These threshold values helped to find the hot area and the cool area and the medium area if the threshold value is high the intensity of image will be high and the temperature of that thresholding level is very high if low level of thresholding mean low intensity and the temperature will be low. The accuracy of this system is about 98.9%, operation speed is very fast, human factor 5%, running cost is very less, and reliability is 99%.

5 Conclusion

Infrared thermography is the best tool for fault diagnosis in power system. The noninvasive system: using thermogram image has been developed for monitoring and controlling the power equipment. The high impact region detected by the watershed segmentation. The threshold value of high impact region obtained and the also image intensity and temperature which can damage the equipment.

The equipment's health is to be monitoring by the threshold value which help to find the hot region and cool region of electrical equipment image and provide the alarming system which helps to indicate that the equipment is in good condition or is in bad condition according to temperature level. This proposed method is gives the satisfactory result and the accuracy of this system is about 98.9%.

References

1. Dengwen, Z., Wengang, C.: Image denoising with an optimal threshold and neighbouring window. *Pattern Recogn. Lett.* 1694–1697 (2008)
2. Vijayakumar, D., Malathi, V.: A flexible scheme for transmission line fault identification using image processing for a secured smart network. *Int. J. Electr. Electron. Eng.* **2**, 74–77 (2015)
3. Om, H., Biswas, M.: A generalized image denoising method using neighbouring wavelet coefficients. In: *Signal, Image and Video Processing*, Springer, Berlin (2013)
4. Kumar, R., Saini, B.S.: Improved image denoising technique using neighboring wavelet coefficients of optimal wavelet with adaptive thresholding. *Int. J. Comput. Theor. Eng.* **4**(3), (2012)
5. Panigrahi, S., Verma, T.: Texture image classification using neurofuzzy approach. *Int. J. Eng. Comput. Sci.* **2**, 2309–2313 (2013)
6. Ahmad, A., Rudrusamy, G., Budiarto, R., Samsudian, A., Ramadass, S.: A hybrid rule based fuzzy-neural expert system for passive network monitoring. *Int. Joint Conf. Artif. Intell.* **5**, 85–92 (2001)

7. Shitole, S.N., Zahran, O., Al-Nuaimy, W.: Advance neural–fuzzy and image processing technique in the automatic detection and interpretation of weld defects using ultrasonic time of diffraction. In: International Conference on NDT, vol. 4 (2007)
8. Varma, H., Namitha, R.M., Sunil, G.N., Kumaraswamy, H.V.: Thermo-vision scanning of vital nodes in a substation using image processing. *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.* **3**, 13881–13888 (2014)
9. Ibarahim, M., Amani, A.: Infrared thermography and distribution system maintenance in alexandria electricity distribution company. In: International Conference on Electricity Distribution, vol. 21, pp 14–18 (2011)
10. Novizon, N., Malek, Z.A., Bashir, N., Asilah, N.: Thermal image and leakage current diagnostic as a tool for testing and condition monitoring of arrester. *Jurnalteknologi* **4**, 27–31 (2013)
11. Laurentys Almeida, C.A., Caminhas, W.M., Braga, A.P., Paiva, V., Martins, H., Torres, R.: Intelligent thermo graphic diagnostic applied to surge arresters: a new approach. In: IEEE Transaction on the Power Delivery, vol. 24, pp 109–120 (2009)
12. Ha, H., Han, S., Lee, J.: Fault detection on transmission lines using a microphone array and an infrared thermal imaging camera. *IEEE Trans. Instrum. Measurement* **61**(1), 267–275 (2012)
13. Lifeng Pan.: Intelligent Image Recognition Research on Status of Power Transmission Lines. *Sensors and Transducers*, vol. 179, pp. 174–179. IFSA Publication (2014)
14. Jun, L., Liu, X.: Heating defect detection system scheme design based on infrared image processing for high voltage plant of substation. In: Advance in Control Engineering and Information Science, pp. 699–703. Elsevier, Amsterdam (2011)
15. Jadin, M.S., Taib, S., Kabir, S., Yusof, M.A.B.: Image processing methods for evaluating infrared thermographic image of electrical equipments. In: Progress in Electromagnetic Research Symposium Proceeding, pp. 20–23 (2011)

Design of Double Fork Shaped Patch Radiator for Ultra Wide Band Applications

Swarnaprava Sahoo, Mihir N. Mohanty and Laxmi P. Mishra

Abstract A double fork shaped patch radiator is suggested for ultra wide band (UWB) applications. The patch antenna is easy to design, low-priced, and small in size. The ultra wide band characteristic covers 3.1–9.8 GHz in UWB range of 3.1–10.6 GHz is obtained by employing dual fork-shaped radiating element and a rectangular strip with truncated partial ground plane. The intended radiator is designed on FR-4 epoxy substrate fed by 50Ω microstrip line having dimensions $(42 \times 24 \times 1.6) \text{ mm}^3$. The return loss, $S_{11} \leq -10 \text{ dB}$ covers 3.1–9.8 GHz. The radiator gives sustainable gain with closely omni directional radiation characteristics over the operating bandwidth in UWB range. The antenna characteristics such as VSWR, gain of the proposed radiator are also studied in details.

Keywords Patch antenna · Triple band · Ultra wide band (UWB) Partial ground plane

1 Introduction

The 10 dB bandwidth of 7500 MHz ranging from 3.1–10.6 GHz was emancipated by the Federal Communications Commission (FCC) with EIRP (Effective Isotropic Radiated Power) whose spectral density $-41.3 \text{ dB meter per Megahertz}$ for many communication applications [1]. So, in wireless world ever the ultra wide band (UWB) technology is a motivation for hypnotized academia and industrial surveillance. Now a day's many researchers are giving more interest for generation

S. Sahoo · M.N. Mohanty (✉) · L.P. Mishra
Department of Electronics and Communication Engineering, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India
e-mail: mihir.n.mohanty@gmail.com

S. Sahoo
e-mail: swarna.sahoo@gmail.com

L.P. Mishra
e-mail: lpmishra@soauniversity.ac.in

of ultra wide band spectrum of 3.1–10.6 GHz for commercialized purposes for its broad bandwidth, less power utilization, higher data transfer rate, and very low price. The patch antennas can cover ultra wide band which are printed and fabricated on a substrate. So many patch antennas are reported for ultra wide band applications like binomial curved shaped, elliptical, rectangular, circular disk [2–10], compact tapered shape slot [11], UWB patch radiator using differential fed [12], the magneto electric dipole radiator which is fed by differential fed [13, 14], UWB fractal patch antenna by using CPW fed [15], shape of U, shape of arc slot on the radiating element [16], employing parasitic components alongside the radiator radiating element to eliminate the identified band [17], placing a slit in the ground plane or feed line [18].

In this demonstration, the double fork shaped patch radiator for ultra wide band applications is proposed which is very easy to design. The triple band performances with desired bandwidth are achieved by inserting a rectangular strip into the bottom fork shaped patch with partial ground plane and the ultra wide band (UWB) characteristics is obtained by placing an extra fork shaped patch to the top of the bottom fork shaped patch with cutting two symmetrical rectangular slots at the opposite corners of the truncated ground plane. Here in both the cases the truncated ground plane is on the rear part of the substrate beneath to the radiating element. The double fork shape is obtained by inserting a rectangular strip in the dual shape of U patch radiator. The bottom fork shaped patch resonates at triple band and the top fork shaped patch resonates over in UWB range. The intended antenna configuration is compact and small in size. Section 2 describes the model conception of the proposed patch radiator; Sect. 3 explains the simulation results. At last Sect. 4 concludes the demonstration work.

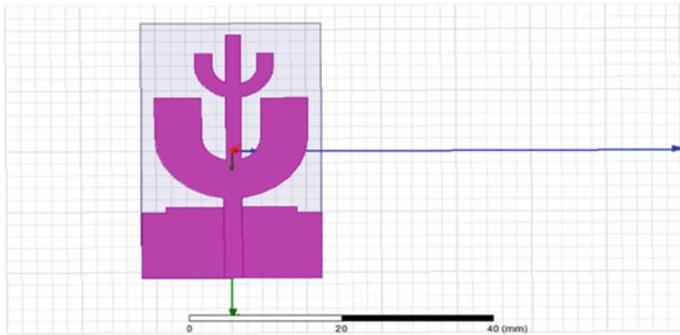
2 Antenna Design

The geometry and the specifications of the intended radiator are shown in Fig. 1. Which is fed by 50 Ω microstrip line feed printed on substrate epoxy FR4 with size 42 mm \times 24 mm of height 1.6 mm having dielectric constant (ϵ_r) as 4.4 as well as loss tangent (δ) as 0.02.

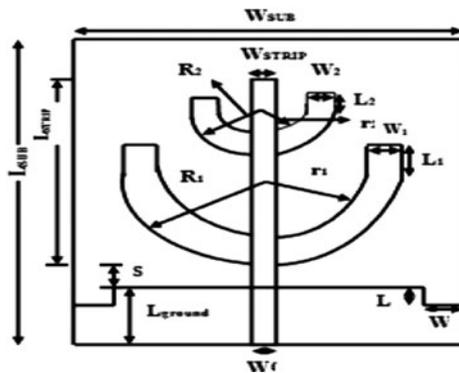
The intended structure of the radiator is a variant of circular patch radiator. From the circular patch [19] the radius R_1 is calculated from the given equation for the lower fork shaped patch.

$$f_{\text{lower}}, \text{GHz} = \frac{7.2}{2.25R_1 + S} \quad (1)$$

where, S is separation across the radiating element and partial ground and f_{lower} is the lower resonant frequency which corresponds to VSWR less than 2. It is decided to design the lower fork shaped patch due to the current concentration along the periphery of the circular patch. Therefore, an annular ring patch antenna is created



(a) Front view



(b) Specifications

Fig. 1 Geometry of the proposed shape of dual fork patch antenna **a** front side **b** specifications ($R_1 = 10.2$, $r_1 = 4$, $R_2 = 5.2$, $r_2 = 3$, L_{ground} , $L = 0.9$ mm, $W = 3.3$, $w_f = 2.4$, $S = 2$, $L_{STRIP} = 20.7$, $L_{SUB} = 24$, $W_{SUB} = 2$, $W_{STRIP} = 2$, $L_1 = 6.2$, $W_1 = 6.2$, $L_2 = 2.1$, $W_2 = 2.1$) (all dimensions are in mm)

with the removal of middle portion of the circular patch with insignificant effect on bandwidth impedance or radiation pattern characteristics. Due to this a semi annular ring structure is depicted and a rectangular strip is inserted vertically at the center of the top of semi annular ring patch that forms a shape of fork patch [20]. For better impedance bandwidth and radiation characteristics another fork shaped patch is designed which is placed on the upper to the lower fork shaped patch whose radius and specifications are half of the lower fork shaped patch. The whole antenna-designed structure is optimized such that its operating bandwidth with return loss ≤ -10 dB covers the range of ultra wide band (UWB) frequency. The rectangular strip is inserted in the middle section of the fork shaped patch to resonate at triple band characteristics. The intended antenna provides UWB operation due to the smaller fork shape patch. The bottom fork shape patch resonates at triple band while the top fork shape (intended antenna) patch resonates over ultra

wide band (UWB). The length (L_{strip}) of the rectangular strip is of $\lambda/4$ long at the middle resonant band frequency.

$$L_{\text{strip}} = \frac{c}{4f_{\text{lower}}} \quad (2)$$

By optimization of dimensions of rectangular strip, dimension of bottom fork shape patch and top fork shape patch the performance of UWB is achieved to resonate over 3.1–9.8 GHz which covers the UWB. The separation among the partial ground plane and radiating element is optimized because it behaves like matching element and enhances the bandwidth of impedance to obtain the ultra wide band characteristics. It influences the resonant frequencies and bandwidths in three bands. Uniform rectangular slots on the opposite corners of the upper edge of the partial ground plane. Two rectangular slots enhance bandwidth impedance for UWB characteristics. The commercial Ansoft High Frequency System Simulator (HFSS) [21] is used to carry out the design process, simulation along with optimization.

3 Simulation Results

The characteristics of the double fork shaped patch radiator depends on the variables like separation (S) across the radiating element and the partial ground, the dimensions ($W \times L$) of the symmetrical rectangular slots on the two corners of the partial ground plane, the dimensions ($W_1 \times L_1$) and ($W_2 \times L_2$) of the rectangular stubs over the bottom and top semi circular annular ring, outer radius (R_1, R_2), inner radius (r_1, r_2) of the semi annular ring and size of partial ground plane. The consequential effect on ultra wide band characteristics is explained with parametric studies.

The separation (S) among the radiating element as well as the partial ground plane influences the bandwidth impedance. It behaves as a matching network. The return loss versus frequency plot of the intended radiator at various values of separation (S) is given in Fig. 2.

The optimal impedance bandwidth is observed when $S = 2$ mm. The capacitance occurs from the separation between the boundary of the partial ground plane and radiating element that reasonably stabilizes the antenna inductance. The initial resonant frequency of the circular patch radiator depends on outer radius (R_1) and separation (S). The impedance bandwidth depends on inner radius (r_1). The UWB operating bandwidth is observed when ($r_1 = 4$ mm and $r_2 = 3$ mm). The return loss (S_{11}) degrades when inner radii r_1 and r_2 are increased beyond 4 and 3 mm. The length of rectangular strip (L_{STRIP}) on the upper side of the two semicircular semi annular rings is responsible for the first resonant frequency. With increasing in L_1 the lower frequency of the UWB slightly decreases. The three different modes correspond to these resonant frequencies. Similarly, with increasing in L_2 the lower

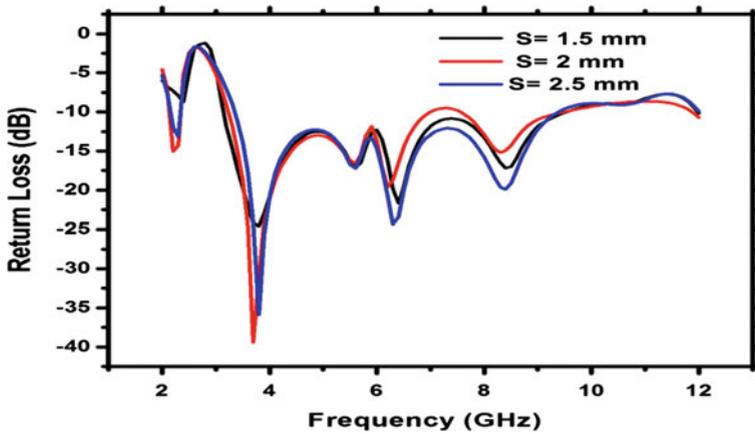


Fig. 2 Return loss versus frequency at various values of separation (S)

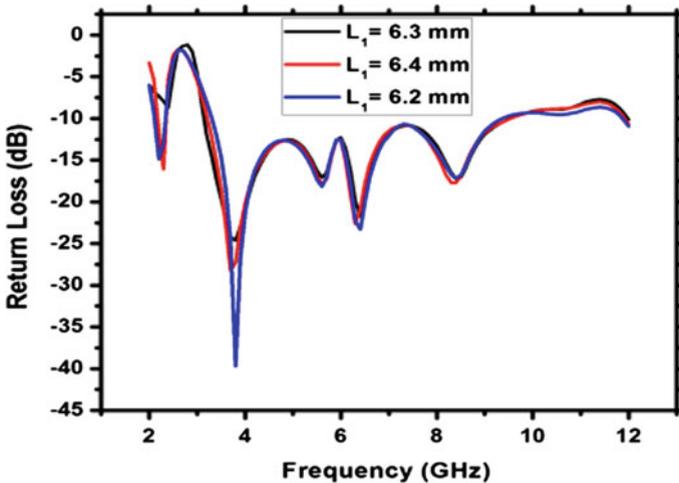


Fig. 3 Return loss against frequency at various L_1

frequency of the first resonant frequency decreases and upper frequency increases. The return loss plot is shown for various values of L_1 and L_2 in Figs. 3 and 4, respectively.

The rectangular strip at the middle of the double fork shaped patch radiator resonates over the ultra wide band frequency range. The middle resonated frequency is determined by the length (L_{STRIP}) of the rectangular strip while the width of rectangular strip (W_{STRIP}) influences impedance bandwidth of the ultra wide band. When L_{STRIP} decreases the lower limit of the first resonant frequency increases, when there is increase in L_{strip} the lower edge limit of the second resonant

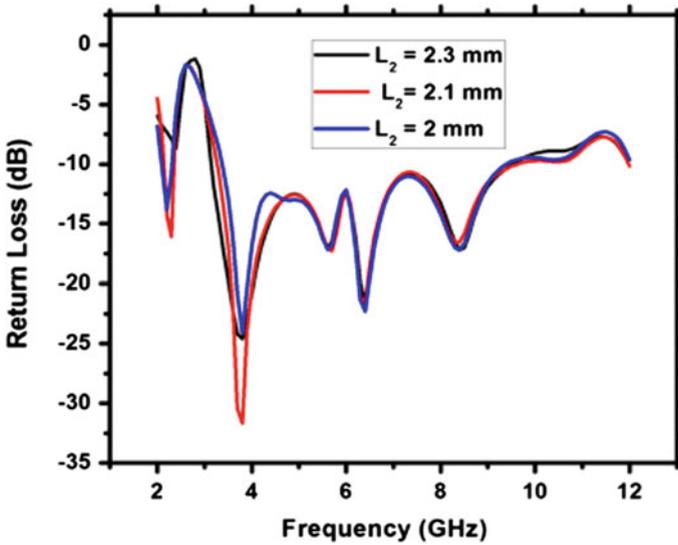


Fig. 4 Return loss against frequency at various L_2

frequency increases and with increasing in W_{strip} , the lower edge limit of the band increases and an extra resonant frequency appears at 2.2 GHz resulting in dual band performance. The length of the rectangular strip is evaluated using formula (2). The length of the rectangular strip is less than the calculated length because of properties of substrate, fringing effect and mutual coupling among the radiating elements that resonate over ultra wide band frequency range. Figures 5 and 6 represent the return loss for different values of length (L_{STRIP}) and width (W_{STRIP}) of the rectangular strip, respectively.

When the length of the two symmetrical rectangular slots in the partial ground plane decreases, the lower limit of the operating band sensitively changes as reported in Fig. 7. Figure 8 explains the simulated return loss plot of the intended radiator for UWB application.

The simulated proposed antenna having bandwidth 6500 MHz with $\text{VSWR} < 2$ is presented in Fig. 9. Figure 10 represents the gain versus frequency graph of the radiator having maximum gain 8.6 dBi and average gain 5.06 dBi.

Figure 11 presents the simulated radiation patterns characteristics at 3.8, and 5.6 GHz in E-(XZ) and H-(XY) planes, respectively.

The radiation characteristics in E-plane and H-plane are normalized with respect to the maximum in their crossing points at each frequency. The intended radiator shows a stable radiation pattern over the operating bandwidth. The radiation characteristics deteriorate at the higher frequency due to the change in radiating area with frequency of the operating bandwidth. Other important parts in the retrogression of radiation characteristics at higher frequency are different phase distribution and consequential magnitude of higher order mode at upper frequencies. The

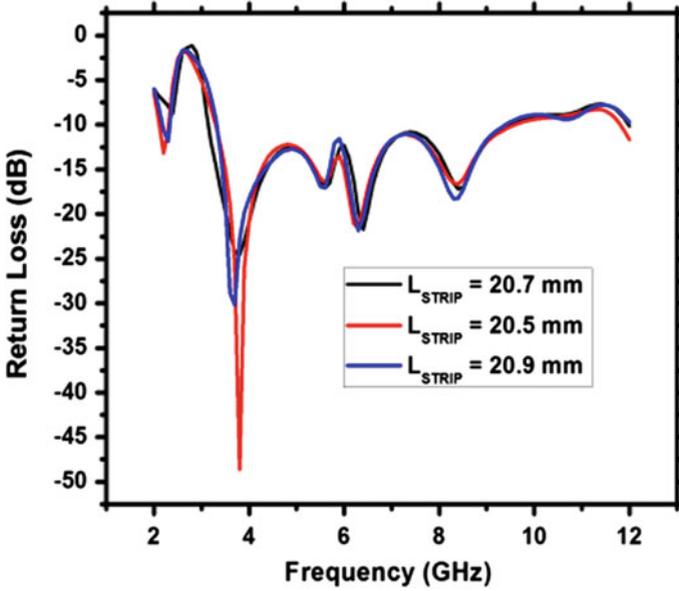


Fig. 5 Return loss versus frequency at various L_{STRIP}

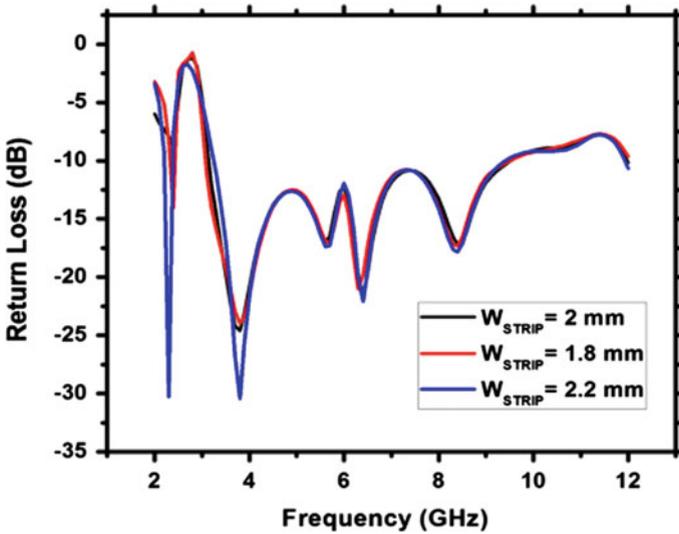


Fig. 6 Return loss versus frequency at various W_{STRIP}

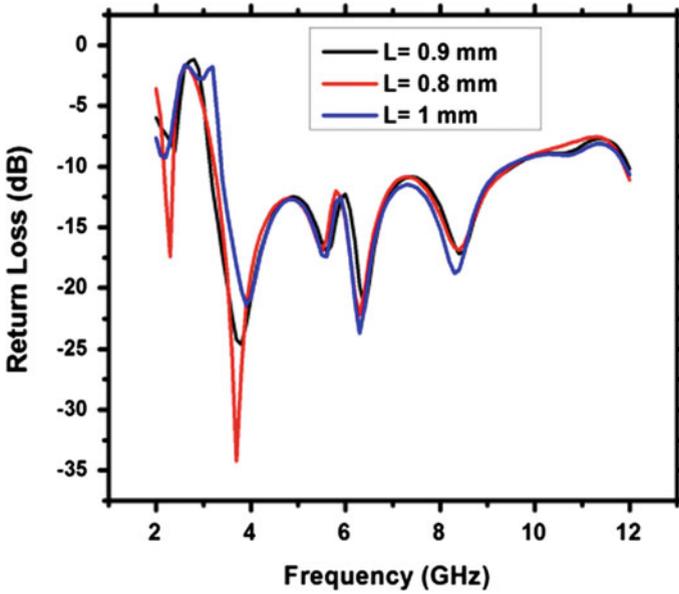


Fig. 7 Return loss versus frequency at various L in partial ground plane

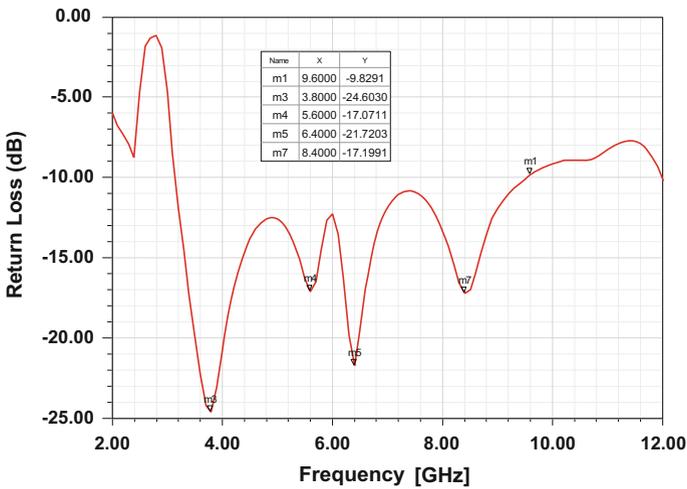


Fig. 8 Return loss plot of the intended radiator

intended radiator has closely omnidirectional radiation pattern in XY-plane and radiation pattern like figure eight shape in the XZ-plane over the operating bandwidth. At frequencies in XZ-plane the cross polarization level is about -20 dB

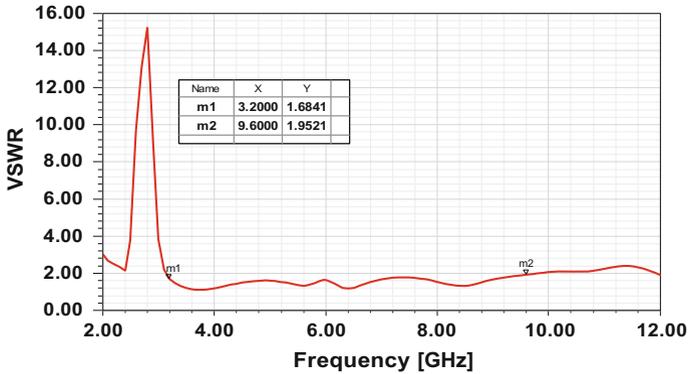


Fig. 9 VSWR graph of the intended radiator

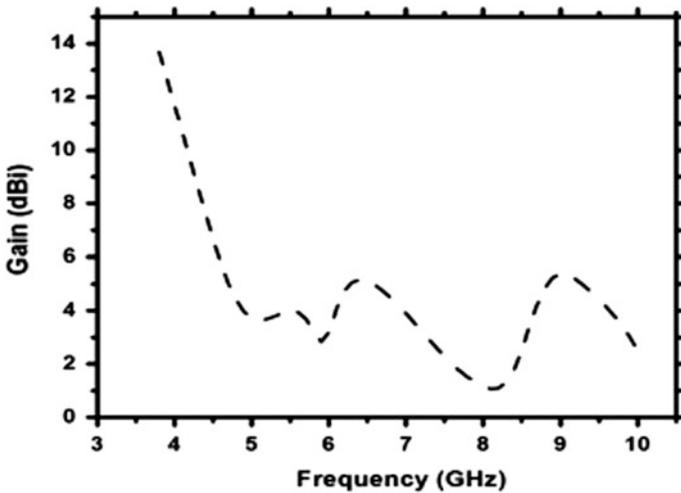
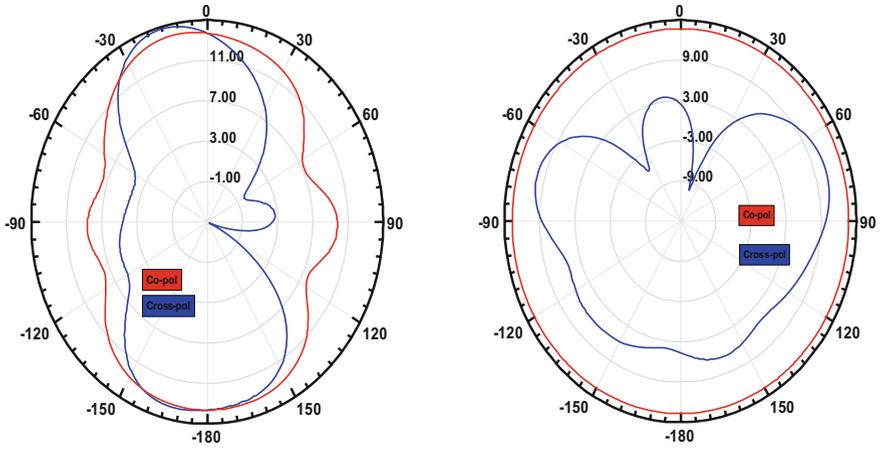


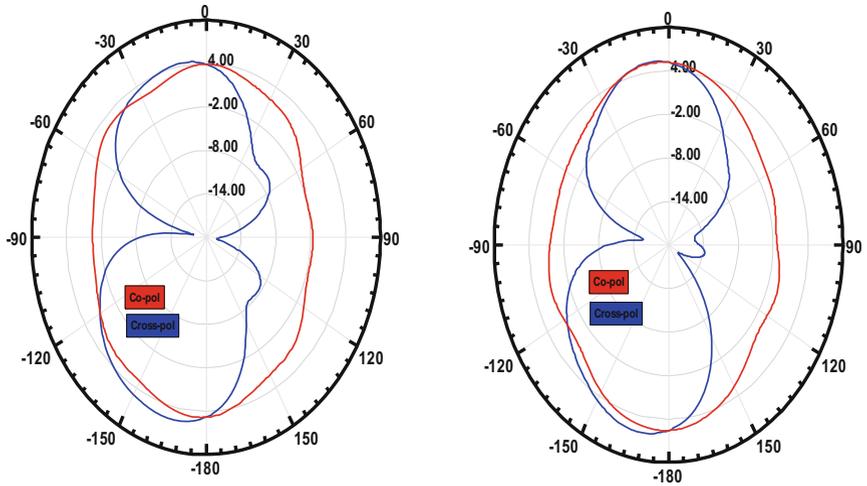
Fig. 10 Gain versus frequency graph of the proposed radiator

below the co-polarization level. The surface current distributions for 3.8, 5.6 GHz are presented in Fig. 12.

From Fig. 12a it is shown that the middle rectangular strip resonates at 3.8 GHz. Again in dual fork shaped patch the current is induced, but around the periphery of the dual fork shaped patch the surface current creates a loop and thereafter the radiations are primarily from the middle rectangular strip. The surface current density component (J_x) is very low which is produced at the periphery of the partial ground plane alongside the radiating double fork shaped patch radiator at 3.8 GHz. From Fig. 12b it is shown that the motion of surface current is in vertical direction in the double fork shaped patch very little current in the rectangular strip. The small



(a) 3.8 GHz



(b) 5.6 GHz

Fig. 11 Simulated radiation patterns of the intended double fork shaped patch radiator at 3.8, and 5.6 GHz frequencies **a** E-(XZ). **b** H-(XY) plane

surface current density (J_x) component is produced at the boundary of the partial ground plane alongside the radiating element. So, the radiation pattern characteristics are closely omnidirectional with relatively less cross polarization constituent.

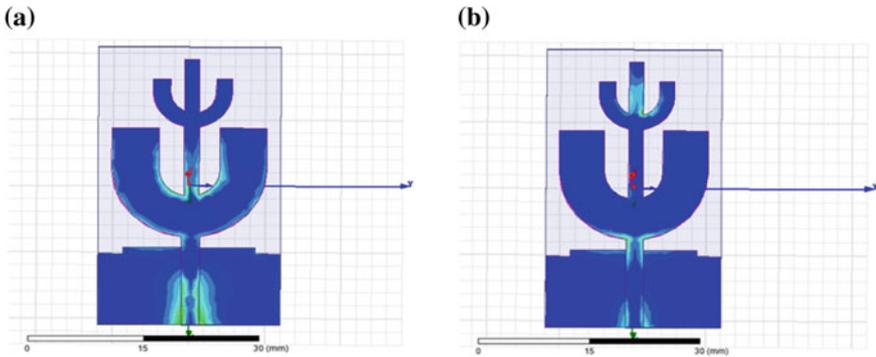


Fig. 12 Surface current distributions of the proposed antenna **a** 3.8 GHz. **b** 5.6 GHz

4 Conclusion

The double fork shaped patch radiator is suggested for ultra wide band (UWB) applications which are very easy in design, low-priced and small in size. The rectangular strip with bottom shape patch governs the triple band and the top fork shaped patch governs the UWB applications. The intended antenna gives 90% antenna radiation efficiency and its maximum gain and average gain are 8.6 and 5.06 dBi, respectively, over UWB. The intended radiator has closely omnidirectional radiation patterns which indicate the good candidate for UWB applications.

References

1. Federal Communications Commission revision of Part 15 of the commission's rules regarding UWB transmission systems FCC, Washington, First Report and Order FCC, 02.V48, (2002)
2. Agrawal, N., Kumar, G., Ray, K.: Wide-band planar monopole antennas. *IEEE Trans. Antennas Propag.* **46**(2), 294–295 (1998)
3. Ammann, M.J., Chen, Z.N.: Wideband monopole antennas for multi-band wireless systems. *IEEE Antennas Propag. Mag.* **45**(2), 146–150 (2003)
4. Clerk, J., Liang, J., Chiau, C.C., Chen, X., Parini, C.G.: Study of a printed circular disc monopole antenna for UWB systems. *IEEE Trans. Antennas Propag.* **53**(11), 3500–3504 (2005)
5. Low, Z.N., Cheong, J.H., Law, C.L.: Low-cost PCB antenna for UWB applications. *IEEE Antennas Wireless Propag. Lett.* **4**, 237–239 (2005)
6. Wong, K.L., Su, S.W., Tang, C.L.: Broadband omnidirectional metal-plate monopole antenna. *IEEE Trans. Antennas Propag.* **53**(1), 581–583 (2005)
7. Ling, C.W., Lo, W.H., Yan, R.H., Chung, S.J.: Planar binomial curved monopole antennas for UWB communication. *IEEE Trans. Antennas Propag.* **55**(9), 2622–2624 (2007)
8. Ammann, M.J., Chen, Z.N.: A wide-band shorted planar monopole with bevel. *IEEE Trans. Antennas Propag.* **51**(4), 901–903 (2003)
9. Li, P.C., Liang, J.X., Chen, X.D.: Study of printed elliptical/circular slot antennas for UWB applications. *IEEE Trans. Antennas Propag.* **54**(6), 1670–1675 (2006)

10. Chahat, N., Zhadobov, M., Sauleau, R.: A compact UWB antenna for on-body applications. *IEEE Trans. Antennas Propag.* **59**(4), 1123–1131 (2011)
11. Azim, R., Islam, M.T., Misran, N.: Compact tapered-shape slot antenna for UWB applications. *IEEE Antennas Wireless Propag. Lett.* **10**, 1190–1193 (2011)
12. Zhaoyang, T., Ruina, L., Yingzeng, Y.: Differential-fed ultra wide band patch antenna with triple band-notched characteristics. *Electron. Lett.* **51**(22), 1728–1730 (2015)
13. Li, M.J., Luk, K.M.: A differential-fed magneto-electric dipole antenna for ultra wide band applications. *IEEE Trans. Antennas Propag.* **61**(1), 92–99 (2013)
14. Wang, J.H., Yin, Y.Z.: Differential-fed ultra wide band microstrip antenna with improved radiation patterns. *Electron. Lett.* **50**(20), 1412–1414 (2014)
15. Moghadasi, M.N., Sadeghzadeh, R.A., Sedghi, T., Aribi, T., Virdee, B.S.: Ultra wide band CPW-fed fractal patch antenna with band notched function employing folded T-shaped element. *IEEE Antennas Wireless Propag. Lett.* **12** (2013)
16. Mehranpour, M.J., Ghobadi, C., Ojaroudi, M.: Dual band-notched square monopole antenna for UWB applications. *IEEE Antennas Wireless Propag. Lett.* **11**, 172–175 (2012)
17. Mardani, H., Ghobadi, C., Nourinia, J.: A simple compact monopole antenna with variable single and double filtering function for ultra wide band applications. *IEEE Antennas Wireless Propag. Lett.* **9**, 1076–1079 (2010)
18. Kim, D.O., Kim, C.Y.: CPW-fed ultrawide band antenna with triple-band notch function. *Electron. Lett.* **46**(18), 1246–1248 (2010)
19. Kumar, G., Ray, K.P.: *Broadband Microstrip Antennas*. Artech House, Norwood (2003)
20. Mishra, S.K., Gupta, R.K., Mukherjee, J.: Parallel metal plated tuning fork shaped omnidirectional monopole antenna for ultra wide band applications. *Microwave Opt. Technol. Lett.* **601–604** (2011)
21. Ansoft High Frequency System Simulator Software vir. 13

Subjective Analysis of Multimedia Traffic in Wireless Multimedia Sensor Networks

A. Sivagami and S. Malarkkan

Abstract Topical advances in Wireless Sensor Networks (WSN) have directed to the consequent generation of networks that let repossessing video, audio and data from the corporal surroundings called Wireless Multimedia Sensor Networks (WMSN) will facilitate novel claims such as multimedia observation sensor networks, environmental and structural screening, etc. Most of these applications entail means to bring multimedia content with not only an echelon of Quality of Service (QoS) but also Quality of Experience (QoE) metrics like, PSNR, SSIM, MOS, etc. Provisioning ample throughput, crouch delay and stumpy packet trouncing rate, while deeming network lifetime are the main QoS prerequisites of multimedia data transmission over WMSN. These setbacks impose us to exploit multipath routing that can afford soaring end-to-end bandwidth while providing high resiliency against the channel and clogging error for video transmission in WMSNs. To sustain a reliable QoE, Hexanode Realtime Power aware Routing protocol (RPAR) with cross layering has been proposed and hole bypassing also been introduced and tested by means of sending multimedia data like, audio, video and data signals and proved to be better while compared with the performance over Pre emptive AODV (PAODV) and Pre emptive DSR (PDSR) protocols.

Keywords QoE · Hexanode RPAR · PSNR · SSIM · MOS

A. Sivagami (✉)

Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Kancheepuram
631561, India
e-mail: arasusivagami19@gmail.com

A. Sivagami

Department of Electronics & Communication Engineering, Sri Manakula Vinayagar
Engineering College, Puducherry 605107, India

S. Malarkkan

Manakula Vinayagar Institute of Technology, Puducherry 605107, India

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_59

1 Introduction

Wireless multimedia sensor networks, encompassed of sensor devices endowed with audio and visual information compilation units [1]. It has the capability to recuperate multimedia data, accumulate or route data in real time. It correlates and blend multimedia data instigated from assorted resources and wirelessly transmit gathered data to preferred destination. To minimize cost factor, energy constraints and bandwidth are considered to be the most important resources in WMSN [2, 3]. The WMSN gives various assents over benchmark observation and control work plots. Routing methods play a prime role in conservation of bandwidth and end-to-end delay.

2 Existing Work

The consolidated selection of protocol work has been shown in Fig. 1.

(i) Selection of Hexanode and Cluster Arrangement

The foremost fondness of hexagonal especially for the coverage performance [2] and it has “*M*” no. of cells with each cell radius, “*a*”.

The total area of hexagon =

$$(3\sqrt{3}a^2)/2$$

In this the base station partitions the intact network to internal cells (IC) and outermost cells, and cluster head (CH) is chosen based on residual energy [15]. After network is established, IC forwards their data to base station (BS). Node which has superior energy is chosen as CH and node once chosen as CH cannot turn out to be again until all the nodes befall to be CH (Fig. 2).

$$\text{Cluster Size} = m * n$$

$$\text{Cluster Area} = N * (3\sqrt{3}a^2)/2$$

$$\text{Total no. of clusters} = M * N$$

where,

m No. of nodes

n Cluster size

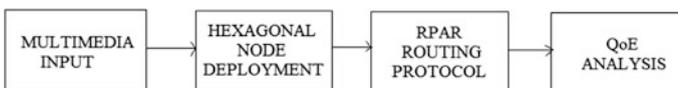
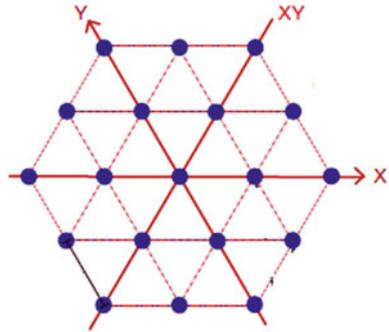


Fig. 1 Hexanode RPAR selection flow

Fig. 2 General hexagonal structure



N No. of hexagon within one cluster
M No. of clusters

(ii) QoE Metrics

A metric is portrayed as “an arrangement of associated measures that persuades the appraisal of a number of specific trademark” [4, 5].

(a) Structural Similarity (SSIM):

The structural similarity (SSIM) index is a one of the subjective measures of foretelling the observed quality of pictures, images and videos [6]. SSIM is employed for quantifying the correlation between two images and is scaled between 0 and 1.

SSIM for the input video signal can be estimated using the formula given below:

$$SSIM(M, N) = \frac{(2\mu_m\mu_n + C_1)(2\sigma_{mn} + C_2)}{(\mu_m^2 + \mu_n^2 + C_1)(\sigma_m^2 + \sigma_n^2 + C_2)}$$

where, *X* and *Y* are original frame and the reference frame, etc. μ_m and μ_n are mean values of *M* and *N* frames. σ_m and σ_n are variance of *M* and *N* · σ_{mn} is covariance value.

Thus the graph illustrates the SSIM value for each set of frames transmitted.

(b) Peak Signal to Noise Ratio (PSNR): Peak signal-to-noise ratio, habitually abbreviated PSNR, is another prime metric which defines the ratio flanked by the greatest achievable power of a signal and the power of humiliating noise that manipulates the steadfastness of its depiction [7]. PSNR works out mean square error (MSE) of each pixel stuck among original and received images. Images with greater similarity outcomes maximum PSNR [8, 9, 10].

(c) Mean Opinion Score (MOS): The mean opinion score (MOS) offers a statistical appraisal of the eminence of human speech at the targeted end of the circuit. The scheme formulates the use of subjective analysis (estimated scores) that are precisely averaged to acquire a quantitative sign of the system concert.

3 Proposed Work

(a) Selection of RPAR

To provide real-time concert and preserve energy consumption and exploits multihop communication [11, 6]. Many wireless sensor network applications must steadfastness the innate variance flanked by energy efficient communication and necessitate to realize required Quality of Experience service such as SSIM and MOS, etc. To tackle this dispute, Real-time Power-Aware Routing (RPAR) protocol is put forward because it attains the communication delays without exchanging periodic control message. RPAR attributes a power-aware forwarding strategy and builds up the neighbour information with a constrained flooding-based mechanism [5]. Moreover, computing the waiting time before flooding the waiting time of each node is different according to residual energy and distance to sink and estimating delay to sink based on one hop neighbour information in the transmitted packet and then forwarding the packet provides real-time performance.

The protocol architecture comprises of four sections as follows:

- (i) Dynamic velocity: This scheme utilizes velocity assignment policy and velocity will be calculated based on slack and velocity towards destination.

$$\text{Slack}(A) = s_y(A) - ((t_h(A) - t_y(A))) \quad (1)$$

$$v_{\text{req}}(A, B) = d(A, B)s_y - (t_h(A) - t_r(A)) \quad (2)$$

where, B is the destination of the packet and $d(A, B)$ is the Euclidean distance stuck between A and B . It is momentous to note that the target is accomplished if the entailed velocity is reached at each hop.

- (ii) Delay Estimator: This module designed based upon waiting time computation, delay estimation and forwarding decision.

$$\text{Delay}(A, (K, q)) = (d_c(A) + d_t) \cdot H(A, (K, q)) \quad (3)$$

- (iii) Forwarding: RPAR makes forwarding decision without periodic control message exchange such as neighbour information. As shown in Eq. (2), the velocity provided by (K, q) is:

$$v_p(A, B, (K, q)) = (d(A, B) - d(K, B)) d(A, (K, q)) \quad (4)$$

- (iv) Neighbourhood manager: The proposed scheme builds up the neighbour information with a constrained flooding-based mechanism which reduces the redundancy in flooding-based operation by means of reducing the no. of forwarding nodes that transmits packet to the node closest to sink.

(b) Cross-layer Design

Unmistakably, cross-layer arrangements include cross-layer flagging which is not particular in the convention style. These flagging systems ought to use as rare property as likely, dropping the overhead (Fig. 3).

Cross-layer (XL) designs have been particularly put forward in wireless communications as an imperative hypothesis to optimize the inadequate wireless bandwidth exploitation. Cross layer is defined as the overall system concert is not independently ascertained by all parameters at each individual layer but is resolved by momentous parameters and their interactions of equivalent layers. They address state of affairs where diverse OSI layers may aid to progress the capability of applications to accomplish assured goals such as QoE guarantees, power cutback, customization according to user preferences, etc. Cross-layer design is optimized by incorporating datalink, network and transport layers.

Bundle headers: In place can be settled in layer headers which can later be second-hand by some other covering to gather the sought data. This can be differentiation to have channel like stream of sign among the layers.

ICMP messages: in IP-based framework, Internet Control Message Protocol (ICMP) messages can be utilized for flagging. Then again, as ICMP mail are constantly exemplified by IP parcels, the messages need to navigate through the net layer, regardless of the possibility that the cooperating layer sets are information connect and physical or convey and utilization.

System benefit: An additional stockpiling in the system can be utilized to store the layer stricture. This stockpiling can be control by some disseminated servers which assemble strait and connection states and serve the layers as required.

Neighbourhood profiles: This weaving machine is like the former one, yet as opposed to utilizing some different areas, layer point of confinement are put away and imparted generally, inside the host.

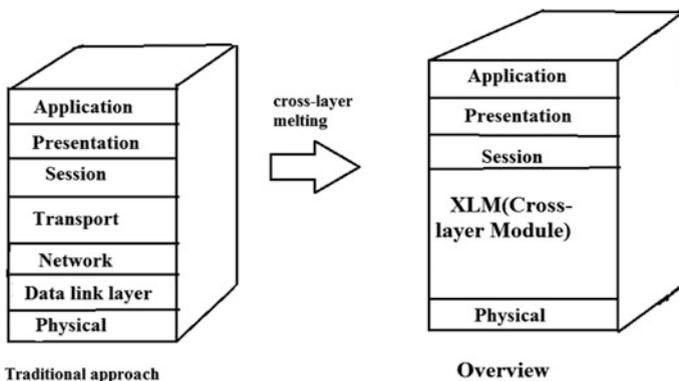


Fig. 3 Cross-layer design

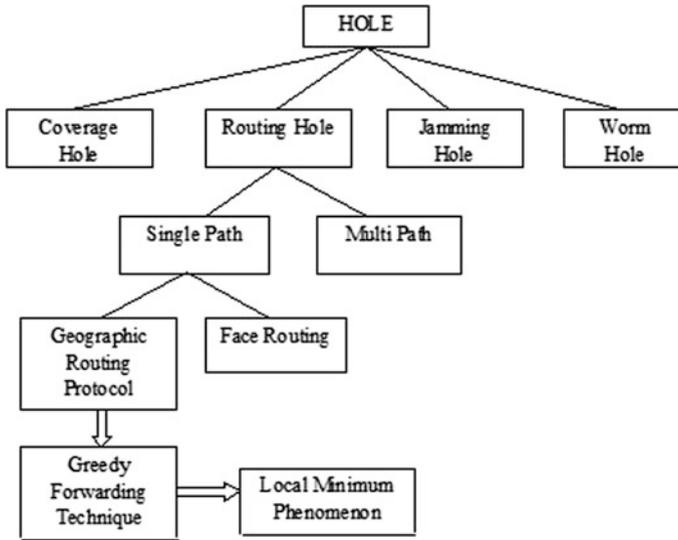


Fig. 4 Types of holes

(c) Hole Management

If the sender has not received any forwarding packet, there would not be any node in the transmission range. This is called hole. Two approaches can be suggested (Fig. 4).

- (i) The transmission range can be increased twice and packet can be retransmitted. This approach is named as latency sensitive approach.
- (ii) The previous sender knows there exists hole around the sender and packet can be retransmitted, i.e. Packet forwarded by other nodes. Among the two approaches, latency sensitive is preferred.

4 Simulation Parameters

NS-2 takes apart control path realizations from the data path realization. The module plan and the vital network module entities in the data path are engraved and piled up with C++ to diminish packet and event processing time. OTcl turns out to have the attribute that C++ lacks. So the amalgamation of these substantiates to be very efficient. C++ is used to realize the meticulous protocol and OTcl is utilized for abusers to organize the simulation state of affairs and program the events (Table 1).

Table 1 Simulation parameters

Property	Values
Set val (chan)	Channel/wireless channel
Set val (prop)	Propagation/two ray ground
Set val (netif)	Phy/WirelessPhy
Set val (MAC)	Mac/802_11
Set val (ifq)	Queue/Drop Tail/Pri Queue
Set val (ll)	LL
Set val (ant)	Antenna/Omni antenna
Set val (ifqlen)	100
Set val (nn)	200
Set val (rp)	RPAR, PADOV, PDSR
Set val (x)	1000
Set val (y)	1000
Set val (stop)	200s
Energy model	Energy model
Initial energy	100 J
Txpower	0.8 mW
Rxpower	0.6 mW
Video sending agent	My UDP
Video receiving agent	My Evalvid Sink

5 Results and Discussion

The simulation is carried out using Network Simulator (NS-2) and scrutiny is presented below. We appraise the concert and validate the efficacy of proposed RPAR through this simulation with two protocols namely PAODV, PDSR are also presented in the graphs below. The simulation is carried out for the network dimension varying commencing (50–250) nodes.

(A) The improvement in transmission of multimedia data is possible by cross layering datalink layer, network layer and transport layers to achieve coordination among layers. Performance improvement arises with RPAR while comparing with other protocols by testing with the transmission of multimedia data.

(i) No. of Nodes versus PSNR:

Inference: Figure 5 illustrates that RPAR furnishes enhanced video quality of 29 dB than erstwhile systems as it depends on communed packet kind that consent to superior precedence frames such as I-Frame. While other systems PAODV shows PSNR of 25 dB, PDSR shows PSNR of 24 dB (Fig. 6).

(ii) No. of Nodes versus SSIM

Inference: SSIM acquires luminance, discrepancy and the structure distinction into account to quantify the likeness of two images. The assessment of SSIM descends

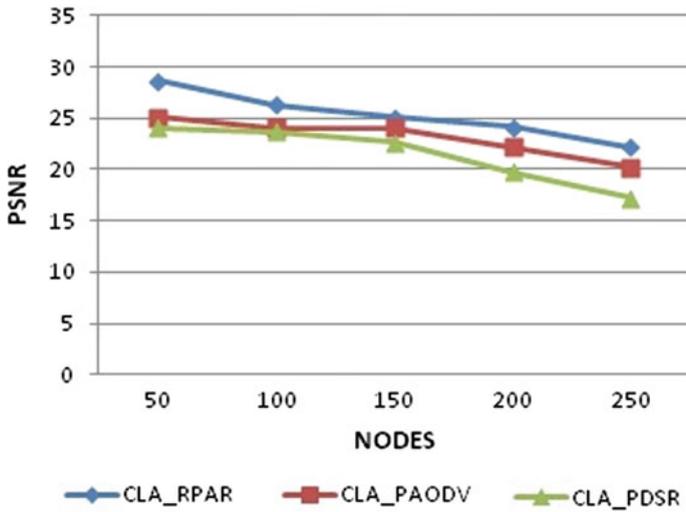


Fig. 5 Nodes versus PSNR

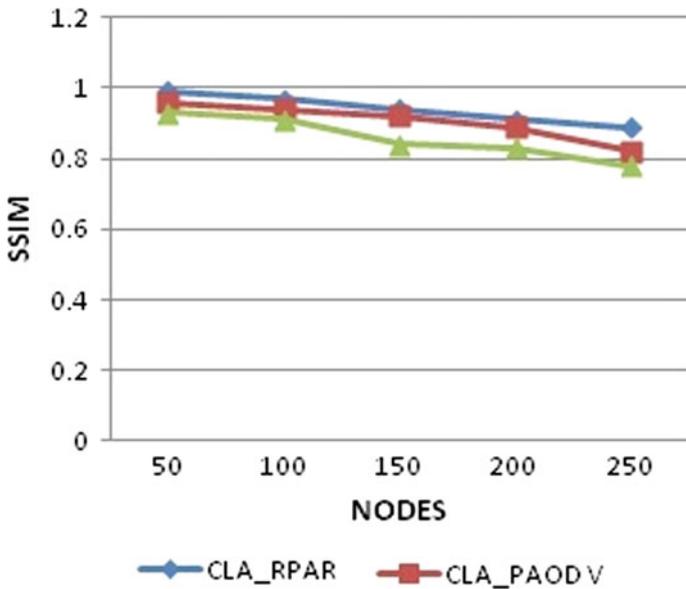


Fig. 6 Nodes versus SSIM

amid 0 and 1. The SSIM of the video transmitted by our proposal was 0.98 without applying cross-layering concept and after applying cross layering, it becomes 0.99.

(iii) No. of Nodes versus MOS

Inference: Figure 7 shows that the mean MOS of all users for the different schemes by varying number of nodes. The RPAR scheme always selects the highest remaining energy path with obstacle overcome path with higher throughput. The RPAR provides a smoother MOS compared to the two other schemes.

(B) Further the improvement has been attained with the hole bypassing concept and the consolidated report has been tabulated.

(i) No. of nodes versus PSNR:

Inference: Figure 8 explains that RPAR gives enhanced video quality of 31.05 dB adaptively makes use of apposite encoding parameters according to existing wireless channel whereas other systems PAODV proves PSNR of 28 dB and PDSR proves PSNR of 26 dB, where their queue scheduling does not maintain lofty priority packets in case of congestion, in accumulation RPAR condenses obstacle problem and energy handling of the network so that additional packets can be passed on (Fig. 9).

(ii) No.of Nodes versus SSIM

Inference: SSIM was utilized to fundamentally mock-up human visual discernment. SSIM takes luminance, disparity and the structure discrepancy into explanation to appraise the similitude of two images. The SSIM of the video transmitted by our scheme was at all times better than 0.9. Regardless of the energy efficient path and obstacle free paths with limited bandwidth in the wireless sensor network, our scheme toils better in forwarding superior quality, visually distinguishable video towards the cluster head.

(iii) No. of Nodes versus MOS

Inference: Figure 10 shows that the MOS of all users for the different schemes by varying number of node. The RPAR scheme always selects the utmost remaining

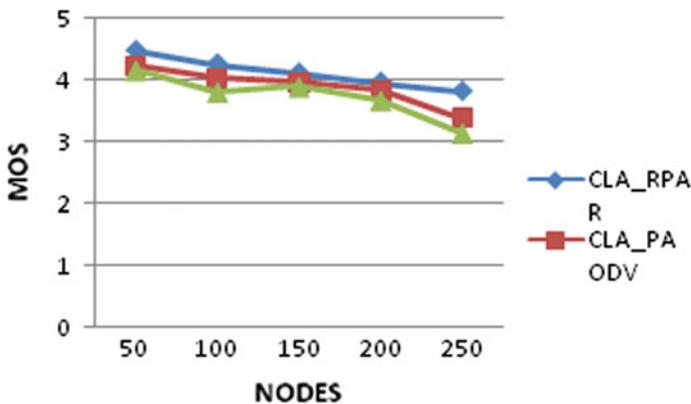


Fig. 7 Nodes versus MOS

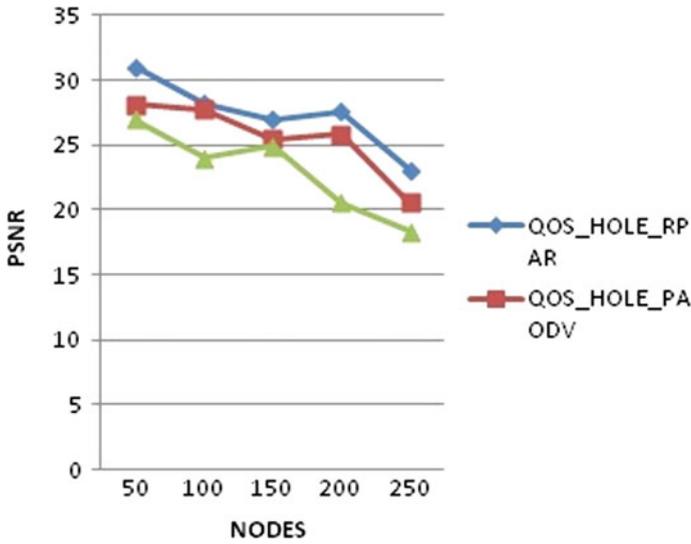


Fig. 8 Nodes versus PSNR

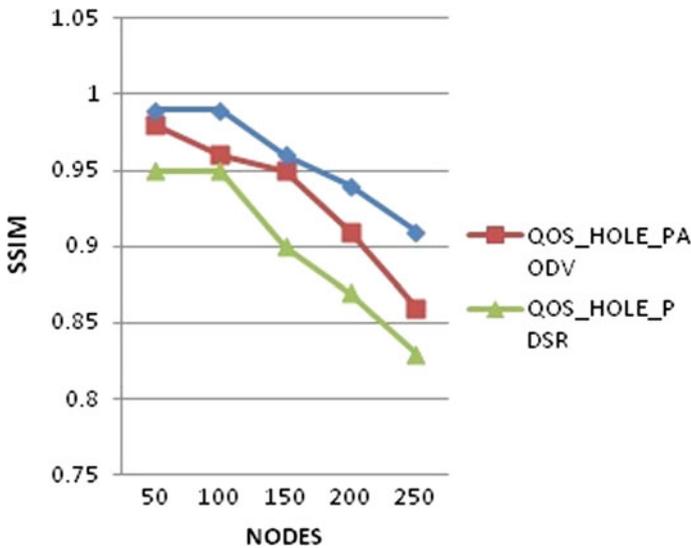


Fig. 9 Nodes versus SSIM

energy path with obstacle overcome path with higher throughput. The RPAR affords a smoother mean MOS in contrast to the two other schemes. It also directs to a plausibly privileged minimum mean MOS in the network. The consolidated performance can be tabulated as shown in Table 2.

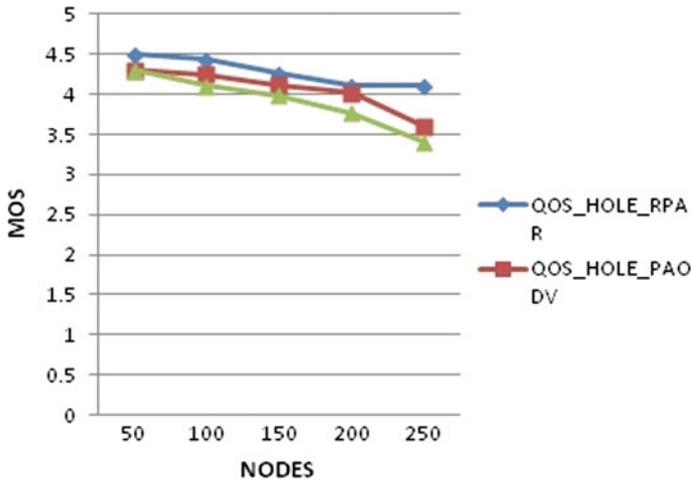


Fig. 10 Nodes versus MOS

Table 2 Consolidated performance

Nodes	Protocol	PSNR	SSIM	MOS
50	RPAR	31.05	0.99	4.5
50	PAODV	28.11	0.98	4.3
50	PDSR	26.95	0.95	4.3
100	RPAR	28.18	0.99	4.43
100	PAODV	27.74	0.96	4.25
100	PDSR	24	0.95	4.1
150	RPAR	27.01	0.96	4.26
150	PAODV	25.48	0.95	4.12
150	PDSR	24.91	0.9	3.98
200	RPAR	27.6	0.94	4.1
200	PAODV	25.79	0.91	4.02
200	PDSR	20.51	0.87	3.77
250	RPAR	23.11	0.91	4.1
250	PAODV	20.56	0.86	3.6
250	PDSR	18.32	0.83	3.4

6 Conclusion

This proposal initiated the real-time power aware routing protocol to afford QoE-aware multimedia transmission of voice, data and video streaming, while deploying nodes in hexagonal type. By examining the simulation results, it was initiated that real-time power aware routing protocol facilitates voice and video delivery with a nominal quality level from a customer’s point of view. This was

endorsed by assessing the intended protocol through renowned objective metrics (SSIM, PSNR and MOS) as well as illustrating video frames. Thus, we can wrap up that the projected real-time power aware routing protocol gives out voice, video and data with QoE assurance.

References

1. Zaidi, S.M.A., Jung, J., Song, B., Lee, H., Youn, H.Y.: Multi-channel multi-path video transmission over wireless sensor networks. In: Proceedings of the IEEE Consumer Communications and Networking Conference, Las Vegas, Nov 2013
2. Serral-Gracia, R., Masip-Bruin, X., et al.: An overview of quality of experience measurement challenges for video applications in IP networks, pp. 252–263 (2011)
3. Gomez, C., Boix, A., Paradells, J.: Impact of LQI-based routing metrics on the performance of a one-to-one routing protocol for IEEE 802.15. 4 multihop networks. EURASIP J. Wireless Commun. Netw. (2010)
4. Ehsan, S., Hamdaoui, B.: A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks. IEEE Commun. Surv. Tutorials **99**, 1–14 (2011)
5. Ben-Othman, J., Yahya, B.: Energy efficient and QoS based routing protocol for wireless sensor networks. J. Parallel Distrib. Comput. **70**(8), 849–857 (2010)
6. Li, S., Neelisetti, R.K., Liu, C., Lim, A.: Efficient multi-path protocol for wireless sensor networks. Int. J. Wireless Mobile Netw. **2**(1), 110–130 (2010)
7. Apostolopoulos, J.G., Tan, W., Wee, S.J.: Video streaming: concepts, algorithms, and systems, pp. 1–34 (2009)
8. Almalkawi, I., Guerrero Zapata, M., Al-Karaki, J., Morillo-Pozo, J.: Wireless multimedia sensor networks. Current trends and future directions. Sensors **10**(7), 6662–6717 (2010)
9. Advanced Video Coding for Generic Audio Visual Services: ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC (2009)
10. Greengrass, J., Evans, J., Begen, A.: Not all packets are equal, part i: streaming video coding and SLA requirements. IEEE Internet Comput. **13**(1), 70–75 (2009)
11. Boluk, P., Baydere, S., Harmanci, A.: Robust image transmission over wireless sensor networks. Mobile Netw. Appl. **16**(2), 149–170 (2011)

Performance Analysis of Collinear Beam-Steering Array Antenna

Shuchismita Pani, Arun Kumar and Malay Ranjan Tripathy

Abstract In the proposed design, a collinear beam-steering array antenna is designed and analyzed. The proposed antenna has used corporate feeding to have improved characteristics of array antenna. A single band at 2.45 GHz is observed for single element antenna whereas multiband response is obtained in case of 1×4 array antenna. The antenna gain of 3.94 and 6.53 dB are obtained for single element and array antenna, respectively. Further beam-steering characteristics is observed and analyzed with variation in phase of feeding signal in case of proposed collinear array antenna. The compact size and beam-steering features make the antenna suitable for intelligent wireless communication.

Keywords Collinear array antenna · Return loss · Radiation pattern
Antenna gain · Beam steering

1 Introduction

In the last few years in the systems development of communication, we need the reduced weight, optimized cost and much efficient antennas which are proficient of sustaining excellent functioning for frequencies of a broad range of spectrum. These technical trends had concentrated on the efforts to the design of a microstrip patch antenna. In wireless communications field microstrip antenna plays an important role which may be microstrip patch antenna, parabolic reflector antenna, slotted

S. Pani (✉) · M.R. Tripathy
Department of Electronics and Communication Engineering, Amity University,
Noida, UP, India
e-mail: spani@amity.edu

M.R. Tripathy
e-mail: mrtipathy@amity.edu

A. Kumar
AITEM, Amity University, Noida, UP, India
e-mail: akumar4@amity.edu

antenna and the folded dipole antenna, etc. All of them are noble in corresponding to their own functioning, applications and properties. Antennas can be treated as the strength and spine which provide its support for the wireless communication and hence it played an important and crucial role in the technological developments. From microstrip patch antenna only dual characteristics, dual frequency operation, circular polarizations (LHCP, RHCP), agility of frequency, flexibility in the feed-line, wide bandwidth, scanning and beam steering can be achieved easily. A linear array generally has more than one element and equally spaced with uniform excitations. In this we get continuous shifts in the phase within the elements of the radiating waves in the electromagnetic spectrum from the plane of microstrip patch array antenna and also the reconfigurable antennas which in the combined form called as to be the multiport antenna. It uses intrusion within the two orthogonal modes which causes interference. It could be prepared either to have the nulls direction under control [1, 2] or to emphasize the beam direction [3–5]. The reconfigurable antennas with the switches have many restrictions which can be overcome by using these antennas in case when we are maintaining small size.

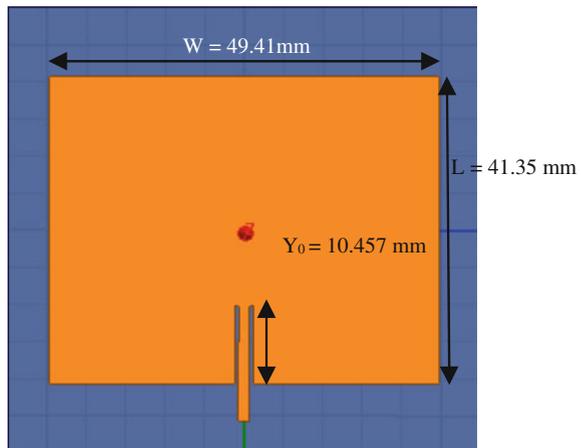
The supreme solution for various applications in the systems which includes control of traffic, radars for avoiding collision and for regulation (*S*-band 3 GHz) which are set up on most ocean ships in order to have detection during the adverse environmental conditions [6, 7] are the beam-steering antennas. Though, a positioning system is being required for mechanical scanning which can be scanned very slowly and costly. Due to this problem, electronic antennas for scanning are called as phased array antennas are being used which can Sweep the beam direction of the radiating element by changing the phase electronically, and hence creating a movable Pattern without parts which are moving. These antennas are steerable electronically, and hence we got a physical and stationary antenna [8, 9]. Microstrip phased array antenna was discussed in the switched phase shifter which are reconfigurable [10]. The phase shifter designed have configurable phase shift for many applications and can be used as beam steering phased array antenna for 360°. The undesirable consequence of isotropic antennas for the design and synthesis had established [11]. Mutual coupling effects and the radiation patterns in the microstrip linear array antenna have been described and analyzed [12]. Large arrays antennas was formulated. Also for the same criterion, the radiation patterns which are active had been taken to be same and equal for each and every elements. Thus, by the classical theory of array antenna we can calculate the overall radiated electric field of the array antenna by the multiplication of array factor with the active radiated radiation pattern of the individual integrated element. This theory is not applied on the microstrip arrays which are small, as the ground plane truncation creates the active edge elements pattern to diverge from the different pattern of the source near the array centre.

2 Antenna Design

Microstrip patch array antennas are very adaptable and versatile and are used to synthesize a radiation pattern which could not be obtained with the single element. Mostly they are used to scan the antenna system beam, directivity improvement and provide various functions which could not be obtained by using a single element. In the arrangement of feed network, the elements can be fed by using single or multiple lines, so the proposed array antenna is suitable for intelligent wireless applications and beam steering. A very important design parameters of a rectangular microstrip patch antenna is the operating frequency (f_0) hence, we choose the suitable value of solution frequency of the antenna at which it resonates. The ISM band Systems uses the 2.4–2.5 GHz frequency range. The resonant frequency selected for the proposed single element antenna design as shown in Fig. 1, is 2.4 GHz and at resonant frequency of 2.68, 3.69, 4.59, 5.09, 5.69 and 5.89 GHz for 1×4 array antenna which is applicable for multiband operation.

In the given work, value of substrate parameter, i.e. Effective relative permittivity = 2.11106 and thickness of the substrate (h) chosen to be 1.5 mm, distance between two patches from centre to centre = 150.00069 mm, distance between two patches from edge to edge = 100.59 mm. The W (width) = 49.41 mm, L (length) = 41.35 mm and $Y_0 = 10.457$ mm, Input impedance of the patch $Z_{in} = 204 \Omega$, as shown in Fig. 2.

Fig. 1 Geometry of single microstrip patch antenna



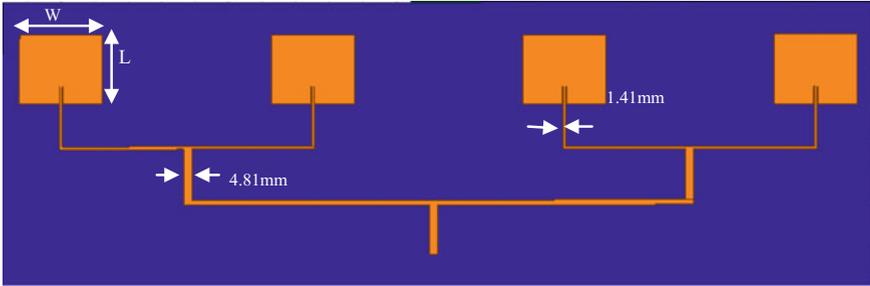


Fig. 2 Geometry of 1×4 collinear array antenna

3 Simulation Results and Validation

In this design all the tests are being carried out by using electromagnetic simulator (HFSS 13). High frequency structure simulator is the method of moment on the integral equations, and is basically a electromagnetic full-wave simulator. It is composed of electromagnetic simulator, layout editor, circuit simulator and schematic editor for the near field calculation program which has format converter and display for field program.

3.1 Return Loss (S_{11})

The measure of the signals reflecting back (S_{11}) and practically it should be as small as possible for an antenna to be feasible. S_{11} basically means signals transmitted from a specific port and reflected back to the same port. So for an optimum antenna, ideally the value of S_{11} should be 0 which is not possible practically due to the losses. This proposed antenna has a return loss -13.65 dB at a resonant frequency of 2.4 GHz for a single patch as shown in Fig. 3 and -15.62 , -17.51 , -17.37 , -17.22 , -18.39 , -18.97 dB at resonant frequency of 2.68, 3.69, 4.59, 5.09, 5.69 and 5.89 GHz, respectively, for 1×4 linear array as shown in Fig. 4.

3.2 Radiation Pattern

It denotes the strength of the electromagnetic radiation in various directions. It has a value of maximum gain of 3 dB for single element as shown in Fig. 5 and gain of 7 dB at 30° for 1×4 microstrip array antenna as shown in Fig. 6.

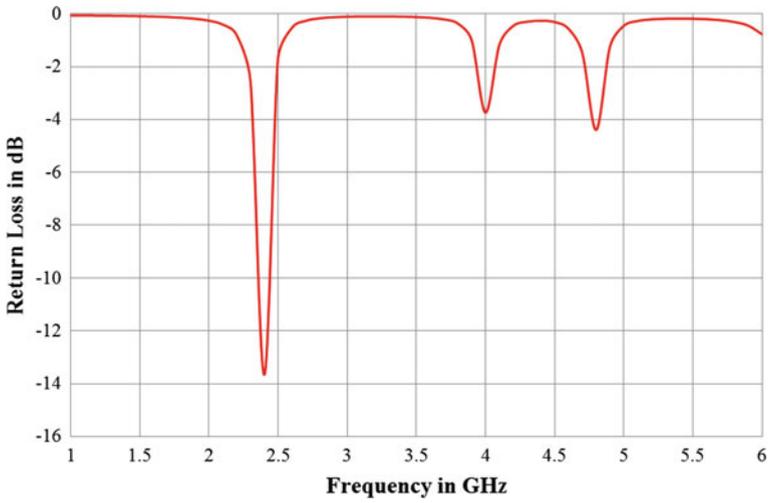


Fig. 3 Showing the S_{11} for a single microstrip patch

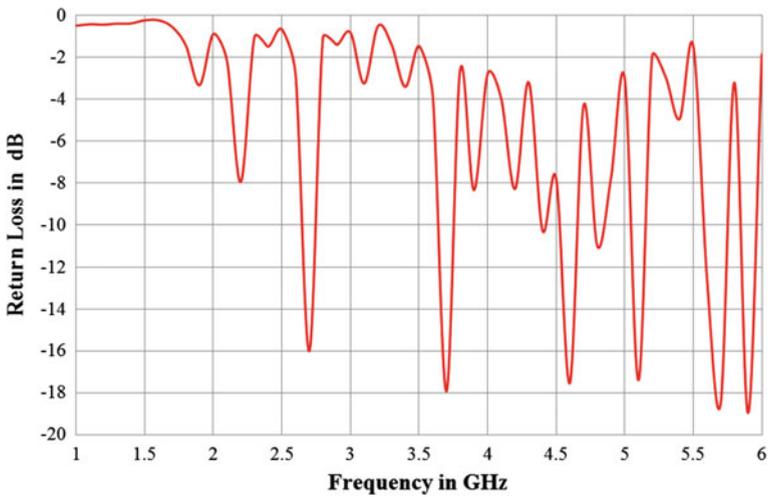


Fig. 4 Showing the S_{11} for a 1×4 linear array antenna

3.3 3-D Polar Plot

3D polar plot is mainly used to evaluate the exact value of the gain of the antenna. Hence for a single patch, the gain is 3.94 and 6.53 dB in case of 1×4 collinear array antenna (Figs. 7 and 8).

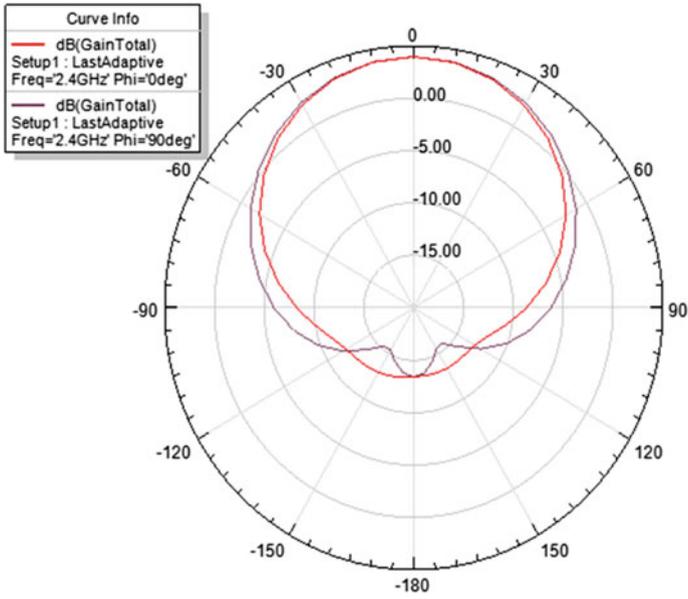


Fig. 5 Radiation pattern for a single patch

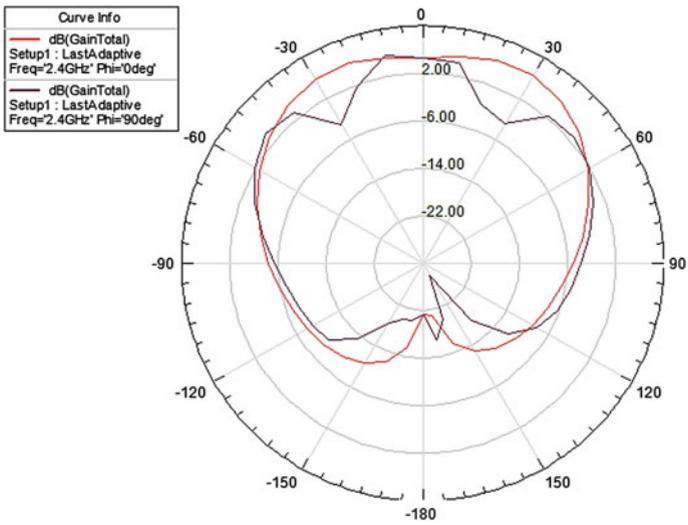


Fig. 6 Radiation pattern for 1×4 array antenna

Fig. 7 Polar plot for a single patch

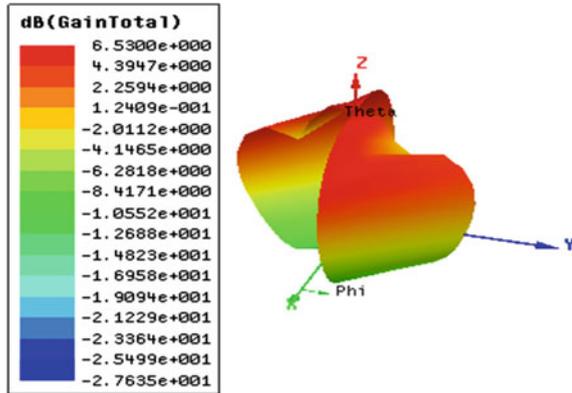
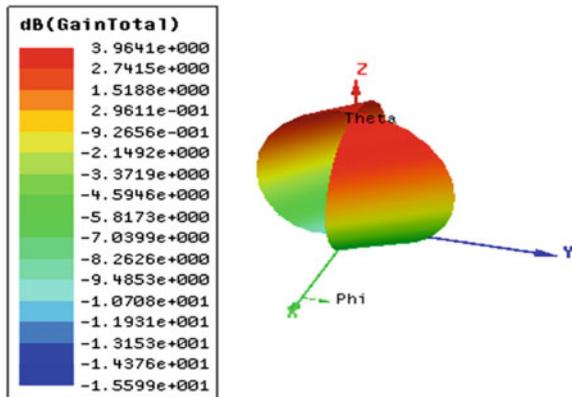


Fig. 8 Polar plot for 1 × 4 array antenna



3.4 Directional Beam

The radiation pattern of the beam-steering planar array with N no. of linear sub-arrays of M patch radiators is the multiplication of the pattern of element and space factor of the planar array [13] as given in below equation

$$F_{\text{planar}}(\theta, \varphi) = F_1(\theta, \varphi) \sum_{m=1}^M e^{jkd_r(m-1)(\cos \theta - \cos \theta_0)} \sum_{n=1}^N e^{jkd_h(n-1)(\sin \theta \cos \varphi - \sin \theta_0 \cos \varphi_0)}$$

where, space difference between two corresponding elements is denoted as d_h . The array antenna beam cannot be steered so far from broadside than a particular angle, as the element pattern has nulls in the grazing direction. So the direction of maximum radiation is different from the value Φ_0 . The original beam inclinations are less than $(90 - \theta_0)$ for the respective array of ideal isotropic radiators placed in free space.

For proposed antenna $m = 1$ and $n = 4$, θ and θ_0 both are 90° . So the beam direction is changing by varying the value of Φ . The analysis is carried out with MATLAB and explained in this research paper. When $\Phi = \Pi/2$, the directional beam is seen exactly at 90° when it is calculated for 2.4 GHz as shown in Fig. 9.

Fig. 9 Directional beam for array at $\phi = \pi/2$

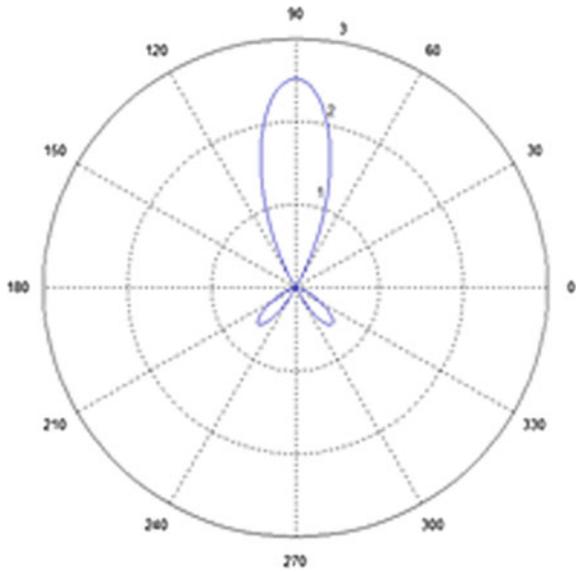
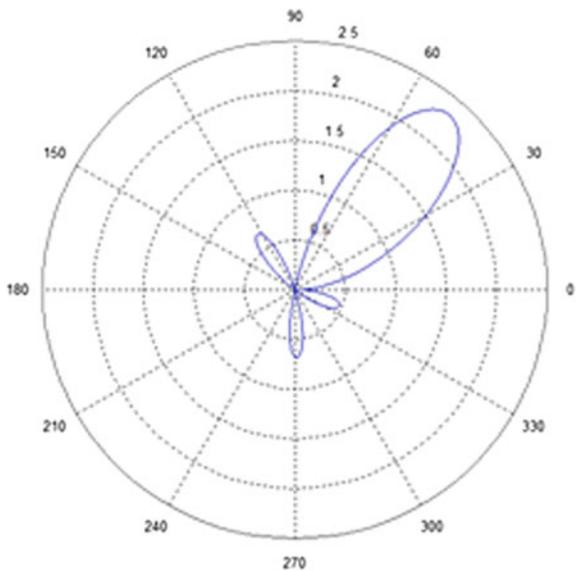


Fig. 10 Directional beam for array at $\phi = \pi/4$



The directional beam is approaching to 50° when the Φ value is changed to $\Pi/4$ as shown in Fig. 10.

Similarly when the resonant frequency at 5.8 GHz calculated and simulated through MATLAB, it is seen that the beam directed to 270° without any side lobes

Fig. 11 Directional beam for array at $\phi = \pi/2$

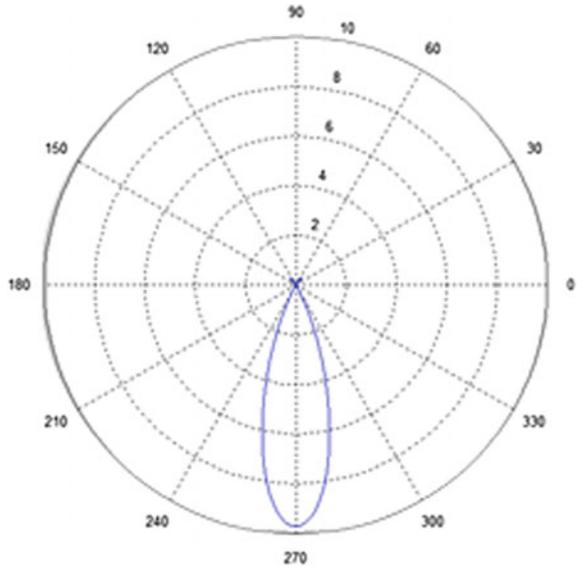
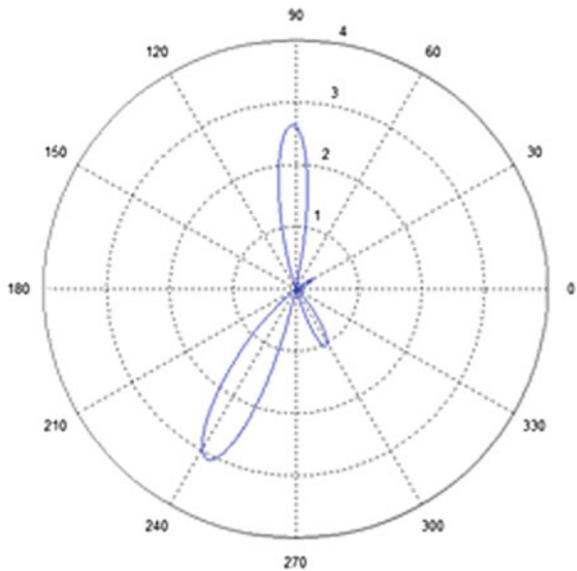


Fig. 12 Directional beam for array at $\phi = \pi/4$



when $\Phi = \Pi/2$ as shown in Fig. 11. When $\Phi = \Pi/4$, the directional beam covers both 90° and 240° shown in Fig. 12, which is very useful for wireless applications.

4 Conclusion

Collinear beam-steering array antenna is designed and analyzed in this paper. It is seen to have better operating frequency, return loss, radiation pattern and antenna gain as compared to single element. The beam-steering feature is further analyzed by using MATLAB software. It is observed that the wide angle (50° , 90° , 239° , 270°) beam steering can be achieved with different phases of feeding signals. This array antenna can be used for future generation intelligent wireless system.

References

1. Labadie, N.R., Sharma, S.K., Rebeiz, G.M.: A circularly polarized multiple radiating mode microstrip antenna for satellite receive applications. *IEEE Trans. Antennas Propag.* **62**(7), 3490–3500 (2014)
2. Deng, C., Li, Y., Zhang, Z., Feng, Z.: A hemispherical 3-D null steering antenna for circular polarization. *IEEE Antennas Wireless Propag. Lett.* **14**, 803–806 (2015)
3. Narbudowicz, A., Bao, X.L., Ammann, M.J., Shakhtour, H., Heberling, D.: Circularly polarized antenna with steerable dipole-like radiation pattern. *IEEE Trans. Antennas Propag.* **62**(2), 519–526 (2014)
4. Narbudowicz, A., Bao, X.L., Ammann, M.J.: Omnidirectional microstrip patch antenna with reconfigurable pattern and polarisation. *IET Microwaves Antennas Propag.* **8**(11), 872–877 (2014)
5. Narbudowicz, A., Ammann, M., Heberling, D.: Switchless reconfigurable antenna with 360° steering. *IEEE Antennas Wireless Propag. Lett.* (2016)
6. Blackman, S.: *Multitarget tracking with radar applications*. Artech House (1986)
7. Pierowicz, J.A.: Intersection collision avoidance using IVHS countermeasures. Task 5 Interim Report: Design of Testbed Systems, vol. 1, Calspan Report No. 8169-8 29 Jan 1997
8. Ehmouda, J., Briqech, Z., Amer, A.: Steered microstrip phased array antennas. *World Acad. Sci. Eng. Technol.* **49** (2009)
9. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: *Microstrip Antenna Design Handbook*. Artech House, Boston
10. Anand, P., Sharma, S., Sood, D., Tripathi, C.C.: Design of compact reconfigurable switched line microstrip phase shifters for phased array antenna. In: *International Conference on Emerging Technology Trends in Electronics, Communication and Networking* (2012)
11. Schlosser, E.R., Heckler, M.V.T., Sperandio, M., et al.: Synthesis of linear antenna arrays for radio base stations. In: *Proceedings of IEEE Antennas Propagation Society International Symposium, Orlando, USA*, pp. 600–601, July 2013
12. Pozar, D.M.: The active element pattern. *IEEE Trans. Antennas Propag.* **42**(8), 1176–1178 (1994)
13. (Tuli) Herscovici, N., Christodoulou, C.: Analysis of beam-steering and directive characteristics of adaptive antenna arrays for mobile communications. *IEEE Antennas Propag. Mag.* **43** (3) (2001)

E-Mail Spam Filtering: A Review of Techniques and Trends

Alexy Bhowmick and Shyamanta M. Hazarika

Abstract We present an inclusive review of recent and successful content-based e-mail spam filtering techniques. Our focus is mainly on machine learning-based spam filters and variants inspired from them. We report on relevant ideas, techniques, taxonomy, major efforts, and the state-of-the-art in the field. The initial interpretation of the prior work examines the basics of e-mail spam filtering and feature engineering. We conclude by studying techniques, evaluation benchmarks, and explore the promising offshoots of latest developments and suggest lines of future investigations.

Keywords Spam · Spam filtering · Techniques · False positive
Machine learning

1 Introduction

E-mail or electronic-mail is a fast, effective, and inexpensive method of exchanging messages over the Internet. Whether it is a personal message from a family member, a company-wide message from the boss, researchers across continents sharing recent findings, or astronauts staying in touch with their family (via e-mail uplinks or IP phones), e-mail is a preferred means for communication. Used worldwide by 2.3 billion users, at the time of writing the article, e-mail usage is projected to increase up to 4.3 billion accounts by 2016 [1]. But the increasing dependence on e-mail has induced the emergence of many problems caused by ‘illegitimate’ e-mails, i.e., *spam*. According to the Text Retrieval Conference (TREC) the term

A. Bhowmick (✉)

School of Technology, Assam Don Bosco University, Guwahati 781017, Assam, India
e-mail: alexy.bhowmick@dbuniversity.ac.in

S.M. Hazarika

Department of Computer Science and Engineering, Tezpur University, Tezpur 784028, Assam, India
e-mail: smh@tezu.ernet.in

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_61

583

‘spam’ is—any unsolicited e-mail that is sent indiscriminately [2]. Spam e-mails are unsolicited, un-ratified, and usually mass mailed. Spam being a carrier of malware causes the proliferation of unsolicited advertisements, fraud schemes, phishing messages, explicit content, promotions of cause, etc. On an organizational front, spam effects include: (i) annoyance to individual users, (ii) less reliable e-mails, (iii) loss of work productivity, (iv) misuse of network bandwidth, (v) wastage of file server storage space and computational power, (vi) spread of viruses, worms, and Trojan horses, and (vii) financial losses through phishing, denial of service (DoS), directory harvesting attacks, etc.

Figure 1 depicts the e-mail architecture and how e-mail works. Spam is a broad concept that is still not completely understood. In general, spam has many forms—chat rooms are subject to *chat spam*, blogs are subject to *blog spam* (splogs), search engines are often misled by *web spam* (search engine spamming or spamdexing), while social systems are plagued by *social spam*. This paper focuses on ‘*e-mail spam*’ and its variants, and not ‘spam’ in general. Prior attempts to review e-mail spam filtering using machine learning have been made, the most notable ones being [2–7]; most recent empirical studies being [8–10]. We extend earlier surveys by taking an updated set of works into account. We present a content analysis of the major spam-filtering surveys over the period (2004–2015). Significant amounts of historical and recent literature, including gray literature were studied to report recent advances and findings. We believe our survey is of complementary nature and provides an inclusive review of the state-of-the-art methods in content-based e-mail spam filtering. Our work addresses the following:

- *First*, we perform an exploration of the major spam characteristics and discuss feature engineering for spam e-mails.
- *Second*, we present a qualitative summary of major surveys on spam e-mails over the period (2004–2015) and taxonomy of content-based approaches to e-mail spam filtering.

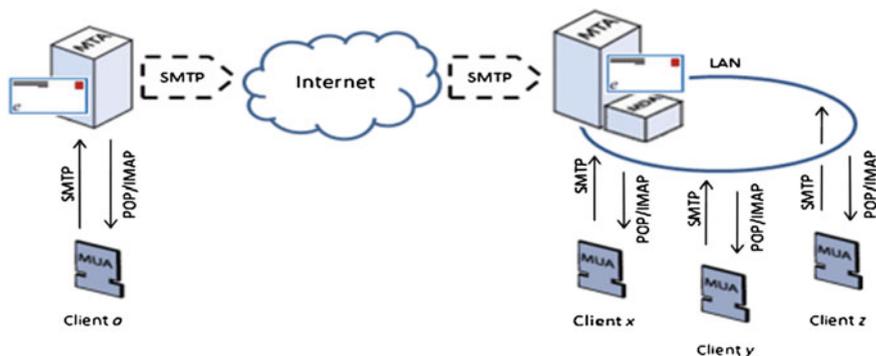


Fig. 1 The e-mail architecture

- *Third*, the article reports on evaluation measures, bench marks and new findings and suggest lines of future investigations for emerging spam types.

2 Feature Engineering

Feature selection is a key issue and has become the subject of much research. It has mainly three objectives: (i) enhancing the classifier’s predictive accuracy, (ii) building effective and economical classifiers, and (iii) obtaining a better understanding of the elementary process involved in generation of data. Dimensionality reduction and feature subset selection are two preferred techniques for lowering the feature set dimension. While feature subset selection involves the extraction of a subset of the original attributes, dimensionality reduction involves linear combinations of the original feature set.

Table 1 presents a summary of feature extraction and selection in popular literature. This review article also examined a number of major earlier surveys on spam filtering over the period (2004–2015). A summary of popular machine

Table 1 A summary of feature extraction and feature selection techniques in popular literature

References	Year	Approaches
[12]	2004	Studied subject line, header, and message body. Employed information gain (IG), document frequency (DF), and chi-square test for selecting features. Found <i>bag of words</i> model quite effective on spam filtering, and header features as important as message body
[13]	2006	Extracted fixed-length character <i>n-grams</i> and variable-length character <i>n-grams</i> . Explored information gain (IG) as a feature selection technique. Character <i>n-grams</i> were noted to be richer and definitive than word-tokens
[14]	2006	Considered features of three types: <i>word, character, structured feature</i> in a feature-based <i>versus</i> feature-free comparison. Employed information gain (IG) as a feature selection technique. Noted feature-free methods to be more correct than the feature-based systems, however feature-free approaches took much longer than feature-based approach in classifying e-mails
[15]	2005	Used behavioral patterns of spammers, Meta-heuristics as features Employed term frequency, inverse document Frequency (TFIDF), SpamKANN for feature selection. Tested SVM, Decision trees, Naive Bayes to get increased prediction accuracy than keywords
[16]	2003	Experimented on features: header (H), textual (T), handcrafted features (HH), etc. Different ways of feature selection for Decision Tree and Naive Bayes models were evaluated. The usefulness and importance of different type of features were discussed in detail in experiments
[17]	2006	Considered subject, body, header, attachment feature. Analyzed strength and weaknesses of document frequency (DF), Information Gain (IG), Chi-square test, and Mutual Information. Presented a deep analysis of feature selection methods. Found e-mail attachments to be useful when integrated with models

learning-based techniques categorizing them according to perspective (Algorithm, Architecture, Methods, and Trends) is presented in Table 2.

Articles classified under ‘*Algorithm*’ reflect research that focused on classification algorithms and their implementations and evaluations. Articles classified under ‘*Architecture*’ concentrated on development of spam filtering infrastructures. Articles classified under ‘*Methods*’ refers to study of the existing filtering methods while ‘*Trends*’ speaks of discourses concentrating on emerging methods and the adaptation of spam filtering methods over time. Limitations listed in the last

Table 2 A summary of popular machine learning-based spam filtering attempts by authors according to perspective with their strengths and limitations

References (Year)	Perspective	Strength and limitations
[18] (2004)	Naive Bayes, k-NN, ANN, SVM Algorithms, methods	Techniques benefits beginners <i>Does not</i> deal with feature selection
[3] (2006)	Naive Bayes, Logitboost, SVM Algorithm, methods, trends	Resulted in— <i>LingSpam</i> and <i>PUI Ignored</i> headers, HTML, attachments
[4] (2006)	Bayesian filtering Methods, architecture	Broad review of implementations <i>Focuses primarily</i> on automated, filters
[5] (2008)	SVM, TF-IDF, boosting Algorithms, methods, trends	Explains feature extraction methods <i>Does not cover</i> neighboring topics
[2] (2008)	SVM, perceptron, OSBF Algorithms, methods, trends	Testing achieves FPR = 0.2% User feedback <i>difficult to simulate</i>
[6] (2009)	Regression, ensembles Algorithms, methods	Focuses on textual and image analysis. Focuses <i>only</i> on application specific aspects
[2] (2010)	SVM, Naive Bayes Algorithms, methods	Proposed Matthews correlation coefficient (MCC). Need for comparison
[3] (2012)	MDL principle, SVM Algorithms, methods	Uses six, well known, large public databases. Bogofilter, SpamAssassin <i>not considered</i>
[15] (2012)	Signature, k-NN, ANN, SVM Methods, architecture	Focuses on distributed computing paradigms. <i>Avoids</i> interoperability issues
[7] (2013)	Statistical analysis, n-grams Trends	Investigated <i>topic drift</i> Limited datasets
[8] (2015)	Naïve Bayes, J48, Clustering Methods	Comparative study of different techniques Limited datasets and tools

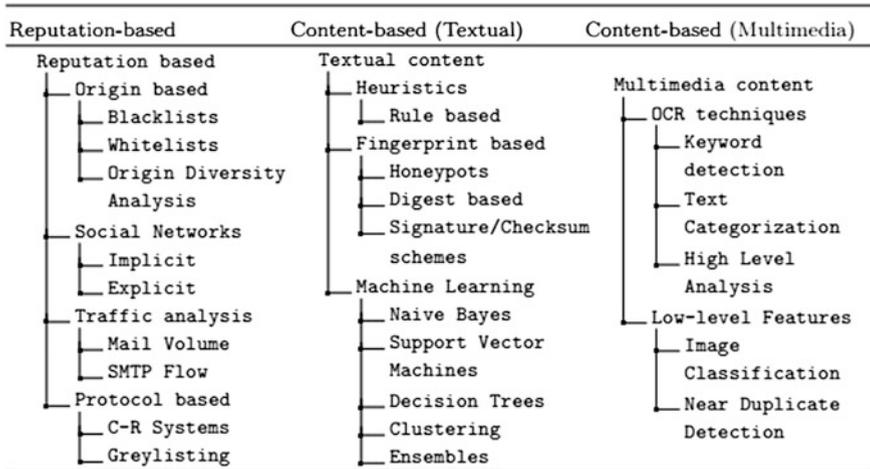


Fig. 2 A taxonomy of e-mail spam filtering techniques

column, corresponding to each article are as acknowledged by the authors themselves. Perusing the different spam techniques and the methods used by researchers to combat spam, taxonomy of spam filtering techniques is presented (Fig. 2) next.

3 Publicly Available Datasets

Most of the datasets publicly available are static datasets with very few concept drift datasets. Many authors construct their own image spam or phishing corpus. Table 3 lists public corpora with associated information used in spam filtering experiments.

4 Future Trends and Conclusion

Models built on old data become less accurate or inconsistent making the rebuilding of the model imperative (called *virtual concept drift*). Spam filtering is a dynamic problem that involves concept drift. While the understanding of an unwanted message may remain the same, the statistical properties of the spam e-mail changes over time since it is driven by spammers involved in a never-ending arms race with spam filters. Another reason for concept drift could be the different products or scams driven by spam that tends to become popular. The dynamic nature of spam is one of its most testing aspects. An effective spam filter must be able to track target concept drift, swiftly adapt to it, and have a successful mechanism to identify the drift or evolution in spam features.

Table 3 Public corpora used in e-mail spam filtering experiments

Corpus name	No. of messages (Spam Ham)		Spam rate (%)	Year of creation	References
SpamAssassin	1897	4150	31	2002	[17]
Enron-Spam	13,496	16,545	–	2006	[19]
Ling Spam	481	2412	17	2000	[20]
PU1	481	618	44	2000	[7]
PU2	142	579	20	2003	[12]
PU3	1826	2313	44	2003	[12]
PUA	571	571	50	2003	[12]
Gen Spam	41,404		78	2005	[17]
Spambase	1813	2788	39	1999	[20]
ZH1	1205	428	74	2004	[12]
TREC 2005	52,790	39,399	–	2005	[4]
TREC 2006	24,912	12,910	–	2006	[5]
TREC 2007	50,199	25,220	–	2007	[18]
Spam archive	>2,20,000		100	1998	[3]
Biggio	8549	0	–	2005	[12]
Princeton spam benchmark	1071	0	–	–	[12]
Spam archive	>2,20,000		100	1998	[3]
Dredze dataset	3927	2006	–	2007	[3]
Phishing corpus	415	0	–	2005	[3]

Content-based spam filtering systems, though widely adopted as a successful spam defense strategy, has unfortunately substituted the spam issue with a false positive one. Such systems achieve a high accuracy but there exists some false positive tradeoff. False positives are *more severe and expensive* than spam. Reduction of false positives is another domain in email spam analysis where much work needs to be done on leveraging existing algorithms. Future researches must address the fact that e-mail spam filtering problem is co-evolutionary, since spammers attempt to outdo the advances in predictive accuracy of the classifiers all the time.

One of the biggest spam problems today even as spam e-mail volumes associated with botnets are receding is the *snowshoe spam*. Snowshoe spamming is a technique that uses multiple IP addresses, websites, and sub-networks to send spam, so as to avoid detection by spam filters. Spammers operate by distributing their spam load across a wide footprint of systems to keep from sinking, just as snowshoe wearers do. With many users today migrating to social networks as a means of communication, spammers are diversifying in order to stay in business.

E-mail prioritization is an urgent research area with not much research done. In addition to basic communication, e-mail systems are used for a wide variety of other tasks such as—business and personal communication, advertisements, reminders, management of tasks, and cloud storage, etc. There is a serious need to

address the information overload issue by developing systems that can learn personal priorities from data and identify important e-mails for each user. Prioritizing e-mail as per its importance or classifying emails into personalized folders as in [9, 11] is another desirable characteristic in a spam filter. Prioritizing e-mail or perhaps redirecting urgent messages to handheld devices could be another way of managing e-mails.

Fortunately, machine learning-based systems enable systems to learn and adapt to new threats, reacting to counteractive measures adopted by spammers. No single anti-spam solution may be the right answer. A multi-faceted approach that combines legal and technical solutions and more is likely to provide a death blow to such spam. As long as spam exists it will continue to have adverse effects on the preservation of integrity of e-mails and the user's perception on the effectiveness of spam filters. Overall remarkable advancements have been achieved and continue to be achieved, however, some outstanding problems in e-mail spam filtering as highlighted above still remain. Till more improvements in spam filtering happen, anti-spam research will remain an active research area.

References

1. Radicati: Email Statistics Report, 2012–2016 Executive Summary. Technical Report 650, Radicati (2016)
2. Cormack, G.V.: Email spam filtering: a systematic review. *Found. Trends Inf. Retrieval* **1**(4), 335–455 (2008)
3. Androutsopoulos, I., Paliouras, G., Michelakis, E.: Learning to filter unsolicited commercial e-mail. Technical Report in National Centre for Scientific Research Demokritos, Athens, Greece (2006)
4. Carpinter, J., Hunt, R.: Tightening the net: a review of current and next generation spam filtering tools. *Comput Secur.* **25**(8), 566–578 (2006)
5. Blanzieri, E., Bryl, A.: A survey of learning-based techniques of email spam filtering. *J. Artif. Intell. Rev.* **29**(1), 63–92 (2008)
6. Guzella, T.S., Caminhas, W.M.: A review of machine learning approaches to spam filtering. *Expert Syst. Appl.* **36**(7), 10206–10222 (2009)
7. Wang, D., Irani, D., Pu, C.: A study on evolution of email spam over fifteen years. In: Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Work sharing (CollaborateCom), Austin, TX, USA (2013)
8. Vyas, T., Prajapati, P., Gadhwal, S.: A survey and evaluation of supervised machine learning techniques for spam e-mail filtering. In: Proceedings of IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 5–7 March 2015
9. Alsmadi, I., Alhami, I.: Clustering and classification of email contents. *J. King Saud Univ. Comput. Inf. Sci.* **27**, 46–57 (2015)
10. Li, W., Meng, W.: An empirical study on email classification using supervised machine learning in real environments. In: IEEE International Conference on Communications (ICC), IEEE, 8–12 June 2015
11. Sethi, H., Sirohi, H., Thakur, M.K.: Intelligent mail box. In: Proceedings of Third International Conference India 2016, vol. 3, pp. 441–450 (2016)

12. Zhang, L., Zhu, J., Yao, T.: An evaluation of statistical spam filtering techniques spam filtering as text categorization. *ACM Trans. Asian Lang. Inf. Process. (TALIP)* **3**(4), 243–269 (2004)
13. Kanaris, I., Kanaris, K., Houvardas, I., Stamatatos, E.: Words versus character n-grams for anti-spam filtering. *Int. J. Artif. Intell. Tools* **16**(6), 1–20 (2006)
14. Delany, S.J., Bridge, D.: Feature based and feature free textual CBR: a comparison in spam filtering. In: *Proceedings of the 17th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'06)*, pp. 244–253 (2006)
15. Yeh, C.Y., Wu, C.H., Doong, S.H.: Effective spam classification based on meta-heuristics. In: *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 3872–3877 (2005)
16. Diao, Y., Lu, H., Wu, D.: A comparative study of classification based personal e-mail filtering. In: *Knowledge Discovery and Data Mining. Current Issues and New Applications*, pp. 408–419 (2003)
17. M'endez, J.R., D'iaz, F., Iglesias, E.L., Corchado, J.M.: A comparative performance study of feature selection methods for the anti-spam filtering domain. In: *Advances in Data Mining. Applications in Medicine, Web Mining, Marketing, Image and Signal Mining*, pp. 106–120. Springer, Berlin, Heidelberg (2006)
18. Tretyakov, K.: Machine learning techniques in spam filtering. In: *Data Mining Problem-Oriented Seminar, MTAT.03.177*, pp. 60–79 (2004)
19. Koprinska, I., Poon, J., Clark, J., Chan, J.: Learning to classify e-mail. *Inf. Sci.* **177**(10), 2167–2187 (2007)
20. Sakkis, G., Androutsopoulos, I., Paliouras, G., Karkaletsis, V.: Stacking classifiers for anti-spam filtering of e-mail. In: *Empirical methods in Natural Language Processing*, pp. 44–50 (2001)

Stealthy Attacks in MANET to Detect and Counter Measure by Ant Colony Optimization

Abhishek Ranjan, Rajalakshmi Selvaraj, Venu Madhav Kuthadi
and Tshilidzi Marwala

Abstract MANET (Mobile Ad Hoc Network) is the active wireless network which can produce communication less connections in which every node can perform as a router. Every node in the MANET is responsible for dynamically creating other nodes for communication. Protection is an important examine for the wireless and wired network communications in MANET. The achievement of MANET (mobile ad hoc networks) strappingly depends upon people's assurance in its sanctuary. Though, uniqueness of MANET poses both opportunity and challenges in the accomplishment of security purpose, like authentication, privacy, access, availability, integrity, non-repudiation, and control. The malicious node presence is being a major challenge in MANET for designing the vigorous security solution which can defend the MANET from different routing attacks. We have categorized the existing security services and security threats in this mobile ad hoc network paper. The proposed approach gives a better solution in the achievement of present security threats in MANET. Based on monitoring system a guard node is recommended in this paper through which every activity occurred inside the network can be constantly monitored. If there is any abnormal action of Stealthy attack occurred inside the network then the Guard node collects and forwards the abnormal information toward the source node. In this guard node an algorithm based on Ant Colony optimization is deployed to recognize the threats which are present in the

A. Ranjan (✉) · T. Marwala
Faculty of Engineering and the Built Environment, University of Johannesburg,
Johannesburg, South Africa
e-mail: abhishek.ranjan@bothouniversity.ac.bw

T. Marwala
e-mail: tmarwala@uj.ac.za

Rajalakshmi Selvaraj · V.M. Kuthadi
Department of Computer Science, BIUST, Gaborone, Botswana
e-mail: selvarajr@biust.ac.bw

V.M. Kuthadi
e-mail: vkuthadi@uj.ac.za

V.M. Kuthadi
Department of AIS, University of Johannesburg, Johannesburg, South Africa

MANET's network. This proposed system detects the stealthy attack and also resists the attackers from the network. Therefore, this approach makes the enhancement in the network packet delivery and its performance ratio, etc.

Keywords Mobile ad hoc networks (MANET) · Stealthy attacks
Guard node · Ant colony-based optimization

1 Introduction

MANET is a type of wireless network having WMN (Wireless Mobile Nodes). It is a distributed method that dynamically and freely self-organizes into temporary, arbitrary, and ad hoc network topologies which allows connection inside the network without centralized administered control management and pre-existing communication infrastructure. Every network is having security vulnerabilities, as like MANET. Every node inside a network performs like router which ahead the data packet to further node. If source node demands route toward destination for which it don't contain any route, it first transmit the route request packet (RREQ) crossways network. After that, RREP send rear to source node, these nodes arrangement forward pointer toward destination. Formerly source node obtains RREP; it begins to promote the data packet to destination. If the source gets RREP with greater no of sequence than already Sended (RREQ) or having the similar sequence with less hop count a Stealthy attack is anticipated to jumble the route and then drop or eavesdrop the packets assuming a possible threat, the Safety property that may bring up to date its routing protocol for destination and use a better route. Through the RREP, Stealthy attack can be directly replied with very high 1 hop count number and sequence number to source node, destination node can also select a route with nominal hop count and reply a route reply packet (RREP).

When there is a stealthy attack then malicious node always gives a positive route request, even though it might not contain suitable route toward destination. As malicious node does not verify its routing access, it is at all times first one that replies route request message. As a result, approximately traffic inside malicious neighbor node is intended toward malevolent node which possibly drops every packet. The ant colony optimization method converts this rule to its mathematical equivalent to spot the stealthy attack in wireless mesh network [1, 2]. Additional advantage of ant colony optimization includes its flexibility and simplicity. It is located inside guard node and has the ability to tackle the problems with incomplete and inaccurate data. The non linear functions of arbitrary complex, based on Ant colony optimization prefers the shortest path procedure which come across shortest length path among entire possible paths in network for reaching destination.

There are two pre-processing steps in Ant colony optimization that are neighborhood construction and data set reduction. The data set reduction assists to decrease processing time and storage requirements. Neighborhood construction removes proximity, density, and local connectivity. Ant colony-based optimization

is stimulated by real ant's behavior. As the ants look for foodstuff into ground and deposited import that are known as pheromone resting on its path. The pheromone focuses on path assists to direct a colony to food sources. By interacting with environment the ant colony effectively discovers food sources.

The Solution construction, pheromone update mechanism, and solution representation are main choice to design Ant colony optimization. If there is a Stealthy attack occurred into the network then the abnormal information will be captured and forwarded by the guard node toward the source node where Ant colony optimization-based algorithm is deployed for the recognition of threats that are present in MANET network.

2 Literature Survey

Few years before, the researchers has been a vigorously explored many different methods to ensure control on data traffic and its security into the wireless network. These methods are generally classified into the various classes—integrity and authentication services, protocol which depends on the path assortment, protocol which needs detailed acknowledgement or uses any statistical methods, protocols which use specific hardware, protocol which overhear its neighbor communiqué.

The techniques of path assortment enlarge route strength by initial determining various path routes [3, 4] and next these paths are used for providing data transmission redundancy among source and destination. The data are subdivided into several distributes and sent toward destination through dissimilar other routes. This scheme is useful in connected network, but doesn't gives sufficient path diversity inspire the network. In addition, these methods are quiet expansive for controlled wireless networks because of data redundancy.

Moreover, these protocol can be weak for route discovery attack, so that the Stealthy attack can prevents the formation of non adversarial route. The malicious node detection is important to find abnormal activities occurred through this Stealthy attack. The technique proposed for detecting a malicious behavior specific data dropping, depends upon clear acknowledgement to receive data using similar channel [4], otherwise channel out-of-band is received [5]. This method cause Stealthy attack drop visible toward the end point. Though, the technique earns a high communication operating cost it is to be expanded with further methods for isolation and malicious node diagnosis.

A natural conservatory will be for reduction in control significance above head after reducing the frequency by one in each N no. of statistics communication (greater than paper $N = 1$). On the other hand, this may cause delay in adversary detection and cause significant damage in result. In addition, every approach may suffer a non convergent performance which means standing of excellent node having difficulties at lower value or the malicious node is wrongly evaluated. Supportive local monitoring is a method applied for organized and control information delivery misbehavior into the MANET [1, 4–8]. Research Work in [4] gives

a method to discover routes having certain advantageous property like as node disjoint. This paper [8] gives reorganization of an extensive class control of stealthy attacks next to static sensor networks. Though, every performance-based process (mutual communication and noncommunication-based) are as utilized by every researchers to reduce Stealthy attack packet reducing obtained in this paper [9, 10]. The pheromone focus of a route quality is finest indication of its quality. After every step of ant its result are compared and for its best result the concentration of pheromone is increased for equivalent keywords. The end of algorithm when there is no extra addition of individual is present at the knowledge base. Under this circumstance, the method Ant colony optimization is more efficient. For applying Ant colony optimization, the optimization difficulties are altered into problem for finding the best route into weighted graph. Artificial ants (here after ants) increasingly develop a solution by moving onto the graph. The process of solution construction is biased and is stochastic by pheromone model, i.e., set of different parameters related with components of graph.

3 Proposed Work

3.1 Overview

The mobile nodes in MANET plays significant role in routing and in hosting, it also maintain the multi hop communication among the nodes. Mobile nodes throw data packets to every further node with the help of routing protocol in MANET. Some of unique feature of MANETs like communication through wireless links, resource limitation (battery power and bandwidth), support among nodes and dynamic topology build it extra susceptible to attacks. Malicious attack performance in the network will not completely forward any of the traffic to its entire neighbor node. When there is a stealthy attack inside the network then the Ant colony optimization is applied for the performance in MANET. This Ant colony optimization protocol checks Stealthy attack and also helps in the improvement of MANET. The Ant colony optimization system helps in the no. of link changes, traffic movement source and several numbers of generations of solicitations through the source.

This system gives the ability of proposal in the stipulations of packet loss and normalizes overhead. This is a very effective and simple way of providing security against the stealthy attack by the technique of Ant colony optimization. The attack prevention technique detects the Stealthy attack and isolate from routing and active data forwarding and it also reacts when sending response to its neighbors. Probability set: the route selection depends upon the probabilistic value of route, because it is based on the ant behavior. The routes probability value depends on the feasibility of path $Z_{ij}(a)$, Heuristic information $X_{ij}(a)$, and Pheromone value $Y_{ij}(a)$ of the route for an ant. Two more parameter c' and c^\wedge are used there for calculating the probability of a route. The used parameters c' and c^\wedge control the desirability with respect to visibility. c' and c^\wedge which are present with heuristic

and pheromone value of routes, respectively. Proposed Ant colony optimization algorithm also helps in getting knowledge of the present node as well as the feasible route from the current node to its next node and its history of already transverse rote.

3.2 Overall Proposed Architecture

Construct the MANET nodes and define the source and destination. To transmit the packet from source to destination node it finds shortest path. Source node request to the nearby node to transmit the packet to the destination and it checks the node whether it stealthy or not. If next node is stealthy then it finds the other shortest path to transmit the packet. Based on the Ant colony technique it finds the shortest path to send the packet. Using route discovery it finds the minimum distance path to transmit the packet. Then the destination packet replay to the source about transmission of packet. The source node finds the shortest path using route discovery. In route discovery, it finds the minimum distance path to transmit the packet to the destination. If any stealthy node is present in the MANET, it chooses another path to transmit the packet (Fig. 1).

Stealthy node is a malicious node it can't able to transmit the packet to destination. Using route discovery, the path distance can be calculated and it finds the minimum distance path to reach the destination. If any node malicious in between the MANET node, the malicious path can be avoided and find another shortest path to transmit the packet to the destination.

3.3 Methodology

3.3.1 Ant Colony Optimization

Routing algorithm Ant colony optimization gets stimulated by ant's activities and by the Ant colony optimization associated field to handle the difficulty of routing into the communication networks. Most significant source of idea is originated in capability of convinced types of ants (example: families of Argentine ants *Linepithema Humile*) for finding the shortest path between their food source and nest by utilization of impulsive chemical material called pheromone. As the ants travel among the food source and the nest depart traces of pheromone and they also moves in preferentially high pheromone intensities. As shortest route can be accomplished firstly, they obtain higher pheromone level as earlier; drawing more number of ants, which in turns increases the number of pheromone. These reinforcement methods allow the colony to come together on the shortest route. It develops the route of the work in the region of Ant colony optimization.

In maximum case, an artificial ant leave specific quantity of pheromone characterized by $\Delta\tau_{p,q}$ after finishing its path and not in additional way at the time of

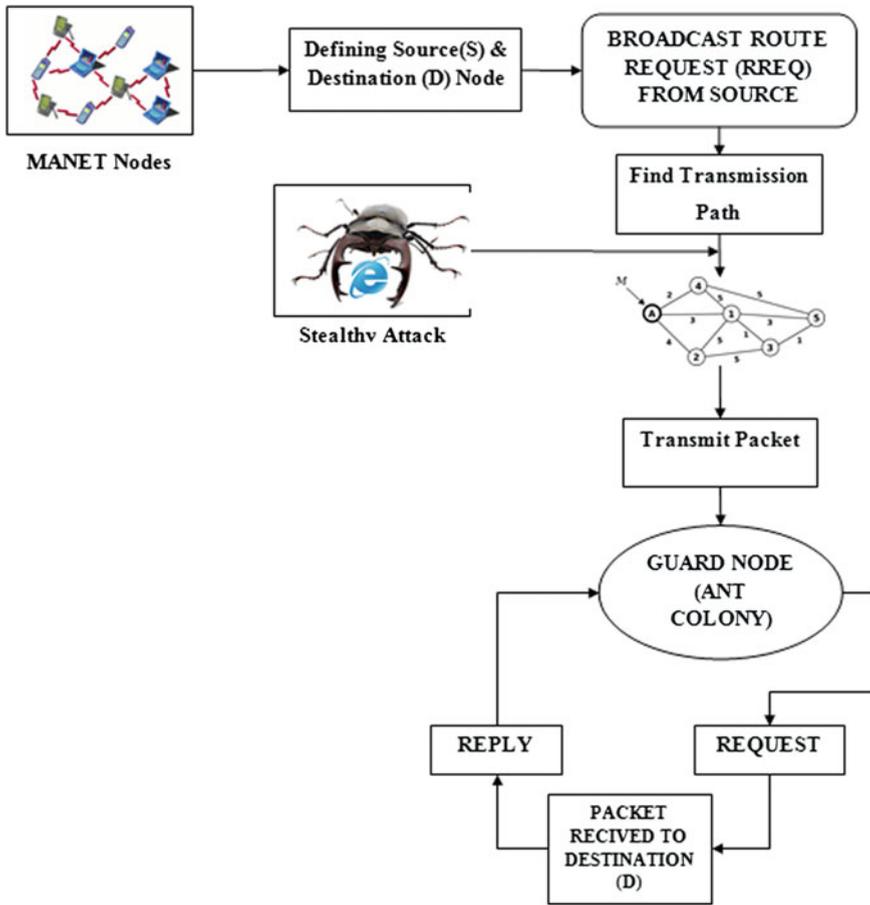


Fig. 1 Proposed architecture

their progression. Pheromone quality is the founded route quality function and also it is the impulsive material. An ant makes change in quantity of pheromone in route (m, n) traveling from the node m to node n as follow:

$$\tau_{p, q} = \sigma \cdot \tau_{p, q} + \Delta\tau_{p, q} \tag{1}$$

where σ is factor of pheromone desertion. Its value must be less than 1 to evade premature convergence and pheromone addition. At point m , ant choose point as n (i.e., to pursue the path (m, n)) as of subsequent probability:

$$P_{p, q} = \frac{(\tau_{p, q})^\alpha \cdot (\eta_{p, q})^\beta}{\sum_{(p, o) \in c} (\tau_{p, o})^\alpha \cdot (\eta_{p, o})^\beta} \tag{2}$$

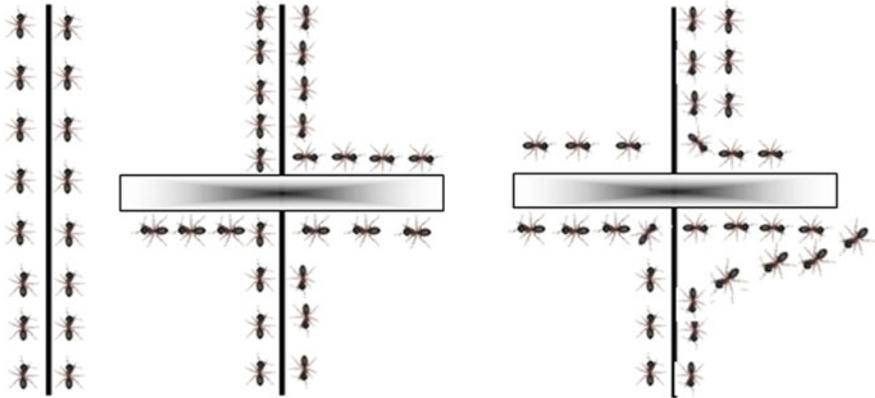


Fig. 2 ANT behavior of searching shortest path

where,

- $\tau p, q$ is pheromone power on the path (m, n)
- $\eta p, q$ is visibility field of ant's on path (m, n) (ant thinks there is foodstuff at end of the route).
- \acute{a} and \hat{a} are parameter that controls comparative pheromone importance intensity evaluated with visibility field of ant.
- C it signifies set of all probable paths initial from the point $I ((p, o)$ is path of C) (Fig. 2).

The characteristics of Ant colony optimization algorithm for routing. The set of core property characterizes the Ant colony instances for the problems in routing:

- (1) Gives multipath routing and traffic-adaptive.
- (2) Depends on both active and passive information gathering and monitoring.
- (3) Using stochastic components.
- (4) Do not allow local estimate to have a global impact.
- (5) Set paths in a less egoistic way than in the pure shortest path favors load balancing.
- (6) Shows less sensitivity to the parameter setting [6].

3.3.2 Path Discovery Phase

STEP I: Suppose source node S is containing some information to throw at the destination D with higher demands of QoS transmission rate, more bandwidth and less delay. A various list of nodes which progressively visit by the ant is known as visited node list. This list makes route R as of source to destination node.

STEP II: Firstly it chooses the source node S . The visited node list is initialized to the (S) source node.

Step III: The source S will initiate a $path_request_ant$ to the destination D by its entire neighbor that is at distance of 1 hop from S . The $path_request_ant$ is having source address, bandwidth, hop count, and destination address.

STEP IV: After this pheromone evaporation of every 1-hop distance nodes are calculated.

Each node (i) preserves a table called “PMtab” pheromone table which identify the quality of pheromone available at each link (Vm, Vn). This quantity is initialized with constant C .

STEP V: After that we calculate pheromone evaporation of every 2-hop node distance.

STEP VI: lastly we will calculate the value of path preference probability of every route from source S by the help of pheromone evaporation for every node. j node from the set of adjacent nodes (n, o, \dots, z) for I is taken as MPR node so that it will cover every 2-hop distance node and thir probability of path preference is better as others.

STEP VII: when the calculated probability of path preference value is better than the required, path will be accepted and stored inside the memory.

STEP VIII: If $path_request_ant$ reach at the destination, it will be changed as $path_reply_ant$ and delivered toward original source. The $path_reply_ant$ will be taken as of same path as of equivalent to $path_request_ant$ but its direction will be reversed.

STEP IX: The route having the greater path preference probability is considered as best route and the transmission occurred at that route.

Algorithm 1: Path Discovery

1. $tw = twMax$
2. While $tw \geq twMin$ do
3. set $A \in As, I(tw), \dots, As, t(tw)$;
4. $io = 0$;
5. While $io \leq L$ and $io \leq imax$ do
6. Path = 0;
7. $io = io+1$;
8. $Plower = 1 - 1 - \sqrt[m]{1 - pu}$;
9. For $\forall neighbor j \in set A$ do
10. If $As, n(tw) \geq Plower$ then
11. Path = path + 1;
12. End if
13. End for
14. If path $\geq io+1$;
15. $\Pi_{end-to-end} = \{(z1, v2, \dots, (zz, vz))\}$;
 {where list is arranged according to local consistency rate or ant colony optimizations problems routing table depends whether it works on Ant colony or

local mode, v1 is neighbor and n1 is neighbor ID corresponding local value or Ant colony optimization probability routing table}

16. rreq = new RREQ (*mo*, *Plower*, *tw*);
17. n = number of Neighbors of source node;
18. numRREQMsg = 0;
19. for I = 1 to n do
20. If $As(tw) \geq Plower$ then
21. Send (RREQ, $\prod_{i \in \text{endtoend}[i].neighbor}$);
22. numRREQMsg++;
23. if num RREQMsg = *mo* then
24. return success;
25. end if
26. end if
27. end for
28. end if
29. end while
30. $tw = 0.9 \times tw$
31. end while
32. return error

Latest *mmax* or total no of neighbors (*L*) is the procedure required to decrease *tw* because of the reality that links could be much more reliable at the shortest time window (line 30). Though, it's favored not to contain small time windows, as source need to throw route to ensure communication to authenticate if end to end reliability still can satisfy *pu* or not. Confirmation invite additional overhead particularly when it is taken at diminutive interval occasionally so, a least threshold is taken for *tw* called *tw*, in (here 20 s), when the line algorithm fail intended for least threshold then its breakdown for request (line 30). or else, source node throw RREQ message to *mo* at maximum reliable relations which having the highest rate in Ant produced table for request and destination but in confined mode there highest consistent links are calculated based on the local data collected by direct communication with its neighbors (line 14–28). Based Upon the communication received on RREQ, every intermediate node runs and handle RREQ message algorithm. The Intermediate nodes firstly checks either it is destination or not. If it's not then handling RREQ Message algorithm try RREQ to forward toward the destination. The intermediary node forwards only *mo* example of RREQ message and remove other example. This is achieved through bookkeeping variable number of forwarded message (line 1–2).

Algorithm 2: to handle RREQ Message (RREQ)

1. When num of Forwarded Message [RREQ] > *mo* then
2. Return;
3. End if
4. $\prod_{acc} \times A_j, i(tw)$;

5. $Lt(tw) = \frac{P_{lower}}{\prod_{acc}}$;
6. n = number of neighbors of intermediate node.
7. $P_{end-to-end} = \{(n1, v1), \dots, (n_n, v_n)\}$;
 {Where set is arranged as for local consistency value or Ant colony optimization probability table depends whether it working on ant colony optimization or local mode}, v_i is neighbor equivalent local reliability value or $n1$ is neighbor ID or Ant colony optimizations probability routing table}
8. For o = 1 to z do;
9. If $P_{end-to-end} [O]$ neighbor \in path Vector then
10. Continue;
11. End if
12. If $A_j, o(tw) \geq Lo(tw)$ then
13. NumOfForwardMsg [RREQ]++;
14. Forward (RREQ, $P_{end-to-end} [O]$ neighbor);
15. Return;
16. End if
17. end for

4 Result and Discussion

The measurement attribute for performing our proposed approach, we conducted some of the experiment and extract the dataset which was done on the following configuration and our work implementation will be with following configuration: (1) Windows 7 (2) Intel Pentium (3) CPU G2020 and (4) Processer speed 2.90 GHz.

4.1 Attack Identification

In Table 1, we are presenting the consequence of our research showing the attack detection and a comparison our proposed system with the several existing system. The accurate attack detection is entirely fulfilled by proposed work.

4.2 Path Discovery Ratio

The result in Fig. 3 shows the accurate performance graph and successful packet delivery through the shortest path which is more accurate than the existing approaches.

Table 1 Attack identification

Techniques	Attack				
	Sinkhole attack	Wormhole attack	HELLO flood	Sybil attacks	Spoofing
AODV protocol	✓	✓	•	•	✓
EMPDSR protocol	•	✓	✓	✓	✓
MPDSR protocol	✓	✓	•	✓	✓
Fuzzy logic	✓	✓	•	✓	•
Ant colony optimization	✓	✓	✓	✓	✓

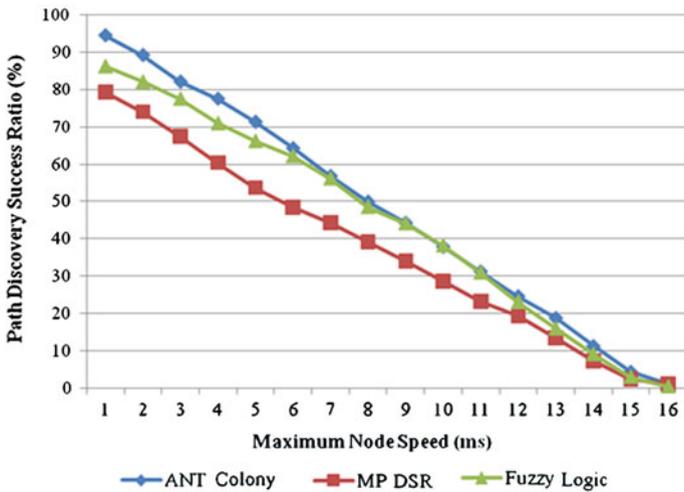


Fig. 3 Path discovery ratios

4.3 Packet Delivery Ratio

The shown result in Fig. 4 gives the shortest path techniques which are helpful in packet delivery than the other existing techniques used for MANET.

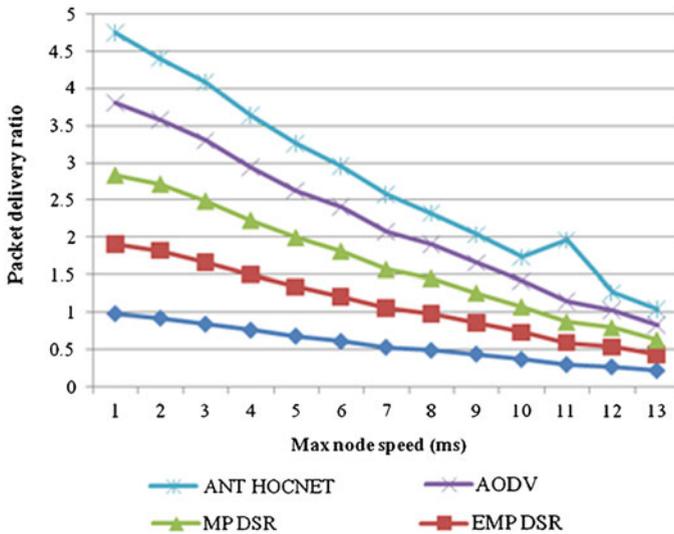


Fig. 4 Packet delivery ratios

5 Conclusion

Here we conclude that in this paper, we briefly introduced the concept of MANET and their vulnerabilities. Then we discussed about various security attacks that can happen in MANETs. We mainly analyzed the problem of Stealthy attacks that are happening in MANET and used Guard Node-based approach to overcome the above mentioned attack. In Guard Node approach ant colony-based optimization algorithm is integrated in order to provide a better solution than different existing approaches. We have taken results from our proposed solution using the NS-2.35 simulator and compared the performance in conditions of packet delay and packet delivery ratio. The simulation consequences illustrate that our projected method is better in detecting the Stealthy attack in MANET. The simulation consequences explain that our projected method is better in terms of packet delay and delivery ratio when compared with existing methods. Our system not only detects the Stealthy attack but also it isolates the attacker from the network.

References

1. Sheela, D., Srividhya, V.R., Asma, B., Anjali, Chidanand, G.M.: Detecting black hole attacks in wireless sensor networks using mobile agent. In: International Conference on Artificial Intelligence and Embedded Systems, Singapore (2012)
2. Kuthadi, V.M., Rajendra, C., Selvaraj, R.: A study of security challenges in wireless sensor networks. JATIT **20**(1), 39–44 (2010)

3. Lee, S.J., Gerla, M.: Split multipath routing with maximally disjoint paths in ad hoc networks. In: IEEE International Conference on Communications, pp. 3201–3205 (2001)
4. Carbanar, B., Ioanisis, I., Cristina., N.: JANUS: towards robust and malicious resilient routing in hybrid wireless networks. In: Proceedings of the 3rd ACM Workshop on Wireless Security, pp. 11–20. NY (2004)
5. Khalil, S.B., Nina, R., Shroff, N.: UNMASK: utilizing neighbor monitoring for attack mitigation in multihop wireless sensor networks. *Ad Hoc Netw. J.* **8**(2), 148–164 (2010)
6. Huang, Y., and Lee, W.: A cooperative intrusion detection system for Ad Hoc networks. In: Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks, pp. 135–147 (2003)
7. Issa, H.: LITEWORP: A lightweight countermeasure for the wormhole attack in multihop wireless networks. In: Proceedings of the International conference on Dependable Systems and Networks. Washington, pp. 612–621 (2005)
8. Kuthadi, V.M., Selvaraj, R., Marwala, T.: An efficient web services framework for secure data collection in wireless sensor network. *Br. J. Sci.* **12**(1), 18–31 (2015)
9. Baruch, A., Reza, C., David, H., Cristina, N., Herbet, R.: ODSBR: An on-demand secure Byzantine resilient routing protocol for wireless ad hoc networks. *ACM Trans. Inf. Syst. Secur.* **10**(4), 1–35 (2008)
10. Selvaraj, R., Kuthadi, V.M., Marwala, T.: Ant-based distributed denial of service detection technique using roaming virtual honeypots. *IET Commun.* **10**(8), 929–935 (2016)

Inset Fed Circular Microstrip Antenna with Defected Ground

Nikhil Kumar Singh, Niraj Sharma, Zakir Ali, Vinod Kumar Singh and Akash Kumar Bhoi

Abstract In this paper, a novel design of defected ground circular patch antenna using glass epoxy materials has been presented. The substrate of the intended antenna was prepared of glass epoxy, whereas patch and ground plane are made out of copper. Enhancement in the geometry is done by increasing the electrical length and volume without growing the size of the patch antenna. The simulations were accomplished using Computer Simulation Technology software and the association of measured and simulated results of anticipated antenna has been studied. Experimental and simulated outcome are in outstanding conformity.

Keywords Partial ground · WLAN · Textile · CST tool

1 Introduction

Presently, the quick improvement of modern communication systems is necessary for transportable devices for some important features which includes easy designing, light weight, small in size, compatible with microwave, millimeter wave integrated circuits, less production cost, and easy fabrication of microstrip antennas.

N.K. Singh · V.K. Singh (✉)
Department of Electrical Engineering, S R Group of Institutions, Jhansi, UP, India
e-mail: singhvino34@gmail.com

N.K. Singh
e-mail: nitinm18@gmail.com

N. Sharma
Department of Physics, S R Group of Institutions, Jhansi, UP, India

Z. Ali
Department of Electronics Engineering, I.E.T.B.U., Jhansi, UP, India

A.K. Bhoi
Department of EEE, Research & Development Section, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, India
e-mail: akash730@gmail.com

The microstrip antenna has abundant useful properties which includes tiny size, low-cost of the fabrication, light weight, ease of setting up but the main limitations of printed antennas remains their narrow bandwidth features which limits the range of frequency over which the antenna can work efficiently. Microstrip antenna comprises three most important parts which is substrate, patch, and ground. A dielectric substrate is sandwiched between radiating patch and ground plane. The conducting patch is located on the dielectric substrate which is used as radiating element. On other side of dielectric substrate there is conducting layer used as ground part [1–6]. Microstrip antenna plays major role in wireless communication system. Microstrip antennas are used in high performance aircrafts, radar, missiles, and other spacecraft. It has many advantages such as its light weight, simple structure, ease of addition and less cost. Microstrip antenna requires very less space for installation as these are simple and small in size. The only space these require is the space for the feed line which is placed behind the ground plane [7–14]. There are a variety of methods for enhancing the bandwidth of textile patch antenna such as enlarge the substrate thickness, utilizing low dielectric material, using numerous feeding techniques and impedance matching. But the thickness of the antenna and the bandwidth is both conflicting property that is enhancement of bandwidth increases the size and thickness of presented antenna [15–22].

2 Antenna Design Configuration

From Eq. (1) the dimension of circular patch were calculated to get the propose design. Where f_r is the resonance frequency in GHz, R is the radius of the circular patch antenna in mm, and ϵ_r is the relative permittivity of dielectric material. The foundation of the proposed antenna was pentagonal slot with patch radius 15 mm (Table 1).

$$R = \frac{87.94}{f_r \sqrt{\epsilon_r}} \quad (1)$$

Table 1 Dimensions of anticipated textile antenna

Design parameters	Value
Relative permittivity (ϵ_r)	4.40
Substrate thickness (h)	1.6 mm
Outer circle patch radius (R)	15 mm
Pentagonal slot dimensions	5 mm
Microstrip feed line ($L_f \times W_f$)	2×32
Substrate dimension ($L_s \times W_s$)	60×60
Partial ground plane ($L_g \times W_g$)	50×27

3 Fabrication of Antenna and Results

The dimension of ground is of 60 mm × 60 mm and is made of copper with a thickness of 0.038 mm. The simulations were accomplished using CST Microwave Studio. Figure 1 describes CST Model of proposed antenna. A 50 Ω microstrip line feed of dimension 32 mm × 2 mm was provided for the proposed antenna feed [15–21]. The proposed antenna is a hardware implementation of a prior published article reported in [1]. Furthermore, Fig. 2 demonstrates the geometry of proposed antenna prototype mounted on the surface of glass epoxy.

The simulation is performed using Computer Simulation Technology (CST) software whereas fabrication and measurements are performed in the Indian Institute of Technology Kanpur microwave lab. The chief apprehension of this article is to improve bandwidth, gain, and return loss. Figure 3 shows the comparison between experimental and simulated reflection coefficient of proposed antenna. A network analyzer is used to determine the reflection coefficient (S_{11}) results of the fabricated antenna. The first band (1.84–2.33 GHz) of designed antenna is suitable to work for Ultra Mobile Telecommunication application and second band (3.36–4.59 GHz) of proposed antenna for Missile and Gunfire-control Radar (3.40–3.65 GHz) applications (Fig. 4).

Fig. 1 Design models on CST **a** top view of antenna **b** back view of antenna

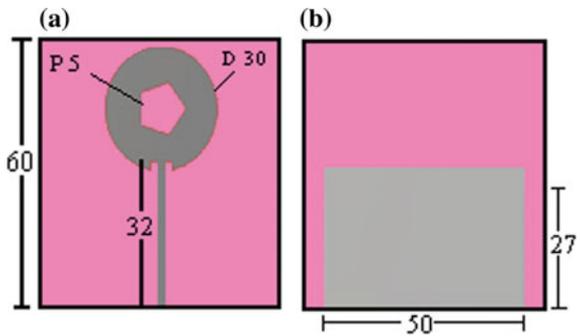
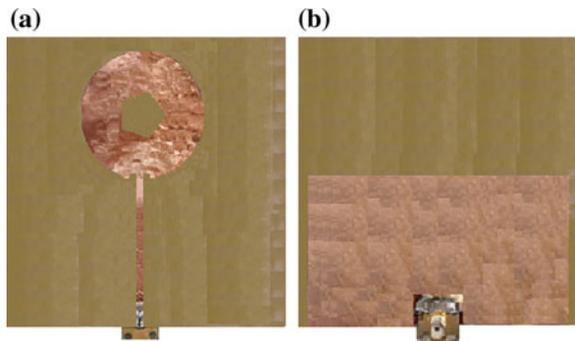


Fig. 2 Prototype microstrip antenna **a** top view of antenna **b** back view of antenna



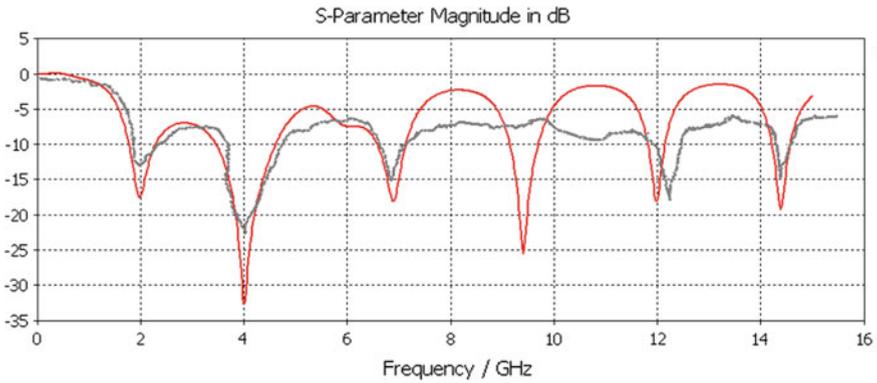


Fig. 3 Comparison between measured and simulated return loss of proposed antenna

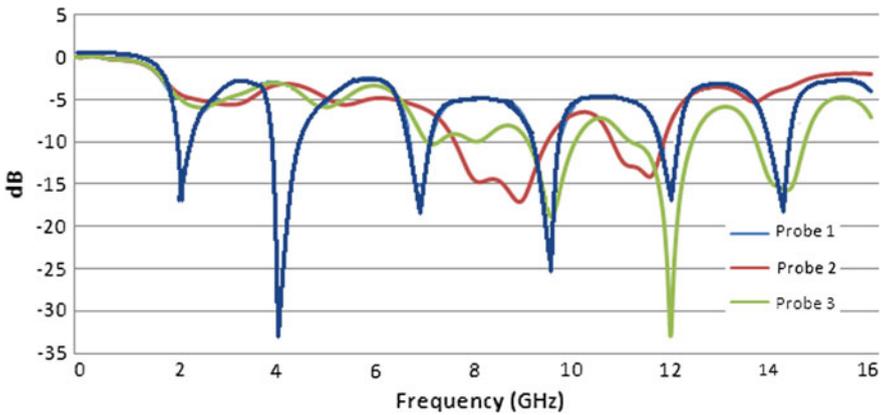


Fig. 4 Return loss versus frequency of proposed antenna at different probes

Figure 5 describes 3-Dimensional radiation pattern of anticipated antenna at 2.0135, 3.991, 6.9393, and 11.991 GHz. The variation of frequencies versus directivity of proposed antenna were demonstrated in Table 2. The maximum directivity achieved was about 6.115 dB at resonant frequency 6.9393 GHz. Our main concern is to improve gain and directivity.

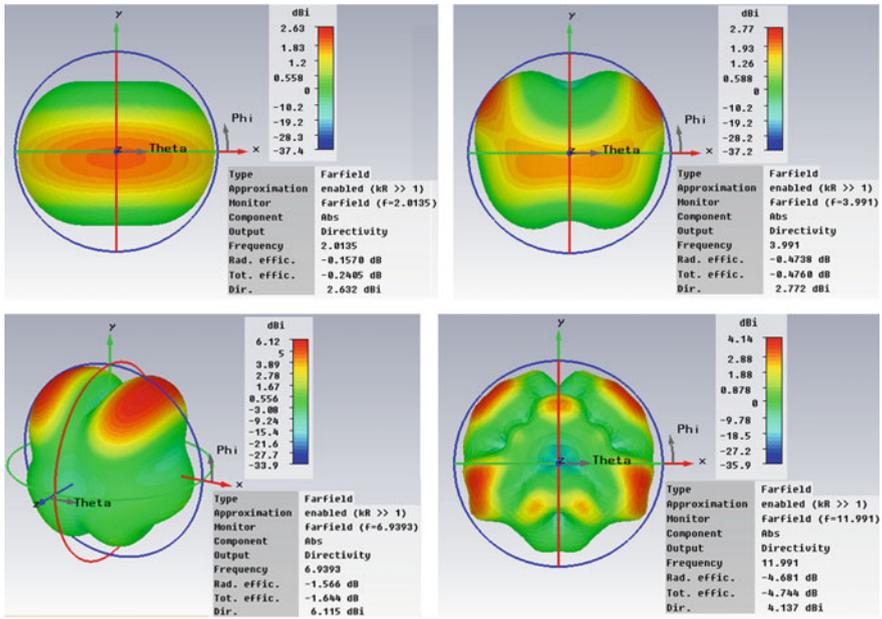


Fig. 5 Radiation pattern of proposed design at 2.0135, 3.991, 6.9393, and 11.991 GHz

Table 2 Measured bandwidth of proposed antenna

Bands	Resonant frequency (GHz)	Directivity (dB)	Bandwidth (%)	Frequency range (GHz)
First	2.0135	2.632	21.90	1.84–2.33
Second	3.9910	2.772	30.94	3.36–4.59
Third	6.9393	6.115	06.71	6.62–7.08
Fourth	12.123	4.137	03.00	11.80–12.16

4 Conclusions

The proposed microstrip antenna has been composed with defected ground plane and customized circular patch with glass epoxy substrate which is appropriate for Ultra Mobile Telecommunication System application. The proposed defected ground microstrip antenna was fabricated and tested. The simulated results show that the presented antenna has wide bandwidth and high gain. In addition of this the compact size of the patch antenna further confirms its appropriateness for transportable devices. The measured results have been compared with the simulated results using commercial software CST. The measured and simulated results in terms of reflection coefficient show good conformity.

References

1. Raghupatruni, V., Krishna, R., Kumar, R.: Design of temple shape slot antenna for ultra wideband applications. *Prog. Electromagn. Res. B* **47**, 405–421 (2013)
2. Balanis, C.A.: *Antenna Theory: Analysis and Design*. Wiley, New York (2004)
3. Wong, K.-L.: *Compact and Broadband Microstrip Antennas*. Wiley. ISBN: 0-471-41717-3 (Hardback) (2002)
4. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: A wide band compact microstrip antenna for GPS/DCS/PCS/WLAN applications. *Intell. Comput. Netw. Inf. (Springer)* **243**, 1107–1113 (2014). ISBN: 978-81-322-1664-3
5. Srivastava, R., Singh, V.K., Ayub, S.: Comparative analysis and bandwidth enhancement with direct coupled C slotted microstrip antenna for dual wide band applications. *Front. Intell. Comput. Theor. Appl. (Springer)* **328**, 449–455 (2015). ISBN: 978-3-319-12011-9
6. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: Bandwidth optimization of compact microstrip antenna for PCS/DCS/Bluetooth application. *Central Eur. J. Eng. (Springer)* **4**(3), 281–286 (2014)
7. Singh, N., Singh, A.K., Singh, V.K.: Design & performance of wearable ultra wide band textile antenna for medical applications. *Microwave Opt. Technol. Lett.* **57**(7), 1553–1557 (2015)
8. Srivastava, S., Singh, V.K., Ali, Z., Singh, A.K.: Duo triangle shaped microstrip patch antenna analysis for WiMAX lower band application. In: *International Conference on Computational Intelligence: Modelling Techniques and Applications (CIMTA-2013)*, Procedia Technology Elsevier, vol. 10, pp. 554–563 (2013)
9. Singh, N.K., Singh, V.K., Naresh, B.: Textile antenna for microwave wireless power transmission. In: *International Conference on Computational Modelling and Security (CMS 2016)*, Procedia Computer Science, vol. 85, pp. 856–861 (2016)
10. Gupta, N., Singh, V.K., Ali, Z., Ahirwar, J.: Stacked textile antenna for multi band application using foam substrate. In: *International Conference on Computational Modelling and Security (CMS 2016)*, Procedia Computer Science, vol. 85, pp. 871–877 (2016)
11. Rawat, A.K., Singh, V.K., Ayub, S.: Compact wide band microstrip antenna for GPS/WLAN/WiMax applications. *Int. J. Emerg. Trends Eng. Dev.* **7**(2), 140–145 (2012)
12. Baudh, R.K., Kumar, R., Singh, V.K.: Arrow shape microstrip patch antenna for WiMax application. *J. Env. Sci. Comput. Sci. Eng. Technol.* **3**(1), 269–274 (2013)
13. Loni, J., Ayub, S., Singh, V.K.: Performance analysis of microstrip patch antenna by varying slot size for UMTS application. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 01–05 (2014)
14. Srivastava, R., Ayub, S., Singh, V.K.: Dual band rectangular and circular slot loaded microstrip antenna for WLAN/GPS/WiMax applications. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 45–48 (2014)
15. Dhupkariya, S., Singh, V.K.: Textile antenna for C-band satellite communication application. *J. Telecommun. Swit. Syst. Netw.* **2**(2), 20–25 (2015)
16. Loni, J., Singh, V.K.: Development of bandwidth enhanced microstrip patch antenna for UMTS application. *J. Microwave Eng. Technol.* **2**(1), 01–07 (2015)
17. Singh, V.K., Naresh, B.: Multi resonant microstrip antenna with partial ground for radar application. *J. Telecommun. Swit. Syst. Netw.* **2**(1), 01–05 (2015)
18. Singh, M., Singh, V.K., Naresh, B.: Rectangular slot loaded circular patch antenna for WLAN application. *J. Telecommun. Swit. Syst. Netw.* **2**(1), 07–10 (2015)
19. Ali, Z., Singh, V.K., Kumar, A., Shahanaz, A.: E shaped microstrip antenna on rogers substrate for WLAN applications. *Proc. IEEE*, 342–345 (2011)
20. Din, N.M., Chakrabarty, C.K., Bin Ismail, A., Devi, K.K.A., Chen, W.Y.: Design of RF energy harvesting system for energizing low power devices. *Prog. Electromagn. Res.* **132**, 49–69 (2012)

21. Osman, M.A.R., Rahim, M.K.A., Azfar, M., Samsuri, N.A., Zubir, F., Kamardin, K.: Design, implementation and performance of ultra-wideband textile antenna. *Prog. Electromagn. Res. B* **27**, 307–325 (2011)
22. Singh, V.K., Singh, N.K.: Compact circular slotted microstrip antenna for wireless communication systems. *J. Microw. Eng. Technol.* **1**(1), 07–14 (2015)

Analysis of Thermal Comfort and Visual Comfort—A Soft Computing Approach

Sandhyalaxmi G. Navada, Chandrashekara S. Adiga
and Savitha G. Kini

Abstract The entry of daylight into the interiors depends on the orientation of windows, size of the windows, time of the day, as well as weather conditions. The position of the blinds also adds to the above factors to allow day light inside. The illuminance inside a room varies throughout the day depending on the above factors. The introduction of daylight inside the room also aids to increase in temperature inside the room, which hinders the thermal comfort of the user inside the room. Thermal comfort can be achieved by installing Air Conditioner (AC) inside the room. But due to the excessive heat inside, the energy consumption increases to maintain the thermal comfort. This paper deals with the various measures taken to reduce the energy consumption. These values are then fed into a fuzzy logic controller to get the optimum result.

Keywords Daylight · Thermal comfort · Visual comfort

1 Introduction

Daylighting is considered as the key element in reducing the consumption of energy. This can be implemented by comfortably using daylighting autonomy. In order to allow adequate daylight into the building, proper placement of windows is very much necessary. The positive effects of daylight are increase in productivity and decrease in the number of absentees. Integration between electric lighting and daylighting becomes a critical part of every building. Of all the energy that is used, nearly one third is consumed by the building sector. In large cities 80% of carbon emission comes from energy used to heat, cool and power buildings. Over the next few decades the amount of energy used by the building sector will increase dramatically. Hence improving the energy efficiency of buildings is vital for

S.G. Navada (✉) · C.S. Adiga · S.G. Kini
Department of Electrical & Electronics Engineering, Manipal Institute of Technology,
Manipal University, Manipal, India
e-mail: sandhya.girish@manipal.edu

greenhouse gas emissions. The amount of heat entry into the building depends mainly on, what the building envelope is constructed with and how it is constructed. The penetration of sunlight into the interiors will cause visual as well as thermal discomfort. Intelligent lighting and shade control solutions are an effective way to increase the well-being, to achieve optimal visual comfort, to enhance the experience and to save energy. By integrating lighting and shading controls we will be able to achieve lower lighting cost as well as lower heating, ventilation, and cooling cost [1].

The energy needs of a building is greatly influenced by the location of the building, orientation of windows, time of the day as well as the season of the year. The day to day well-being of a user can be improved by having a thermally comfortable environment. Thermal comfort is achieved when perfect balance between body temperature and surrounding environment is established. The well-being of a user is also effected by visually comfortable environment, which can be achieved by providing appropriate light inside the interiors and avoiding glare. Energy can be saved by balancing the natural and artificial light inside a room. This can be achieved by using shading devices and automating them [2].

Excessive solar gain inside a building can be reduced by effectively using shading devices. One of the reasons for solar heat gain is the window element. Shading on windows will reduce the solar heat gain and reduce energy consumption. The overheating of a building can be reduced by direct radiation from the sun, which can be achieved by the use of shading devices [3].

The amount of heat entering a building can be reduced to a larger amount, by implementing proper shading devices. This helps in substituting electric light with daylight and hence reduces energy consumption. In order to achieve this, automation is required, which will integrate daylight into the building along with dimming of electric light [4].

The comfort of the user inside a building as well as the energy demand are greatly influenced by the orientation of windows. By employing automated shading devices, the reduction in energy consumption can be achieved, along with visual and thermal comfort [5].

The energy consumption inside a building due to AC can be reduced by changing the set point temperature. But again doing this very frequently may have adverse effect on the comfort of the user [6].

In order to reduce energy consumption, it is advisable to place the windows appropriately during the construction phase of the building. The energy usage can be minimized by integrating daylight with artificial light. This has an adverse effect of introducing heat into the interior. Thus an efficient lighting system can be designed by taking into consideration all the above factors [7].

Reduction in energy consumption of a building as well as occupant satisfaction are the goals to develop an intelligent building façade. To obtain occupant satisfaction, the window blinds need to be automated. Adaptive fuzzy controller has been developed to determine the window blind position, based on solar radiation, visual, and thermal comfort [8].

The two discomfort indexes as defined by ASHRAE standard are Predicted Mean Vote (PMV) and Percentage People Dissatisfied (PPD). The factors affecting P.M.V are; clothing insulation, air temperature, metabolic rate, activity, average occupancy, mean radiant temperature and relative humidity and the range of this index is from -3 to $+3$. According to ASHRAE standard, the comfort zone with respect to the P.M.V model lies between -0.5 and $+0.5$ [9].

2 Methodology

By taking into consideration, entry of daylight into the building, an adaptive controller is proposed. This controller also takes into account the thermal as well as visual comfort of a user inside the office. By controlling the blinds, the required amount of daylight is allowed inside, so as not to overload the air conditioner. The Energy saving hence achieved is noted.

The block diagram of the proposed work is as shown in Fig. 1.

A Temperature sensor is used to sense the available temperature in the room. This is fed into the Fuzzy Logic Controller. The set point value for the temperature is also fed into the Fuzzy Logic Controller. The blinds are opened to allow optimum amount of daylight, so as to maintain the temperature inside the room very near to the set point temperature. This in turn will reduce the excessive load on the air conditioner. A lux meter is used to detect the prevailing light level, inside the room. Lux meter is used to integrate an electric lighting system with a daylighting system, so lights operate only when daylight is insufficient.

The amount of daylight available for the present blind position is determined from the lux meter. The available daylight is compared with the required illumination level as fed into the Fuzzy Logic controller. If the available daylight is less than the required lux level, then the artificial light has to be used. The desired blind position for the fed set point temperature and the prevailing temperature in the room

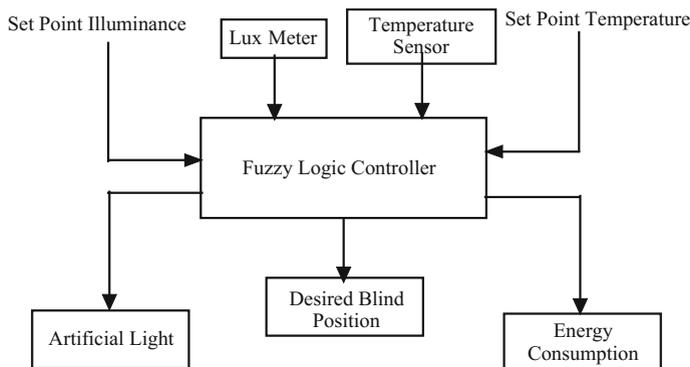


Fig. 1 Block diagram of the proposed work

is calculated by the Fuzzy Logic Controller. The Fuzzy Logic Controller also calculates the Energy Consumed.

The dimension of model room considered are 3.75 m × 3.75 m × 2.35 m. The room consists of two opposite windows each facing east and west side, respectively. The room is divided into nine equal grids to measure the illuminance. Measurement of internal illuminance and temperature for different blind position at a work bench of height 0.8 m in the model room with/without air-conditioning system is carried out. These values were taken at every half hour from 9 am to 5 pm. The different blind positions are 0°–90° (fully closed) in increment of 30°.

Air-conditioning system was installed in the model room in order to carry out thermal analysis. Energy meter readings were noted down at every half an hour interval from 9:00 am to 5:00 pm at different blind position ranging from 0° to 90°. The air-conditioning system was set at three different set point temperatures, i.e., 22°, 24°, and 26°. From the data obtained, the optimum blind position for minimum energy consumption was analyzed.

2.1 Analysis Without AC

The illuminance readings are recorded throughout the day from 9 am to 5 pm for every half an hour. The weighted average of the recorded readings at the 9 points is then calculated. Figures 2, 3, 4, 5, 6 and 7 show the graph of illuminance versus time of the day for various blind positions.

It is observed that for certain time of the day, for particular blind position the illuminance falls below 500 lx (for writing, reading, typing, data processing the minimum illuminance required is 500 lx according to Chartered Institute of Building Services Engineers, code for interior lighting). In order to maintain 500 lx inside the room, the artificial light has to be used and dimmed accordingly. Thus 2 LED luminaire were installed in the model room in order to maintain 500 lx.

Fig. 2 Illuminance versus time with no blinds

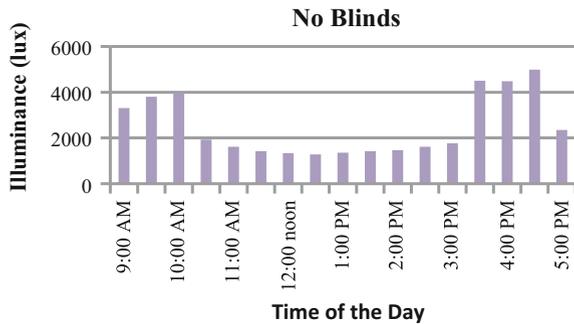


Fig. 3 Illuminance versus time with 30° blinds

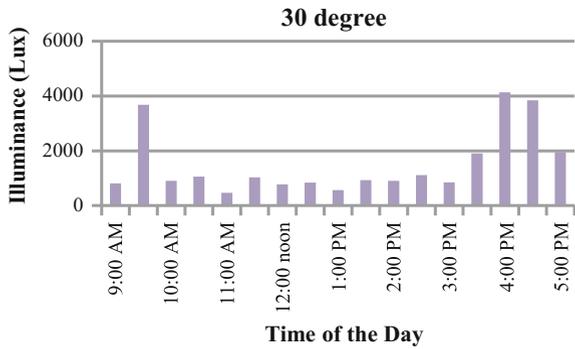


Fig. 4 Illuminance versus time with 60° blinds

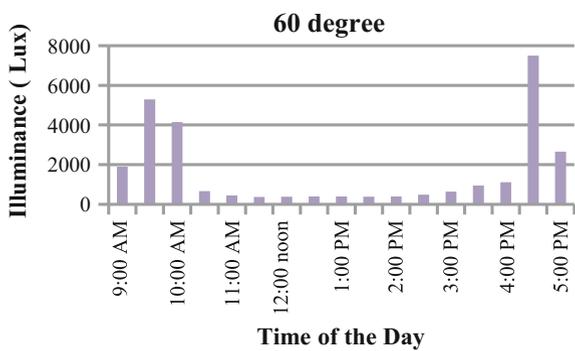
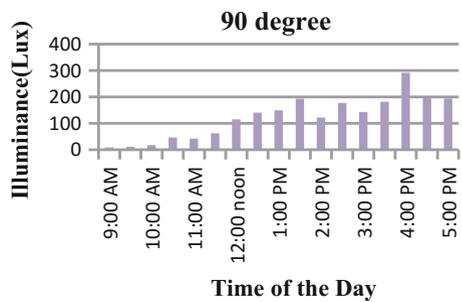


Fig. 5 Illuminance versus time with 90° blinds



2.2 Analysis with AC

The analysis is then done by connecting AC. The AC is set to different set point temperature. The energy meter readings are taken separately for lighting as well as AC load. The readings are tabulated in Tables 1, 2 and 3. The total energy consumption including both AC as well as lighting load is also calculated.

Table 1 Energy consumption at 22 °C

AC temperature -22 °C				
Blind position (wrt vertical)	AC (kWh)	Artificial light (kWh)	Used (h)	Total (kWh)
0°	17.1	0	0	17.1
15°	16.8	0	0	16.8
30°	16.2	0	0	16.2
45°	15.7	0	0	15.7
60°	15.5	0.17	1.5	15.67
75°	15.35	0.581	3.5	15.931

Table 2 Energy consumption at 24 °C

AC temperature -24 °C				
Blind position (wrt vertical)	AC (kWh)	Aritificial light (kWh)	Used (h)	Total (kWh)
0°	16.8	0	0	16.8
15°	16.6	0	0	16.6
30°	16	0	0	16
45°	15.4	0	0	15.4
60°	15.25	0.21	2	15.46
75°	15.2	0.487	4	15.687

Table 3 Energy consumption at 26 °C

AC set point temperature -26°				
Blind position (wrt vertical)	AC (kWh)	Aritificial light (kWh)	Used (h)	Total (kWh)
0°	16.6	0	0	16.6
15°	16.53	0	0	16.53
30°	15.87	0	0	15.87
45°	15	0	0	15
60°	14.8	0.22	2	15.02
75°	14.67	0.489	4	15.159

It is observed that energy consumption including both artificial light and air conditioning are minimum at 45° and 60° blind positions. And for air conditioning maximum energy consumption is at 22 °C and minimum energy consumption is at 26 °C. For thermal comfort the parameter which is of paramount importance is temperature. To measure temperature LM—35 temperature sensors are used. To measure humidity, a humidity sensor DH—22 is used. According to ASHRAE standard, the ambient temperature of 24 °C is the most ideal temperature amongst those specified in the comfort range [9]. It is also found that the PMV and PPD values fall within the specified comfort range for this value of temperature. The humidity also falls within the comfort range at 24 °C.

Table 4 Hourly energy consumption reading at a set point temperature of 24 °C

Hourly energy consumption reading (kWh) (24 °C)									
Time	Blind position (considering vertical as 0°)								
	45°			60°			75°		
	AC	Light	Total	AC	Light	Total	AC	Light	Total
9:00–10:00	1.88	0	1.88	1.78	0	1.78	1.7	0	1.7
10:00–11:00	1.98	0	1.98	1.62	0	1.62	1.6	0	1.59
11:00–12:00	1.82	0	1.82	1.57	0.11	1.68	1.5	0.12	1.64
12:00–13:00	2.31	0	2.31	2.21	0.12	2.33	2.2	0.13	2.29
13:00–14:00	2.52	0	2.52	2.42	0.12	2.44	2.4	0.17	2.57
14:00–15:00	2.34	0	2.34	2.2	0.11	2.31	2.1	0.23	2.33
15:00–16:00	2.21	0	2.21	2.01	0.11	2.21	2	0.23	2.23
16:00–17:00	2.1	0	2.1	1.86	0.11	1.97	1.9	0.11	2.01

Therefore, the hourly energy consumption for different blind positions are noted for this ideal temperature of 24 °C and are represented in Table 4.

From the hourly energy consumption reading taken at a set point temperature of 24 °C it is observed that the energy consumption varies according to the blind position. The hourly energy consumption is noted down for different blind position from 15° to 90° throughout the day. The readings for 45°, 60° and 75° is shown in Table 4. It is observed that the minimum energy consumption is at either 60° blind position or 75° blind position for different time of the day. Hence in order to obtain minimum energy consumption, the position of the blinds has to be changed during the entire day.

2.3 Development of Fuzzy Logic

From the thermal and visual analysis, the desired blind position and amount of energy consumption by both air conditioning and artificial light at different AC set temperature has been obtained. A soft computing approach is considered for further analysis. Illuminance value and AC set temperature are taken as input variable. The input variable illuminance has 2 membership functions low and high of type trimf, while temperature has 3 memberships function named 22, 24, 26 of type trimf.

Desired blind position, energy consumption by AC and use of artificial light are taken as output variable. 8 rules base are formed. The formulated rules along with the ranges for various outputs considered are fed into the fuzzy logic tool box in MATLAB.

Figure 6 shows that when illuminance is less than 500 lx, then there will be use of artificial light and desired blind position is 60° and with ac set temperature at

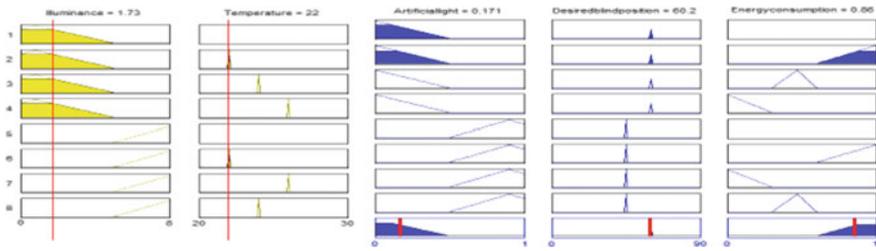


Fig. 6 Rule base when illuminance is less than 500 lx and temperature is at 22 °C

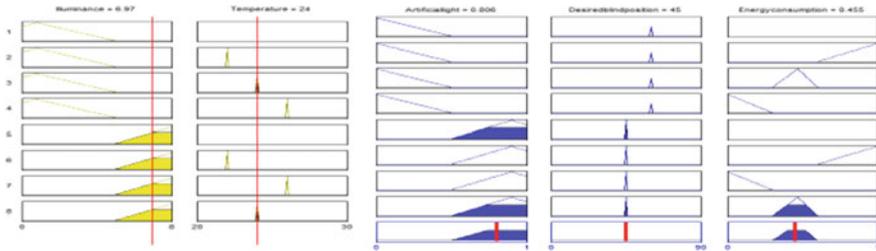


Fig. 7 Rule base when illuminance more than 500 lx and temperature is at 24 °C

22 °C, energy consumption is quite high. Figure 7 shows that when illuminance is more than 500 lx, then there will be no use of artificial light and blind position can be kept at 45° and with ac set temperature at 24 °C, energy consumption is less than that of 22 °C.

3 Conclusion

The analysis proved that daylighting is an effective option to reduce energy consumption. Window blinds can be used and adjusted effectively according to availability of daylight at particular time so that necessary amount of light can be utilized without causing visual discomfort and hence maintaining thermal comfort. From thermal and visual analysis it is inferred that with the change in the angle of blinds, energy consumption is changed. The optimum angle of blinds for minimum energy consumption is found out. It is concluded that the minimum energy consumption can be obtained at different time of the day by adjusting the blinds accordingly. The optimum temperature for thermal comfort to be achieved is found out to be 24 °C. The PMV and PPD value for this temperature is found to be within the comfort range. The Fuzzy Logic Controller gives the optimum value of blind position for fed set point temperature and the availing temperature inside the room. The amount of illuminance needed from artificial light is specified by the Fuzzy

Logic Controller. For the above conditions, the energy consumed is also calculated by the Fuzzy Logic Controller.

References

1. Kensek, K., Suk, J.Y.: Daylight factor (overcast sky) versus daylight availability (clear sky) in computer-based daylighting simulations. *J. Creative Sustain. Architect. Built Env. CSABE* **1** (2011)
2. Balancing Energy Efficiency with Comfort and Well-Being, Somfy Solutions. <https://www.somfypro.com/documents/531668/d7e3cb48-e7ba-434b-9649-29a4d0c3e262>
3. Ylitalo, H., Ip, K., Marshall, D.: Thermal performance of external roller blinds retrofit for offices in the United Kingdom. <https://enviroblinds.files.wordpress.com/2013/03/eppm2012-paper-thermal-performance-of-external-roller-blinds-retrofit-for-offices-in-the-united-kingdom.pdf>
4. Inkarojrit, V.: Balancing comfort: occupants' control of window blinds in private offices. A Dissertation Submitted in Partial Satisfaction of the Requirements for the Degree of Doctor of Philosophy. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.66.3256&rep=rep1&type=pdf>
5. Wienold, J.: Dynamic simulation of blind control strategies for visual comfort and energy balance analysis. In: *Proceedings: Building Simulation (2007)*
6. Gao, P.X., Keshav, S.: SPOT: A smart personalized office thermal control system, e-Energy'13, Berkeley, California, USA. Copyright 2013 ACM 978-1-4503-2052-8/13/05. 21–24 May 2013
7. Wu, G.: Studies in preliminary design of fenestration: balancing daylight harvesting and energy consumption. In: *PLEA2012—28th Conference, Opportunities, Limits & Needs Towards an Environmentally Responsible Architecture*, Lima, Perú 7–9 (2012)
8. Inkarojrit, V.: Multivariate predictive window blind control models for intelligent building façade systems. In: *Proceedings: Building Simulation (2007)*
9. Kim, J.-H., Min, Y.-K., Kim, B.: Is the PMV index an indicator of human thermal comfort sensation? *Int. J. Smart Home* **7**(1) (2013)

Spoken Dialog System in Bodo Language for Agro Services

Aniruddha Deka and Manoj Kumar Deka

Abstract In this work, a cost effective spoken dialog system and a dialog manager is developed for accessing information like price, weather, and fertilizer agricultural commodity in Bodo language. For the development of SR models, we have collected data from different dialect regions of BTC area, Assam using mobile phone where native language is Bodo. The target speakers are mainly farmers. Our proposed system consists of bluetooth enabled mobile for communication channel, Asterisk-based IVR module, a dialog manager, and a speech recognition module. Here, we discuss our system development, call flow design and performance evaluation of the speech recognition module. Our result indicates similar performances between 16 and 32 GMMs/state. The raw performance of the developed SDS system in terms of word recognition rate (WRR) is 83.29% in training and 79.12% in testing.

Keywords Spoken dialog system · Speech recognition · Asterisk Chain mobile · IVR

1 Introduction

There is over 80% of the population directly engaged in agricultural in BTC area of Assam where most the people are Bodo Tribe and cultivate different commodities. Although the Ministry of Agriculture, Government of India, designed some policies for the benefit of cultivators, but most of policies are internet-based. So farmers have to well verse literate to use those benefits. In BTC area of Assam we found that farmers are semi-literate and they are not able to use computer. Another

A. Deka (✉) · M.K. Deka
Department of Computer Science and Technology, Bodoland University,
Kokrajhar, Assam, India
e-mail: dekaaniruddha@gmail.com

M.K. Deka
e-mail: manoj.deka1@gmail.com

problem with the web-based system is that it is not native language. For this problem we try to investigate other solution through which we can solve this issue. We find some of the SMS, DTMF, IVR, and web-based solution but those are not cost effective and they are not designed in Bodo language. During survey we find that, the mobile penetration throughout the world is so high that the semi literature persons also can handle the mobile phone. From technical point of view it is easy to build voiced-based telephony architecture by using automatic speech recognition system with telephone network [1, 2]. In 2010 Saurabh Goel and Mahua Bhattacharya described a method for integration of telephone network with speech over Asterisk PBX server [3]. For a small organization or research point of view we proposed and design a Bluetooth-based IVR solution for our prototype model which is very cost effective and easy to maintain.

Bodo is a Sino Tibetan family language mainly spoken in Northern parts of Assam, the North Eastern state of India. It is also known as tonal language. There are 22 phonemes 16 consonant and 6 vowel phoneme [4]. The basis features of this language are agglutinating, juncture, and Intonation. Mostly used monosyllabic words are found in this language. Unrounded /w/ vowel is frequently used in this language. Bodo language used Devanagari script in writing and it is lesser rich in literacy. Due to this problem a very few work has been reported in speech recognition purpose. Few works reporting the recognition of limited set of vowels and digits in Bodo, there is no work reporting a moderate vocabulary Bodo SDS system. This work is the first of its kind in Bodo language targeting the practically deployable spoken dialog system. Also, to the best of our knowledge, there is no publicly available Bodo speech corpus in agricultural commodity large enough for training an HMM-based SDS system. The SDS system is developed using hidden Markov models (HMM) having triphone-based acoustic modeling and equiprobable word net covering a vocabulary size of 31 words. The raw performance of the developed SDS system in terms of word recognition rate (WRR) is 83.29%. The performance of the developed system is also evaluated using a measure which quantifies the efficiency of the call flow for the decimation of the information to the user.

2 Proposed Work

A spoken dialog system is computer agent that interacts with people by understanding spoken language. The input in spoken dialog systems consist of small set of spoken words like yes and no, digits such as 0–9 or combination of both [4]. In our work we mainly concentrate on building of dialog manger and evaluating the performance of speech recognizer engine. In our proposed system we take the input by making a call, recognize to word with our sphinx-based recognizer and then language model generate the word based on our recognizer output. Dialog manger

takes the output from the language analyzer module and passes it to the language generator. It also controls the structure of the dialog.

Our proposed system consists of Bluetooth enabled mobile phone, Asterisk server, speech interface, dialog manager and MySQL database. Asterisk is open source software on Linux/Unix platform that enables the computer to be used as a telephone network server [3]. It consists of computer telephone interface (CTI) card and IVR module. In our proposed system we have used Bluetooth connectivity in place of CTI card. The developed system mainly transmits the prerecorded voice responses and records and processes the voice query from the user. In our proposed architecture user first call to a mobile number which is connected to our asterisk server. When 2nd mobile get the call then its blue tooth adaptor connect the asterisk server, with the help of Chan mobile module. Chan mobile is asterisk module which helps to connect asterisk server with Bluetooth enabled mobile and it will help to transfer the communication. Chan mobile transfer the call to Dahdi, it is a collection of open source drivers, for Linux, which are used to interface with a variety of telephony related hardware. Here Dahdi will identify the channel API and finally land the call to dialplan. The configuration file named as extension conf is stored in dialplan. Asterisk gateway interface (AGI) program, shell script is used to integrate Asterisk server with speech recognizer module.

3 Call Flow Design

For designing dialog manger the developer must design a call flow for his program. The intended user of our system interacts with a predefined call flow. As our system mainly deals with isolated word hence system expect isolated word from the user. Otherwise the performance of the system will be decreased. As our SDS is designed to give price, fertilizer and weather information for a particular commodity hence the user have to prompt one, two, three for respective information in Bodo language and have to prompt zero to hang up the call. The user has to prompt yes/no in Bodo for increase the confidence of the reorganization engine. Then user has to prompt the commodity name. The user will get the information of the asked commodity provided the recognition of the queries success. If the recognition of the commodity name or the option selection is failed then the system will be terminate. Following snapshot gives the flow of our proposed system (Fig. 1).

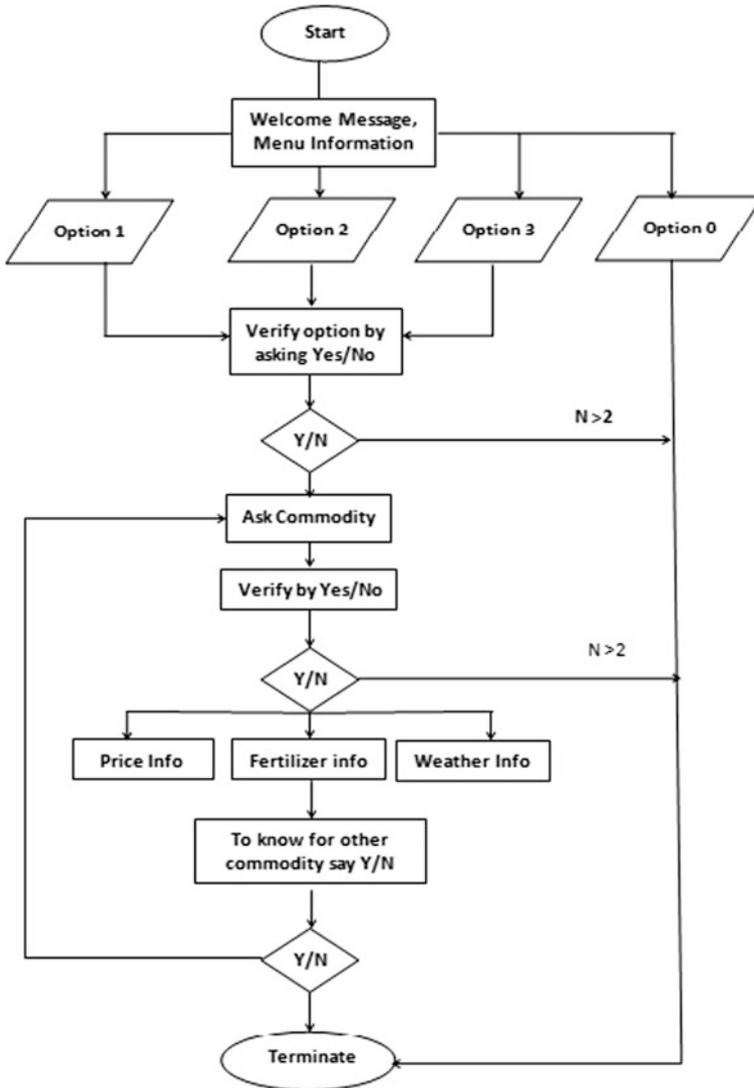


Fig. 1 Call flow of the proposed system

System: Welcome to Agricultural information system. To know the price of a commodity Say 1, to know about the fertilizer say 2, to know the weather please say 3 and to end up the call please say Zero.

User: one

System: You have chosen one. To confirm it say yes or to deny say no.

User: yes

System: What is the commodity name for which you want to know the price?

User: Potato

System: You have said Potato. To confirm it say yes or to deny say no.

User: Yes.

System: The whole sale price is 1000 per quintal and the retail price is 12 per kg.

System: For another commodity please say yes or to go to main menu say no.

User: No

System: Welcome to Agricultural information system. ...

User: Zero

System: Thank You.

4 Data Collection and Preparation

As per requirement of our system we mainly required three types of data, commodity name, digit, and yes/no. Our speech corpus consist of 31(25 + 4 + 2) unique Bodo words, out of which 25 commodity name, 4 digit (0, 1, 2, 3) and Yes/No word. We take 70 males speaker and 30 female speaker of different age group for data recording purpose. The phoneme duration information is not required for the training sentences in the Sphinx toolkit [5] as the differences in timing in the pronunciation of the training sentences are not important. The sphinx toolkit learns to recognize the words through fitting the word transcriptions on the training set. These transcriptions are used for all realizations of the same word, even though there might be variation between speakers relative to the transcription. A total of 7908 words in speech data were collected digitized at 16 bits/sample at a sampling rate of 16 kHz.

A transcription file, pronunciation dictionary and a filler dictionary also required to develop the system. A transcript is needed to represent what the speakers are saying in the audio file. In the transcription file we have used starting tag <s>, ending tag </s>), followed by the file id which represents the utterance. In Sphinx, pronunciations are specified as a linear sequence of phonemes. A phoneme is the basic sound unit which distinguishes one word from another in a particular language. Pronunciation specification is contained in the dictionary file.

5 Feature Extraction

Speech waveforms cannot be processed directly by speech recognition tool. In the feature extraction process the speech waveform is converted into feature vector for further analysis and processing. This representation and analysis is called acoustical

Table 1 MFCC feature extraction parameters

Parameter	Default value	Modified value
Pre emphasis coefficient	0.97	0.97
No. of filters in filter bank	40	31
Lower frequency (lowerf)	200 Hz	130
Upper freq (upperf)	6800 Hz	3500
Sampling rate	8000 Hz	16000
Base feature dimension	13	13
Window length	0.00256 s	0.00256 s
Frame rate	100 per s	100 per s

analysis. We have used MFCC technique for feature extraction. After obtaining the feature vector, acoustic model is build. Acoustic model is used to score the unknown voice sample. The computation steps of MFCC include framing, windowing, extracting. Speech recognition toolkit Sphinx-3 is used for building the ASR system. The tools Sphinx Train-1.0 and Sphinx 3.0.8 are used for training and decoding purposes, respectively. MFCC features of 39 dimensions, comprising of 13 base features with its first and second derivatives, are used for speech parameterization. The system is trained using five state left-to-right HMMs with two non-emitting states. This system was trained using 6358 files of commodity and digit words collected from 100 speakers. Table 1 summarizes the different parameters used for feature extraction.

6 System Training and Testing

During training time using sphinx system verify several things in seven different phases. In first phase it checks dictionary and filler dictionary with phone list file. In the second phase it checks to make sure that there is not any duplicate entry in the dictionary. In the third step checks the general format and utterance length of the training file. Utterance length must be positive. The matching of total number of lines between the transcript file and the control file is done in the fourth step. In fifth steps it checks the amount of training data and verifies that the number of `n_tied_states` is reasonable for those training data or not. In the sixth step it verifies that all the transcript words are present in the dictionary or not. In the seventh step it verifies all the phone present in the transcript file is appear in the phone list or not. We have taken 6358 utterance words in the training phase and 1550 words in the testing phase. After completion of their training we have tested by new utterance words by the new input different Bodo speakers. For evaluating performance of speech system we have used the following equations where N is the number of words in test set, D is the number of deletions, S is number of substitutions and I is the number of insertions [6].

Table 2 List of training word, occurrences in training set, and % of accuracy

Entity Name	Mapping to Bodo	Dictionary Word	Occurrence Count	Accuracy (%)
Paddy	माय	Mai	196	80
Rice	मायरं	Mairang	196	92
Potato	आलू	Aalu	198	77
Onion	सामब्राम गोजा	Saamraam-goja	196	84
Garlic	सामब्राम गूफूर	Saamraam-gufur	190	67
Ginger	हाइजें	Haijeng	197	73
Greenchilly	बानलु गोथां	Banlu-guthang	196	98
Tomato	बायगन	Baigang	193	97
Apple	आपेल	Aapel	198	84
Cabbage	बान्दा कबि	Bandakabi	203	100
Brinjal	फान्थाव	Phanthao	199	67
Radish	मुला	Mula	198	78
Carrot	गाजर	Gajar	193	67
Cucumber	थाइबें	Thaibeng	199	76
Mango	थाइजौ	Thaijoa	197	67
Masturd	बेसर	Beshar	194	84
Ridge Gourd	जिंखा	Zinkha	199	67
Banana	थालिर	Thalir	199	67
Guava	सुमफ्राम	Sumphram	200	83
Watermelon	तरमूज	Tarmuj	200	81
Pineapples	रैमालि	Raimali	199	79
Jackfruit	खान्थाल	Khanthal	200	81
Mussambi	मौसूमि	Mausumi	198	89
Coconut	नारैखल	Narikhhal	193	91
Orange	खमला	Khamla	195	84
Zero	लतिखो	Lathikho	198	96
One	नई	Nai	198	97
Two	से	Se	198	93
Three	थम	Tham	198	95
Yes	नंगौ	Nongwo	324	97
No	नडा	Nonga	316	91

Table 3 Word accuracies with diff GMM sizes and fixed no. of tied states

Decoder	No. of GMM/state				Total test files
	4	8	16	32	
Commodity	66.73	68.45	62.69	62.63	1000
Bodo-digit	83.4	81.6	87.3	83.4	500
Yes/no	91.3	94.3	92.1	93.5	50

Table 4 Word accuracies with diff no. of tied states and 16 GMM versus 32 GMM

Decoder	Sennon								Total test files
	1000		1500		2000		2500		
	16 GMM	32 GMM							
Commodity	70.23	69.5	71.34	62.8	72.38	72.8	72.69	72.6	1000
Bodo-digit	87.3	82.1	84.6	80.7	86.3	80.2	86.2	82.2	500
Yes/no	93.1	89.2	93.2	88.3	89.3	91.3	93.4	92.4	50

$PC = (N - D - S)/N \times 100$ where PC gives word correction rate.

$PA = (N - D - S - I)/N \times 100$ where PA gives word accuracy rate.

Word Error Rate (WER) = $100\% - \text{Percentage Accuracy (PA)}$

We have evaluated the performance by grouped all the item in one cluster and got WER as 83.29 in training mode and 79.12% in testing mode (Table 2).

7 Results and Discussions

The system performances in terms of word recognition accuracies with GMMs of different sizes and fixed number of tied states are given in Table 3 Subsequently the number of tied states is also varied for 16 and 32 GMMs/state systems as given in Table 4. As the SDS system with 16 and 32 GMMs/state are found to give similar performances and hence we have used 16 GMMs/state in the deployed system.

8 Conclusions

In this work we design and develop a spoken dialog system for Bodo speaking people through which they can access the information of agricultural commodities. The integration of Asterisk server, Sphinx module using with Bluetooth communication is also described. Finally we evaluate the performance of the system in terms of WER. As a part of future work, we would explore other advanced techniques like LDA dimensionality reduction, speaker adaptation methods and n-gram-based language modeling for improving the system performance.

References

1. Bachhav, N.B: Information needs of the rural farmers: a study from maharashtra, India: A Survey, in *Library Philosophy and Practice (LPP)* (2012). ISSN 1522-0222
2. Shahnawazuddin, S., Thotappa, D., Sarma, B.D., Dekka, A., Prasanna, S.R.M., Sinha, R.: Assamese spoken query system to access the price of agricultural commodities (NCC-2013)
3. Brahma, B., Barman, A.K., Boro, S.K.S.B.: Corpus building of literary lesser rich language—bodo: insights and challenges. In: *24th International Conference on Computational Linguistics* (2012)
4. Dua, M., Aggarwal, R.K., Kadyan, V., Dua, S.: Punjabi automatic speech recognition using HTK. *Int. J Comput. Sci Issues (IJCSI)* (2012)
5. <http://cmusphinx.sourceforge.net/>. Accessed 15 Sep 2016
6. Williams, J.D.: Spoken dialogue systems: challenges, and opportunities for research. In: *Proceeding of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Merano, Italy, (Dec 2009)

Random-Valued Impulse Denoising Using a Fast l^1 -Minimization-Based Image Inpainting Technique

Mayuri Kalita and Bhabesh Deka

Abstract In this chapter, an image inpainting approach based on l^1 -norm regularization is presented for the estimation of pixels corrupted by the random-valued impulse noise. It is a two-stage reconstruction scheme. First, a reasonably accurate random-valued impulse detection scheme is applied to detect the corrupted pixels. Next, the corrupted pixels are treated as missing pixels and replaced by using an image inpainting technique. The inpainting method is based on the fast iterative shrinkage thresholding algorithm (FISTA). The proposed method is fast and experimental results show that it is robust to non-Gaussian and nonadditive degradations like the random-valued impulse noise. It also outperforms similar random-valued impulse denoising schemes in terms of computational complexity while preserving the image quality.

Keywords Random-valued impulse noise · Image inpainting · Two-stage reconstruction · l^1 -regularization

1 Introduction

Digital images are usually corrupted by impulse noise (IN) due to malfunctioning of camera sensors and distortion in communication channels during acquisition, storage, or transmission. Presence of IN in images causes degradation in image quality as well as features, like, edge details, sharpness, etc. So, removal of IN from images is very important and is an active area of research. IN is random and has two

M. Kalita · B. Deka (✉)

Computer Vision and Image Processing Laboratory, Department of Electronics and Communication Engineering, Tezpur University, Tezpur 784028, Assam, India
e-mail: bdeka@tezu.ernet.in

M. Kalita
e-mail: mayurik@tezu.ernet.in

common types, salt and pepper noise (SPN) and random-valued impulse noise (RVIN). Noisy pixels of an 8-bit gray-scale image corrupted by the SPN will have either 255 or 0, while the RVIN is uniformly distributed in the interval [0 255]. Removal of SPN is comparatively easier as the intensity values of the noisy pixels are quite different from the noise-free pixels. In case of the RVIN, it is often seen that the differences are not sufficient enough thereby inducing difficulty in removing it. Here, we will deal with restoration of images degraded by the RVIN. Various techniques have been proposed in this direction including the adaptive center weighted median (ACWM) [1], the directional weighted median (DWM) [2], the signal dependent rank-ordered mean (SDROM) [3], etc. In these techniques, a noise detector distinguishes noisy and clean pixels so that only the noisy pixels are restored keeping other pixels unchanged. The efficacy of such filters depends on the capability of the impulse noise detectors. As these methods fail to detect most of the noisy pixels with increase in noise levels, so the reconstruction performance is poor when the noise level is high.

The problem of estimating missing information in an image is called the image inpainting. It is mainly used in restoring damaged photographs, removing undesired objects from an image, etc. Inpainting is an ill-posed inverse problem that can be solved by regularization-based approaches where the missing pixels are estimated from the available clean pixels of the input image along with some constraints. In this work, we have considered random-valued impulse denoising as an inpainting problem. Gray-values corresponding to locations of the noisy pixels are considered as missing pixels in the image and would be restored by image inpainting using the information of the available noise-free pixels. Edge preserving l^1 -regularization term is used along with a nonsmooth data fidelity term for the restoration of images corrupted by the RVIN [4]. They assume that the RVIN is non-Gaussian and additive. Therefore, a l^1 -norm-based data fidelity term which is effective in describing the heavy tailed nature of the RVIN is used for impulse denoising without noise detection.

In this chapter, however, we have taken the RVIN model adopted by the detection-based median filters, including the ACWM [1], the DWM [2], SDROM [3] filters, etc. These works assume that the impulse noise is non-Gaussian as well as nonadditive in nature. We consider a sparse denoising model based on l^2 -norm data fidelity term constrained by l^1 -norm regularization for image inpainting. The proposed technique is basically a two phase method. In the first phase, detection of RVIN is carried out in the corrupted image. This is done by the impulse detection stage of the ACWM filter or the adaptive switching median (ASWM) filter [5]. Once noise locations in a corrupted image are detected, they are then replaced by zeros. In the second phase, these zeros are replaced by applying the l^1 -norm regularization-based image inpainting method. Simulations are carried out using standard test images to which RVIN at different levels are added artificially. We have also compared the results of the proposed work with similar techniques for the removal of RVIN. The rest of the paper is divided into the following sections.

Section 2 gives brief accounts of the RVIN model and the sparse representation-based image inpainting technique. Section 3 illustrates the proposed method. Simulation results of the proposed method along with comparisons with other methods are reported in Sect. 4. Finally, Sect. 5 draws conclusions from the proposed work.

2 Prior Art

2.1 Impulse Noise Model

Consider a gray-scale image \mathbf{x} of size $\sqrt{n} \times \sqrt{n}$ transformed into a one dimensional vector \mathbf{x} of size $n \times 1$ by considering its rows from left to right and top to bottom. Let $x(i)$ be the gray level value of \mathbf{x} at the i th location whose values occupy the gray-scale range $[n_{\min}, n_{\max}]$. Let $y(i)$ be the gray value of a noisy pixel at the i th location, therefore,

$$y(i) = \begin{cases} r(i), & \text{with probability } p \\ x(i), & \text{with probability } 1 - p \end{cases} \quad (1)$$

where p is the noise density and $r(i) \sim U[n_{\min}, n_{\max}]$ [2]. For an 8-bit gray-scale image $n_{\min} = 0$ and $n_{\max} = 255$.

2.1.1 Sparse Representation-Based Image Inpainting

In sparse representation (SR), a signal or an image can be written as a linear combination of a few atoms or elementary signals from an over complete dictionary $\mathbf{A} \in \mathbb{R}^{m \times n}$ where $n > \underline{m}$. Thus, the observed image \mathbf{y} may be written as:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} \quad (2)$$

where $\boldsymbol{\alpha}$ is the sparse vector containing the representation coefficients of a signal \mathbf{y} and obtained by solving the following unconstrained minimization problem:

$$\min_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha}) = \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (3)$$

where λ is the Lagrange multiplier. The task of image inpainting using SR is done in [6–8]. We solve a similar problem as shown in Eq. (4), where we seek a

representation from the noise-free pixels in the corrupted image \mathbf{y} and fill in the missing pixels using the estimate $\hat{\mathbf{x}} = \mathbf{A}\hat{\boldsymbol{\alpha}}$. The proposed technique minimizes Eq. (4) to estimate the missing pixels using an efficient l^1 minimization algorithm namely, the fast iterative shrinkage-thresholding (FISTA) [9]. It iteratively finds quadratic approximations $Q(\boldsymbol{\alpha}, \mathbf{z})$ of $F(\boldsymbol{\alpha})$ in Eq. (4) around an appropriately chosen point \mathbf{z} and minimizes $Q(\boldsymbol{\alpha}, \mathbf{z})$ instead of $F(\boldsymbol{\alpha})$, that is the equivalent minimization problem is:

$$\min_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha}, \mathbf{z}) = \min_{\boldsymbol{\alpha}} \left\{ \frac{L}{2} \|\boldsymbol{\alpha} - \mathbf{u}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \right\} \quad (4)$$

where $\mathbf{u} = \mathbf{z} - \frac{1}{L} \nabla h(\boldsymbol{\alpha})$ and $h(\boldsymbol{\alpha}) = \|\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}\|_2^2$. The closed-form solution of Eq. (5) is given by:

$$\min_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha}, \mathbf{z}) = \text{soft}(\mathbf{u}, \frac{\lambda}{L}) \quad (5)$$

The detail algorithm for the FISTA is given in [9].

3 The Proposed Method

In this work, we propose a two-stage method which at first detects the noisy pixels from the observed image degraded by the RVIN. The noise corrupted pixels are detected using the ASWM [5]. This method is one of the most efficient noise detection methods proposed in the recent time. We have used this method as it is able to give reasonably less values of both miss detection and false alarms compared to some other well-known RVIN detection methods including the DWM and the ACWM. Miss detection (MD) means a true corrupted pixel but could not be detected as noise. On the other hand, false alarm (FA) means a true noise-free pixel but detected as noise.

After impulse detection, the noisy candidates from the observed image are replaced with zeros whereas the noise-free candidates are kept unchanged to produce the impulse detector output. From the output of the ASWM filter, a binary mask (\mathbf{M}) is generated which contains zero values for noise and ones for true pixels in the observed noisy image. Our aim here is to fill in the missing pixels using Eq. (4) with the FISTA. The proposed approach is summarized in Algorithm 1. To solve FISTA, the measurement matrix (\mathbf{A}) is generated as follows: $\mathbf{A} = \mathbf{M}\mathbf{W}^t$, where \mathbf{M} is the noninvertible linear observation operator defined above and \mathbf{W}^t denotes the inverse discrete wavelet transform.

Algorithm 1 Proposed inpainting algorithm
<p>Input: \mathbf{y}</p> <ul style="list-style-type: none"> • Impulse Detection <ul style="list-style-type: none"> ▪ Obtain the mask \mathbf{M} ▪ Obtain $\mathbf{y}^{obs} = \mathbf{M} \otimes \mathbf{y}$ • Inpainting holes using the FISTA <p>Initialize: $\mathbf{y} = \mathbf{y}^{obs}$, $k \leftarrow 1$, $L_0 \leftarrow 1$, $\lambda_0 \leftarrow 1$, $\beta \in (0, 1)$, $\bar{\lambda} > 0$</p> <ol style="list-style-type: none"> 1. While not converged do 2. $\mathbf{z}^k \leftarrow \boldsymbol{\alpha}^k + \frac{t_{k-1} - 1}{t_k} (\boldsymbol{\alpha}^k - \boldsymbol{\alpha}^{k-1})$ 3. $\mathbf{u}^k \leftarrow \mathbf{y}^k - \frac{1}{L_k} \mathbf{A}^T (\mathbf{A} \mathbf{z}^k - \mathbf{y})$ 4. Update L_k using $F(\boldsymbol{\alpha}) \leq Q_L(\boldsymbol{\alpha}, \mathbf{z})$ 5. $\boldsymbol{\alpha}^{k+1} \leftarrow \text{soft}(\mathbf{u}^k, \frac{\lambda_k}{L_k})$ 6. $t_{k+1} \leftarrow \frac{1 + \sqrt{4t_k^2 + 1}}{2}$ 7. $\lambda^{k+1} \leftarrow \max(\beta \lambda_k, \bar{\lambda})$ 8. $k \leftarrow k + 1$ 9. end while <p>Output: $\hat{\mathbf{x}} = \mathbf{W}^T \hat{\mathbf{a}}$</p>

Convergence of the proposed algorithm is governed by the global convergence rate of the FISTA which is analytically guaranteed. We have experimentally found that on an average the proposed system converges after 106 iterations for a gray-scale image of 512×512 with noise levels varying from 10 to 70%. This would require an average computational time of approximately 104 ms per iteration, considering an error tolerance of 10^{-4} .

4 Experimental Results

We have conducted several experiments with standard gray-scale test images collected from the USC-SIPI image database. These are representatives of image data containing varying detail levels. For example, the images of “Lena” and “Barbara” contain high textures whereas “Cameraman” and “Boat” are low textured images. These four test images are of size 512×512 . For preparation of noisy images, original test images are degraded artificially with the RVIN at varying densities from 10 to 70% using a MATLAB function. For our simulations, we have used the Daubechies wavelet “db2” with 4 levels of decomposition as this wavelet gives reasonable smoothing (Fig. 1).

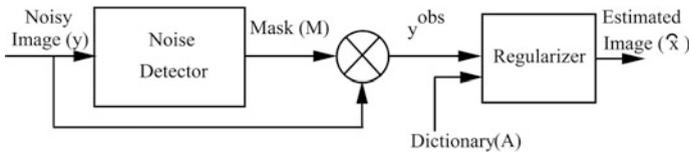
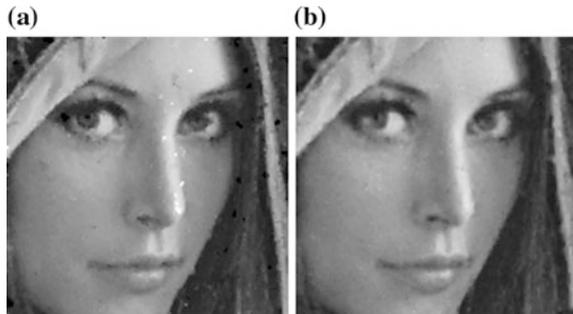


Fig. 1 Block diagram of the proposed system

Table 1 Comparison of noise detection performance for ‘Lena’

Noise (%)	ACWM			ASWM		
	MD	FA	Sum	MD	FA	Sum
50	5149	1563	6712	3828	2254	6082
60	6905	1791	8696	5157	2387	7544
70	9470	2301	11771	7216	2343	9559

Fig. 2 A reconstructed portion of ‘Lena’ image using detection schemes of **a** the ACWM and **b** the ASWM for 50% noise level with the reconstruction algorithm of the DWM [2]



4.1 Detection Performance

To detect positions of impulses in the corrupted image, ASWM detection method is used as it has better noise detection ability compared to many other existing RVIN detection algorithms. From this detector, we have obtained the sensing matrix (\mathbf{M}). The reduction of rows of \mathbf{M} depends on the percentage of noise levels. If $m \times m$ be the size of the noisy image then for 50% noise level the size of \mathbf{M} will be $0.5 m \times m$. To evaluate the performance of the ASWM detection scheme, we have measured the number of missed detections (MD) and the number of false alarms (FA) at varying noise densities. Results presented in Table 1 show the efficacy of the ASWM method over another state-of-the-art RVIN detection scheme, namely, the ACWM. It is seen that the ASWM filter is able to detect more number of the RVIN correctly than the ACWM and also gives the minimum value for the sum of MD and FA at any noise level. To see the impact of the detection scheme on image restoration, we have also presented restored images using a common filtering strategy in Fig. 2. From visual inspection, it is found that the ASWM filter clearly offers better reconstruction compared to the ACWM.

4.2 Restoration Performance

Reconstruction performances of the proposed system is measured by the peak signal-to-noise ratio (PSNR), the mean structural similarity index (MSSIM) and the computational complexity. Results are compared with similar techniques, like, [2, 6, 7], as shown in Tables 2, 3 and 4 for “Lena” corrupted with varying noise densities. All the experiments are repeated 10 times and the average of these 10 results is considered. Methods, [6, 7] have used patch-based SR using the OMP [10] for finding the missing pixels. The SR was carried out patch-wise in order to restrict the size of dictionaries to a reasonable size. Therefore, requirement of computational times by those techniques are more compared to the proposed method. From, Tables 2 and 3, we observe that the proposed filter gives almost similar and sometimes even better PSNR and MSSIM results compared to other methods. However, it requires much reduced computational time as shown in Table 4. On an average, the patch-based methods require computation time of 320 s and give PSNR of 27.96 dB whereas the proposed method requires only approximately 20 s and give PSNR of 27.25 dB for varying noise levels. Similarly, on an average MSSIM values of the proposed method are very close to those of others at

Table 2 Comparison of PSNR (dB) for different methods at varying levels of RVIN

Method	10%	20%	30%	40%	50%	60%	70%
Filter [6]	33.46	31.77	30.07	28.40	26.50	24.22	21.32
Filter [7]	31.34	27.42	24.83	22.70	23.63	21.14	18.68
Filter [2]	31.43	30.58	29.57	28.59	24.96	22.88	20.14
Proposed	33.51	31.62	29.86	28.01	27.07	24.99	22.18

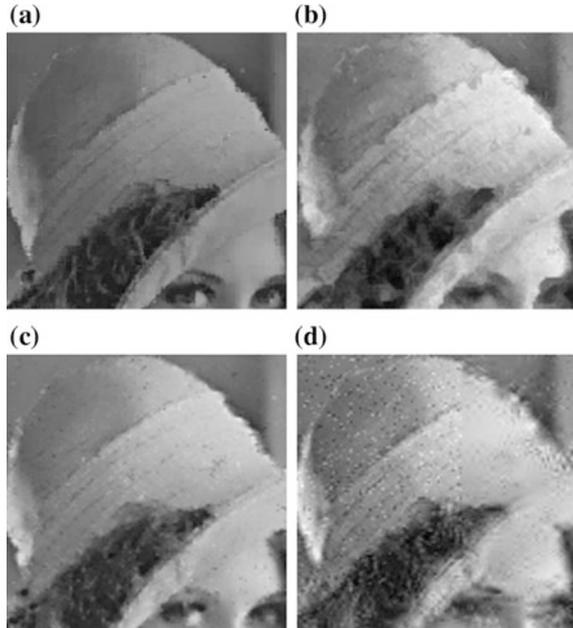
Table 3 Comparison of MSSIM values for different methods at varying levels of RVIN

Method	10%	20%	30%	40%	50%	60%	70%
Filter [6]	0.96	0.94	0.91	0.85	0.82	0.78	0.55
Filter [7]	0.87	0.73	0.61	0.61	0.49	0.50	0.44
Filter [2]	0.95	0.94	0.91	0.89	0.74	0.62	0.46
Proposed	0.93	0.89	0.86	0.93	0.89	0.81	0.65

Table 4 Comparison of CPU time (sec) for different methods at varying levels of RVIN

Method	10%	20%	30%	40%	50%	60%	70%
Filter [6]	93.65	86.71	74.66	81.44	61.33	87.84	76.36
Filter [7]	358.30	375.26	399.19	412.53	671.13	1035.50	1040.70
Filter [2]	74.04	73.70	73.73	73.57	83.95	85.58	82.96
Proposed	29.19	29.92	30.07	28.01	27.95	26.03	28.34

Fig. 3 Cropped portions of a reconstructed “Lena” image by **a** the proposed, **b** [2], **c** [6], **d** [7] at 50% noise density



varying noise levels. Therefore, the proposed inpainting-based method gives comparable results with others but at a very less computation time. The results on PSNR and MSSIM may be further improved with some more tuning of the inpainting algorithm and its parameters.

Cropped portions of a reconstructed image from the “Lena” corrupted with 50% noise are shown in Fig. 3. From the visual inspection, it is seen that the proposed method is able to give satisfactory performance as compared to others in terms of preservation of edges and other details.

5 Conclusion

This chapter investigates an efficient inpainting technique for the restoration of images corrupted by the RVIN. It provides performances similar to methods based on SR both objectively and visually even at high noise levels. Computational complexity of the proposed scheme is much lower compared to SR methods which may be an advantage for its use in multimedia and video streaming applications.

References

1. Chen, T., Wu, H.: Adaptive impulse detection using center-weighted median filters. *IEEE Signal Process. Lett.* **8**(1), 1–3 (2001)
2. Dong, Y., Xu, S.: A new directional weighted median filter for removal of random-valued impulse noise. *IEEE Signal Process. Lett.* **14**(3), 193–196 (2007)
3. Abreu, E., Mitra, S.: A signal-dependent rank ordered mean (SD-ROM) filter—a new approach for removal of impulses from highly corrupted images. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2371–2374 (1995)
4. Dong, Y., Hintermuller, M., Neri, M.: An efficient primal-dual method for L1-TV image restoration. *SIAM J. Imag. Sci.* **2**(4), 1168–1189 (2009)
5. Akkoul, S., Ledee, R., Leconge, R., Harba, R.: A new adaptive switching median filter. *IEEE Signal Process. Lett.* **17**(6), 587–590 (2010)
6. Deka, B., Bora, P.: Removal of random-valued impulse noise using sparse representation. In: *Proceedings of National Conference on Communications (NCC)*, pp. 1–5 (2011)
7. Saikrishna, P., Bora, P.: Detection and removal of random-valued impulse noise from images using sparse representations. In: *Proceedings of 20th IEEE International Conference on Image Processing (ICIP)*, pp. 1197–1201 (2013)
8. Deka, B., Handique, M., Datta, S.: Sparse regularization method for the detection and removal of random-valued impulse noise. *Multimedia Tools Appl.* pp. 1–34 (2016)
9. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.* **2**(1), 183–202 (2009)
10. Pati, Y., Rezaiifar, R., Krishnaprasad, P.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proceedings of IEEE Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, pp. 40–44 (1993)

Comparative Analysis of DFT–DCT-Spreading Strategies-Based SC-FDMA

G.B.S.R. Naidu and V. Malleswara Rao

Abstract The most upcoming technology for uplink transmission is Single-carrier FDMA (SC-FDMA) for LTE standard than OFDMA. The most preferable benefit of SC-FDMA is its slight Peak to Average Power Ratio (PAPR) for multiuser communication without affecting BER. The subcarrier mapping techniques are utilized for transmissions that are Distributed SC-FDMA and Localized SC-FDMA. The distributed mode of FDMA which is equidistant over entire bandwidth is Interleaved FDMA. In this paper, the PAPR and BER for both IFDMA and LFDMA also carried out for SC-FDMA versus OFDMA simulated and analyzed for different shift keying techniques.

Keywords OFDMA · SC-FDMA · IFDMA · LFDMA · PAPR · BER · LTE

1 Introduction

In OFDM structure, the transmitted signals have huge peak power in the time domain due to their subcarriers. So the OFDM has high PAPR than the single-carrier system and it is most determined weakness in this system. In order to enhance quality of the service, the OFDM utilizes some adaptive strategies [1–3]. The PAPR is the major drawback in the uplink transmission, meanwhile the power amplifier efficiency is problematic because of limited battery at the mobile terminal. The power amplifier enforces the nonlinear distortion due the saturation characteristics. This weakness avoided to for SC-FDMA [4, 5].

G.B.S.R.Naidu (✉) · V. Malleswara Rao
Department of ECE, GITAM Institute of Technology, GITAM University,
Visakhapatnam, AP, India
e-mail: mail2gbsrphdk@gmail.com

V. Malleswara Rao
e-mail: mraoveera@yahoo.com

The PAPR is expressed as the ratio of maximum power to average power of complex signal $S(t)$ and specified by

$$\text{PAPR}[\tilde{s}(t)] = \frac{\max|\text{Re}(\tilde{s}(t)e^{jw_c t})|^2}{E\{|\text{Re}(\tilde{s}(t)e^{jw_c t})|^2\}} = \frac{\max|s(t)|^2}{E\{|s(t)|^2\}} \tag{1}$$

The higher order digital modulations like M-ary shift keying techniques, that are PSK and QAM well suitable for OFDM. The behavior of these modulation techniques performs better for BPSK/QPSK than the next order, as increases the order the M-PSK performs poor in terms of PAPR than the lower orders of M-QAM. The PAPR for 64-PSK > 16-QAM and also the expected PAPR is more for M-QAM (where $M > 4$) [6, 7]. The probability of maximum power of the signal decreases with increase in N (subcarriers).

2 OFDM System Model

In OFDM structure, the distribution of output signal for IFFT is considered. While, the NFFT input signal have finite and independent magnitudes. These are distributed uniformly for QPSK/QAM. From the theorem of central limit, the complex signals consist of asymptotically Gaussian distribution for more number of subcarriers sufficiently [8, 9].

Let us the probability density function is

$$f_{z_n}(z) = \frac{z}{\sigma^2} e^{\frac{-z^2}{2\sigma^2}} = 2ze^{-z^2}, \quad n = 0, 1, 2, \dots, N - 1 \tag{2}$$

where $E\{z_n^2\} = 2\sigma^2 = 1$. Here Maximum value of Z_n is equal to the crest factor (CF). Assuming Z_{\max} as CF, where $Z_{\max} = \max n = 0, 1, 2, \dots, N - 1 Z_n$.

The cumulative distribution function (CDF) of Z_{\max} can be written as:

$$\begin{aligned} F_{Z_{\max}}(z) &= P(Z_{\max} < z) \\ &= P(Z_0 < z) \cdot P(Z_1 < z) \dots P(Z_{N-1} < z) \\ &= \left(1 - e^{-z^2}\right)^N \end{aligned} \tag{3}$$

here

$$P(Z_n < z) = \int_0^z f_{z_n}(x)dx, \quad n = 0, 1, 2, \dots, N - 1.$$

Now to express the P(CF) (probability of CF) which exceeds z , take the complementary CDF (CCDF) as:

$$\begin{aligned}
 \tilde{F}_{z_{\max}}(z) &= P(Z_{\max} < z) \\
 &= 1 - P(Z_{\max} > z) \\
 &= 1 - F_{z_{\max}}(z) \\
 &= 1 - \left(1 - e^{-z^2}\right)^N
 \end{aligned} \tag{4}$$

In (3) and (4) equations, these N samples are self-determining (where N is considerably large). These samples do not hold decent for bandlimited/oversampled signals. This can arise for the reason that a sampled signal may not contain the peak value of the original signal (continuous-time). And also, it is challenging in extracting the exact CDF of the oversampled signals. To overcome this, the resulting simplified CDF can be followed [10, 11]:

$$F_z(z) = \left(1 - e^{-z^2}\right)^{\alpha N} \tag{5}$$

where α is given through the actual function in which theoretical Cumulative distribution function is fitted.

The AWGN channel is preferably applicable to simulate the corresponding noise of the channel.

3 PAPR Reduction for SC-FDMA

The methods for PAPR reduction are categorized in various ways: that are signal distortion, scrambling (probabilistic), channel coding, and DFT–DCT–spreading techniques [12, 13].

The DCT (similar to DFT) is a Fourier-related transform which uses only real arithmetic. The DCT-based structure with formats of complex valued modulation requires a bandwidth similar to that of DFT-based structure with same number of sub carriers [14, 15].

The subcarrier mapping schemes: DFT–DCT–spreading technique is based on distributed and localized FDMA in Fig. 1. This is spreading of input signal with DFT/DCT. And this can later be taken into IFFT/IDCT [8, 16] which modifies the PAPR of OFDM signal to single-carrier transmission. This is exclusively used for Single-Carrier-FDMA (SC-FDMA), and implemented for uplink transmission in LTE, LTE-A, and IMT-A standard in 4G cellular communications because of its resource allocations and PAPR behavior [11, 17–20].

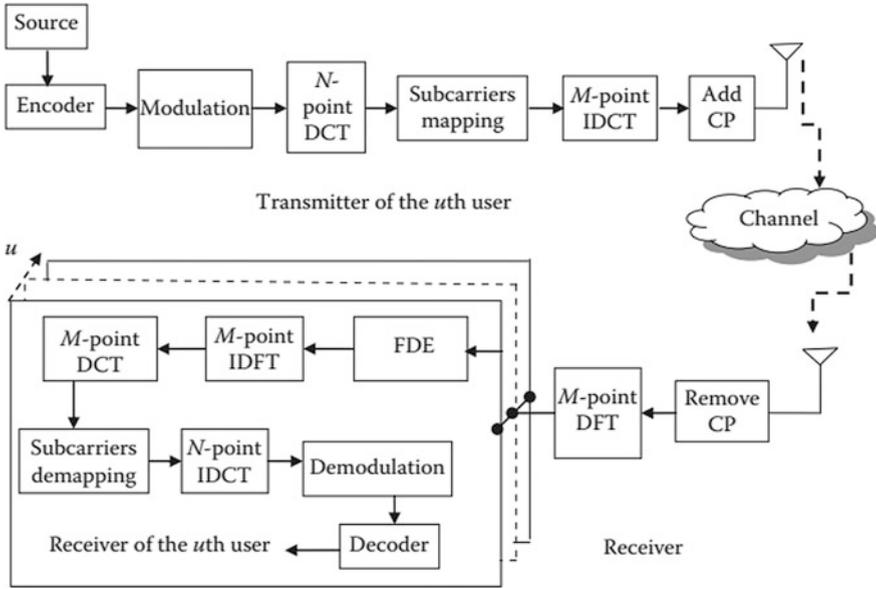


Fig. 1 DFT-DCT-based-SC-FDMA structure

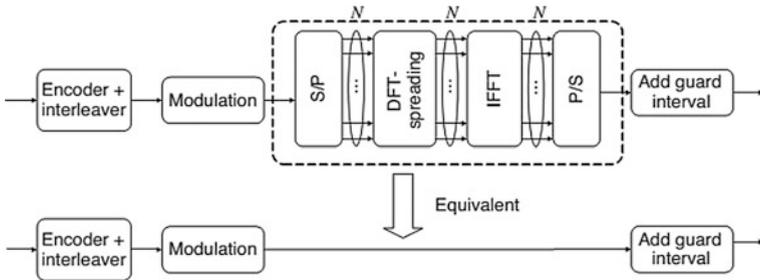


Fig. 2 The OFDMA equivalence with DFT-spreading to single carrier

4 DFT-DCT-Spreading Strategies

If we assume the spreading code for IDFT is same as that DFT size, then the OFDMA structure becomes same as that of Single-Carrier-FDMA (SC-FDMA) structure (shown in below Fig. 2). This is due to the virtual cancellations of DFT and IDFT operations. A single-carrier system has same PAPR value as that in transmitted signal [2, 6].

5 Simulation Results

5.1 DFT–DCT–Spreading for IFDMA and LFDMA

The DFT versus DCT spreading scheme is applied to SC-FDMA and OFDMA techniques having QPSK, 16/64-QAM are compared in Fig. 3 and the PAPR reductions are observed. From that the PAPR concert of the DFT as well as DCT-spreading technique are same. PAPR varies on the basis of subcarrier allocation scheme. In 16-QAM, the PAPR values with subcarrier mappings IFDMA, LFDMA and OFDMA for Complementary CDF of 1% are 3.6, 8.2 and 10.6 dB, respectively.

Similarly, the PAPR reduction of DFT versus DCT-spreading method with IFDMA and LFDMA is shown in Fig. 4. These reductions vary with respect to the roll-off factor of the RC filter (Raised-Cosine) for pulse shaping after IFFT. The increase in roll-of-factor from $\alpha = 0-1$ improves the PAPR performance of IFDMA, in contrast with LFDMA which has no significant affect due to pulse

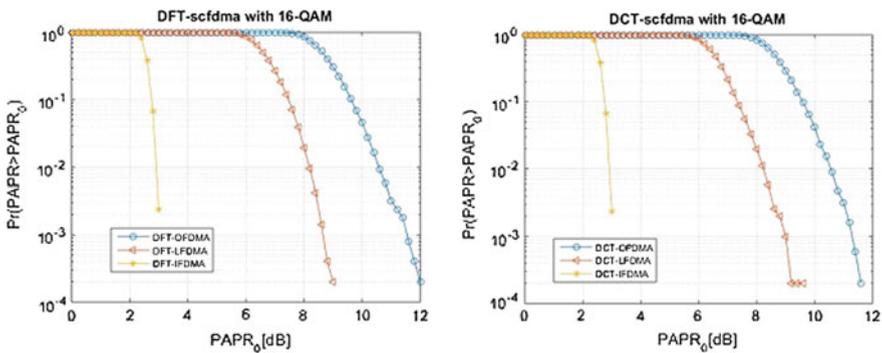


Fig. 3 DFT and DCT-based PAPR for subcarrier mappings

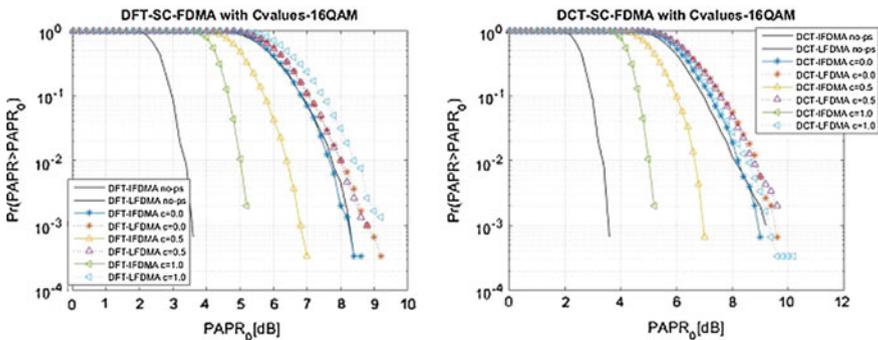


Fig. 4 DFT and DCT-spreading based PAPR with pulse shaping factor

shaping. This infers IFDMA has a trade-off between excess bandwidth and PAPR (due to linear relationship between excess bandwidth and roll-off factor).

Figures 5 and 6 illustrates the DFT versus DCT-spreading’s based PAPR for LFDMA with respect to number of subcarriers (M) allocated to each subscriber. With the increase in m value, the roll-off factors $\alpha = 0.5$ is degraded. For suppose $M = 4-256$. In this, 64 QAM is employed to SC-FDMA model having 256-point FFT.

5.2 BER Versus SNR for OFDMA and SC-FDMA

The plots show that BER versus SNR of OFDMA and SC-FDMA is almost same and are given in bellow Fig. 7.

In Fig. 7, the explanations are observed for a specific value of BER ($1e-3$). Both OFDMA and SC-FDMA will have same SNR values of 6.8 and 6.5,

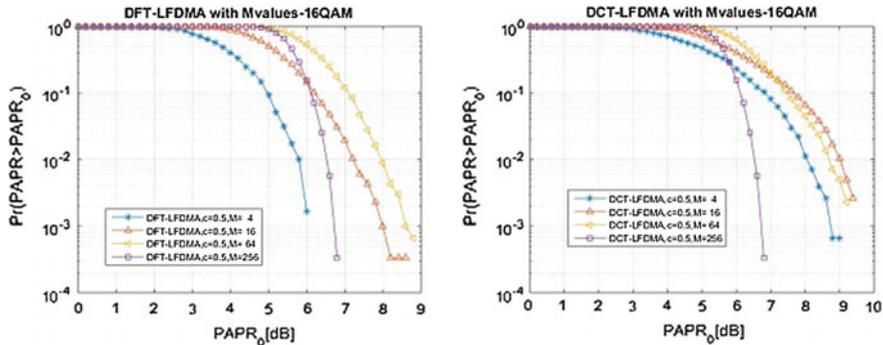


Fig. 5 DFT and DCT-spreading based PAPR with 16-QAM, when M varies up to 256

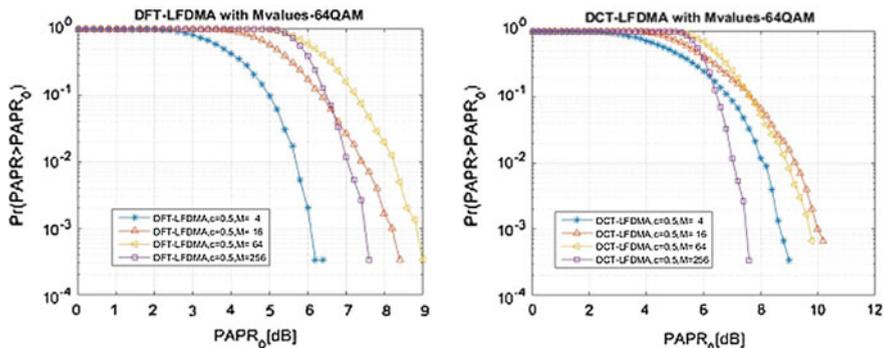


Fig. 6 DFT and DCT-spreading-based PAPR with 64-QAM, when M varies up to 256

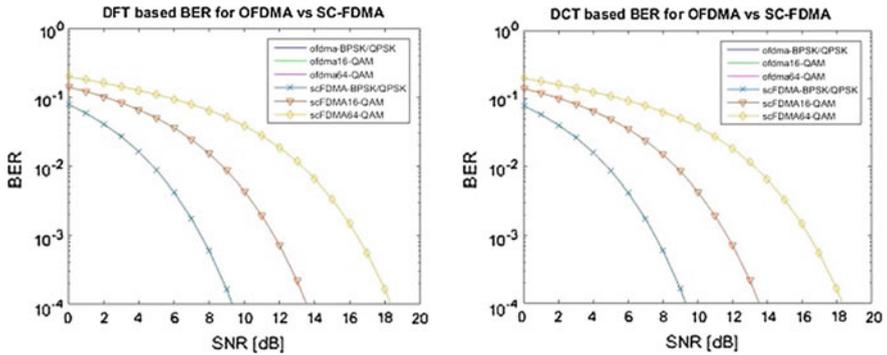


Fig. 7 DFT–DCT-based BER of OFDMA versus SC-FDMA with shift keying

respectively, for BPSK and QPSK, but an unexpected change occurs in 16-QAM and 64-QAM. The 64-QAM is more efficient in terms of BER due to their highest SNR (16.4 dB).

5.3 OFDMA Versus SC-FDMA with PAPR for Higher Order Modulations

5.3.1 Mary-PSK Scheme

The PAPR of OFDMA and SC-FDMA for $M = 2, 4, 32$ and 64-PSK modulations are analyzed. Here, it is observed that PAPR of SC-FDMA is almost less compared to the OFDMA i.e. <6 dB. Whereas the BPSK is better than the QPSK and 32/64-PSK.

5.3.2 Mary-QAM Scheme

The PAPR of OFDMA and SC-FDMA for $M = 16, 64, 128,$ and 256-QAM modulations will have performance better than higher order M-PSK are given in Fig. 8.

From Fig. 8, it can be noticed that with the increase of the order of modulation, the PAPR value for SC-FDMA performs better than the OFDMA. Hence for SC-FDMA the PAPR is <3 dB in case of 256-QAM, whereas for DFT–DCT-spreading strategies will perform same with higher modulation (from 16 to 256-QAM) but 16/64-QAM will have better than 16/64-PSK.

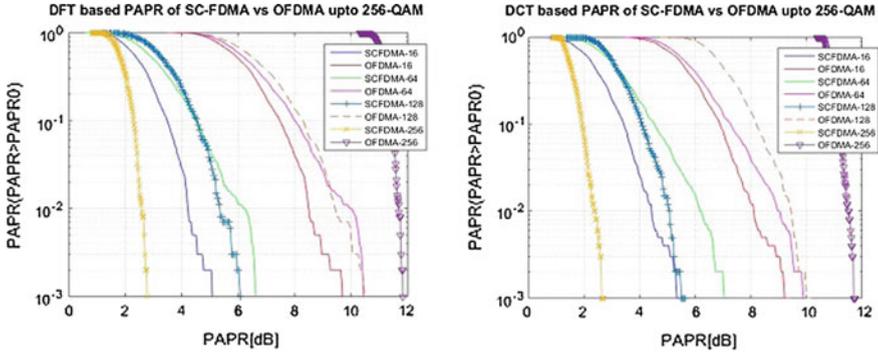


Fig. 8 DFT–DCT-based PAPR of SC-FDMA versus OFDMA for 64/256-QAM

6 Conclusion

It is concluded that when compared to OFDMA the SC-FDMA structure comprising of IFDMA and LFDMA yields better performance results of PAPR. This tremendous nature has been implemented for uplink in LTE and IMT-A. Even though the PAPR value for IFDMA is low compared to LFDMA, the LFDMA is more preferable. This is because their subcarriers are allocated in equidistant manner in the entire band of IFDMA and is not easy to implement, because it needs guard band and pilots in excess and thereby the complexity is increased. Both DCT-based SC-FDMA and DFT-based SC-FDMA are similar according to their PAPR results but DCT is preferable than DFT because of their well-organized Bandwidth.

The QoS measure at Physical layer of LTE cellular systems is indicated by the major factor known as BER. In this paper, the analysis of BER increases for a specific value of SNR as considering the higher modulations (Mary-QAM and for M up to 256) in both multiple accesses (OFDMA and SC-FDMA) utilized in LTE. The lower modulations improve and perform better way in terms of BER and PAPR. Finally, to conclude, DFT–DCT-based shift keying modulations have a significant effect on PAPR of both OFDMA and SC-FDMA. In SC-FDMA the PAPR decreases and slightly decreases in OFDMA for higher modulations. Therefore, SC-FDMA will gives better PAPR than OFDMA in all modulation systems, hence it is adopted for uplink in LTE and LTE-A.

References

1. Nogueroles, R., Bossert, M., Donder, A. Zyablov, V.: Improved performance of a random OFDMA mobile communication system. In: Proceedings IEEE VTC'98, vol. 3, pp. 2502–2506. Ottawa, ON (1998)

2. Bruninghaus, K., Rohling, H.: Multi-carrier spread spectrum and its relationship to single-carrier transmission. In: IEEE VTC'98, vol. 3, pp. 2329–2332 (1998)
3. Rahmatallah, Y.: Peak-to-average power ratio reduction in OFDM systems: a survey and taxonomy. *IEEE Commun. Surv. Tutorials.* **13**(4) (2013)
4. Ochiai, H., Imai, K.: On the distribution of the peak-to-average power ratio in OFDM signals. *IEEE Trans. Commun.* **49**(2), 282–289 (2001)
5. Galda, D., Rohling, H.: A low complexity transmitter structure for OFDM-FDMA uplink systems. In: IEEE VTC'02, vol. 4, pp. 1737–1741 (2002)
6. Wang, L., Tellambura, C.: An overview of peak-to-average power ratio reduction techniques for OFDM systems. In: Proceedings of IEEE International Symposium on Signal Processing and Information Technology, pp. 840–845 (2006)
7. Han, S.H., Lee, J.H.: An overview of peak-to-average power ratio reduction techniques for multicarrier transmission. *IEEE Wirel. Commun.* **12**(2), 56–65 (2005)
8. Lim, D.W., Heo, S.J., No, J.S.: An overview of peak-to-average power ratio reduction schemes for OFDM signals. *J. Commun. Netw.* **11**(3), 229–239 (2009)
9. Kou, Y., Lu, W., Antoniou, A.: New peak-to-average power-ratio reduction algorithms for multicarrier communications. *IEEE Trans. Circuits Syst.* **51**(9), 1790–1800 (2004)
10. Eshwariaiah, H.S., Chockalingam, A.: SC-FDMA versus OFDMA: sensitivity to large carrier frequency and timing offsets on the uplink. In: IEEE Communications Society Proceedings, GLOBECOM-09 (2009)
11. Wang, H., You, X., Jiang, B., Gao, X.: Performance analysis of frequency domain equalization in SC-FDMA systems. In: Proceedings ICC'2008, pp. 4342–4347 (2008)
12. Al-Dhahir, N., Minn, H.: A new multicarrier transceiver based on the discrete cosine transform. In: Proceedings of the IEEE Wireless Communications and Networking, vol. 1, pp. 45–50. New Orleans, LA (2005)
13. Myung, H.G., Lim, J., Goodman, D.J.: Single carrier FDMA for uplink wireless transmission. *IEEE Veh. Tech. Mag.* **1**(3), 30–38 (2006)
14. Wu, F., Abu-Rgheff, M.A.: Efficient subcarrier allocation in downlink multiuser MIMO OFDM systems. In: Proceedings of International Symposium on Wireless Communication Systems (ISWCS), pp. 51–55 (2010)
15. Wu, F., Abu-Rgheff, M.A.: An efficient suboptimal subcarrier allocation algorithm for multiuser OFDM system. In: Proceedings of International Conference on Wireless and Mobile Communications, pp. 190–195 (2009)
16. Ji, T., Zhou, C., Zhou, S., Yao, Y.: Low complex user selection strategies for multi-user MIMO downlink scenario. In: Proceedings IEEE WCNC'2007, pp. 1534–1539 (2007)
17. Raghunath, K., Chockalingam, A.: SC-FDMA for multiuser communication on the downlink. In: 5th IEEE proceedings COMSNETS-13 (2013)
18. Ciochina, C., Mottier, D., Sari, H.: An analysis of three multiple access techniques for the uplink of future cellular mobile systems. *Eur. Trans. Telecommun.* **19**, 581–588 (2008)
19. Jiang, T., Wu, Y.: An overview: peak-to-average power ratio reduction techniques for OFDM signals. *IEEE Trans. Broadcast.* **54**(2), 257–268 (2008)
20. Priyanto, B.E., Codina, H., Rene, S., Sorensen, T.B., Mogensen, P.: Initial performance evaluation of DFT-spread OFDM based SC-FDMA for UTRA LTE uplink. In: Proceedings IEEE VTC'2007, pp. 3175–3179 (2007)

A Review on Spread Spectrum Radar

Nima Donka Tamang, Samarendra Nath Sur, Soumyasree Bera
and Rabindranath Bera

Abstract Radar is used for diverse applications, and there is an enormous progression in the development of different new technologies based on its uses. It is ever growing and evolving. An overview of some of the technologies used in Radar such as spread spectrum, digital beamforming and phased array technology, and also Pulse Doppler Radar and Single Channel Spread Spectrum Digital Beamforming Radar are discussed in this paper.

Keywords Radar · Spread spectrum · Digital beamforming · Phased array
Pulse doppler radar · Spread spectrum digital beamforming

1 Introduction

RADAR's main job is to detect a target. Radio waves are transmitted and a return signal called as an ECHO is used for finding the range, angle, direction, and velocity of the target [1]. There are different approaches to improve the features such as anti-jamming, high SNR, elimination of clutter, multipath, interference, concealing the signals, CSWAP (cost, size, weight, and power) reduction, etc., of RADAR. Some of these features can be improved or gained using different approaches such as Spread spectrum, Digital beamforming, Phased array, and

N.D. Tamang (✉) · S.N. Sur · S. Bera · R. Bera
Department of Electronics and Communication Engineering, Sikkim Manipal
Institute of Technology, Sikkim Manipal University, Majitar, Sikkim, India
e-mail: nimadonka18@gmail.com

S.N. Sur
e-mail: samar.sur@gmail.com

S. Bera
e-mail: soumyasree.bera@gmail.com

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

Spread Spectrum Digital Beamforming techniques. This paper is a summary of these techniques used in contemporary Radar.

2 Overview of the Different Technologies Used in RADAR

2.1 Spread Spectrum

SS is a technique where the original data is multiplied with a code resulting in an increase of bandwidth in the frequency domain. This expanded signal is transmitted and at the receiver the original data is retrieved by de-spreading the received signal. Some of the features of spread spectrum are message protection by encrypting signals, noise, and multipath rejection, interference rejection, anti-jamming feature, high resolution ranging, etc. [2–6].

Majumder et al. has made a new Doppler tolerant and orthogonal on both transmit and receive waveform used in MIMO Radar which is a waveform diversity radar and called it as Spread spectrum coded Linear Frequency Modulation (SSCL) signaling [7]. The features of this kind of waveform is (a) such kind of waveform will remain orthogonal on transmit as well as receive (b) it is Doppler tolerant (c) and it approves two approaches of operation of received signal, firstly de-spreading to obtain original LFM signal, therefore getting resolution capability of LFM signal and then secondly if the received signal is multiplied with SS code, ultra-high resolution ability achieved to detach closely spaced targets. To develop such a new and exceptional waveform with such features, direct sequence spread spectrum (DSSS) coding techniques is integrated with linear frequency modulated (LFM) signals. Finally resulting in novel technology, i.e., Spread spectrum coded LFM (SSCL) signal [8].

Designing a radar system that will help to avoid crashes of vehicles and thus accidents has also been in great demand. Hanada et al. has proposed an intravehicular radar system which can measure the range to, the radial velocity of and (DOA) the direction of arrival of received waves from many targets. This is achieved by using DS-SS technique. SS is used for vehicular Radar because of its significant attributes, i.e., it requires low cost, it has the ability to measure long distance, and also intravehicular communication and ranging quality [9].

Digital communication is concealed in linear chirp radar signals by using Bio motivated RF steganography system. It is a remarkable method to disguise communication and is used in applications such as military operations and other secret and hidden tasks where it is necessary to stop an enemy, etc., from sensing the presence of RF communication. Reduced phase shift keying (RFSK) modulation is used in RF steganography scheme to create the modulated chirp signal develop into a similar unmodulated chirp signal. Other feature such as variable symbol durations is used for additional improvement in the security of concealed communication signal by removing the cyclostationary feature [10].

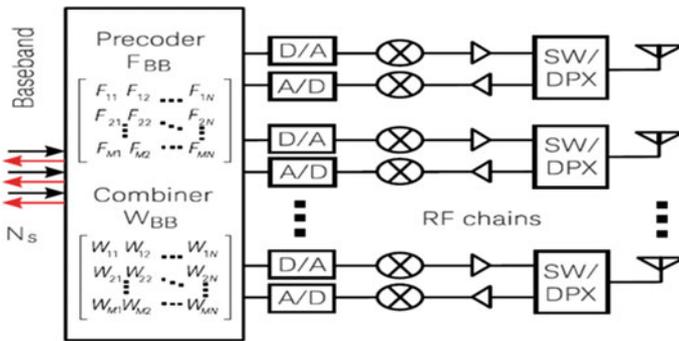


Fig. 1 Digital beamforming [12]

2.2 Digital Beamforming

The likelihood of tracking multiple targets and high SNR makes the digital beamforming approach very suitable for some applications. Digital beamforming is performed at the base band. All the operations are performed digitally. DSP chips or beamforming chips are used. It is not CSWAP (cost, size weight, and power) enabled system because it requires T/R modules for each antenna element. So, if digital beamforming is done with hundreds of antennas then hundreds of power hungry A/D converters and D/A converters are required, it will arise to hardware complexity, more power consumption and heat making the system impractical [11] (Fig. 1).

Fredrick et al. has proposed SMILE (Spatial Multiplexing of Local Elements), a novel type of smart antenna array receiver with one RF channel and Digital Beamforming skills [13]. This novel approach shrinks the necessity of bulk RF hardware for most smart antenna systems. It also decreases power dissipation as well as circuit size. This is achieved by sampling the envelope of received signal at each antenna element, followed by multiplexing, down-converting, de-multiplexing, filtering, and then finally digitizing to be processed under lesser speeds [14, 15].

2.3 Phased Array

The evolution of phased array technology started since the early 1900s. Earlier the beams of antenna array were steered using mechanical rotation later after mid-1940s the technological advancement led to the steering of the beam by adjusting the phase mechanically. Herd et al. discusses the steady advancement of phased array technology with the development of solid-state microwave IC's, analysis and designing tools, and fabricating process [16]. Due to this, the phased

array technologies find its extensive use in military systems; however, the cost is high. Also, with the swift growth in digital processor technology, leading to entirely digital arrays catering to substantial enhancements in its functioning and ability and reduction in cost. There are three types of phased array passive, active, and digital.

3 Passive Phased Array

The early electronically steered passive phased array consist of central high power transmit amplifier and receiver that fed the array. At each element a low loss power combining network and a phase shifter was connected.

The constraint of this array is that the low loss beam former networks was huge and hefty therefore it was difficult to incorporate it into airborne or space borne platforms. Secondly, the centralized transmit power amplifier was a failure. The RF losses generated by the passive beam former and phase shifter limited the radar performance [16] (Fig. 2).

4 Active Phased Array

In an active phased array, each element had individual transmit and receive amplifier. This configuration reduced the RF losses between transmit amplifier and free space. The result was the noise figure was reduced by a factor of two or more of the array thus leading to better sensitivity and extended operating ranges (Fig. 3).

Fig. 2 Passive phased array architecture [16]

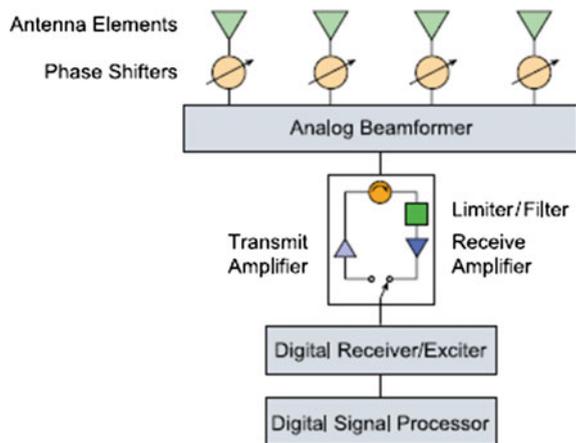
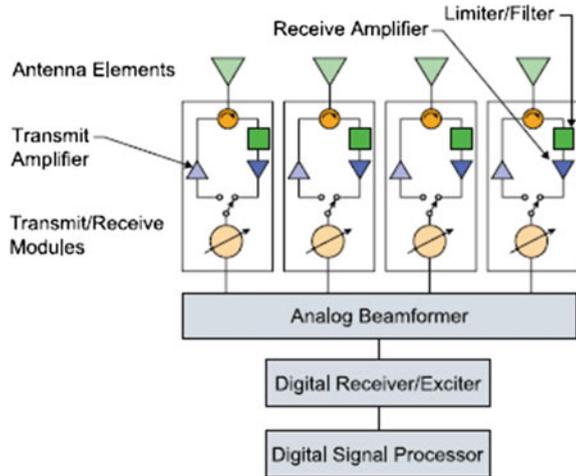


Fig. 3 Active phased array architecture [16]

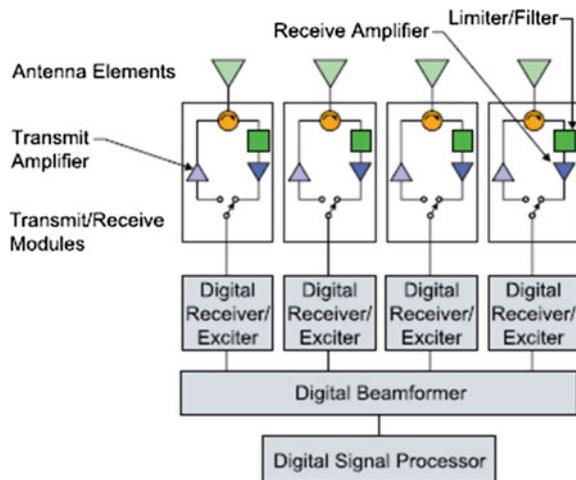


5 Digital Phased Array

In digital phased array the RF signals were first transformed to digital at element level and the beamforming was done by DSP. Two categories of digital phased array (a) Sub-array digital phased array (b) Element level digital phased array (Fig. 4).

The Sub-array digital phased array has both analog beam former and digital beam former. In Element level digital phased array the signals are converted to digital at the element level and processing is done by digital computer to form

Fig. 4 Element level digital phased array architecture [16]



digital beams. The challenge toward employing DBF is the use of RF transceiver and digitizers at each element of the array [17–19].

6 Pulsed Doppler Radar

Doppler radar has been used for a wide variety of applications, among them are measurement of speed of vehicle and storm tracking. The latest evolvement in biomedical and healthcare uses of Doppler radar is remotely detecting heartbeat and respiration in human beings [20].

Xampling techniques have been projected by Cohen et al. to successfully deal with rate bottleneck that disrupt the connection among radar signal bandwidth and sampling rate since conventional DSP is done on received samples at Nyquist rate. It is done in frequency domain [21].

Pulse Doppler radar is used to detect target under severe clutter environment and is generally used in military and commercial applications. Some, e.g., are weather, anti-ship missiles, and low-flying aircraft. It is specifically used for discovering small moving targets concealed by severely cluttered environment. It has the capability to detect range, angle and velocity but it does not have anti-jamming features [22].

6.1 Block Diagram of Pulse Doppler Radar System

I. RF Transmitter

The RF transmitter includes radar source, oscillators, modulations, amplifiers, and filters. The radar signal is a function of pulse repetition frequency (PRF), power, and pulse width. The radar transmitter has the radar source which has a pulse width of 500 ns, PRF of 10 kHz, and BW of 47 MHz. The source data is sampled at 120 MHz. This data is then sent through the modulator having the carrier frequency of 10 GHz. The up-converted signal is then amplified using an amplifier (Fig. 5).

II. Antenna and Target Return Signal with Clutter

Radar Antenna: Signal Incident Angle is 90° and antenna rotation angle is 60° which is specified to form the beam direction and propagation. Target return signal: Doppler frequency 8125 Hz and channel delay are generated to describe the target return signal with different velocities and distances.

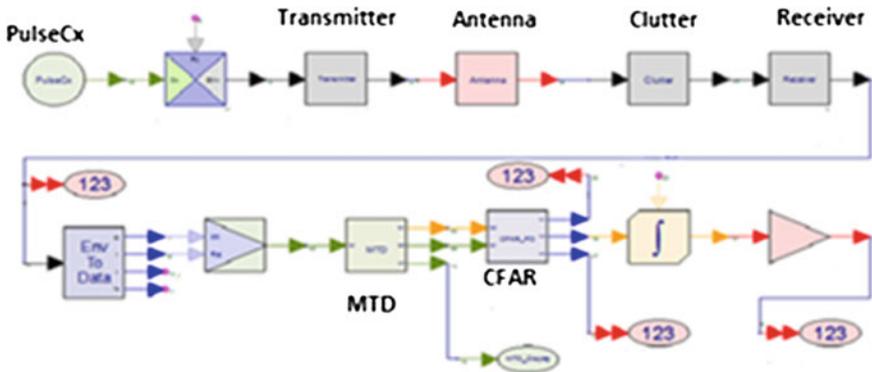


Fig. 5 Block diagram of pulse doppler radar system [23]

III. Receiver

The RF signal received via the channel is demodulated and the carrier is removed. The baseband signal is then converted to a complex signal using rect to cx block which is sent for further processing.

IV. Moving Target Detection (MTD)

MTD is an important processor for PD radar. The moving target detection (MTD) is used for detecting moving target more efficiently. FFT operators or a bank of Doppler filters cover all likely probable target Doppler shifts. The MTD output is used for CFAR processing.

V. Constant False Alarm Rate (CFAR)

The input to CFAR is the output of MTD. As Contemporary radar needs auto-detection, CFAR is used by PD Radar to govern the false alarm rate. Or else the radar will not work. It can be done in frequency or time domains or both. The averaging amplitude value of reference cells is used as the threshold to stop false alarms [24].

6.2 Results and Plots

See Figs. 6, 7 and 8.

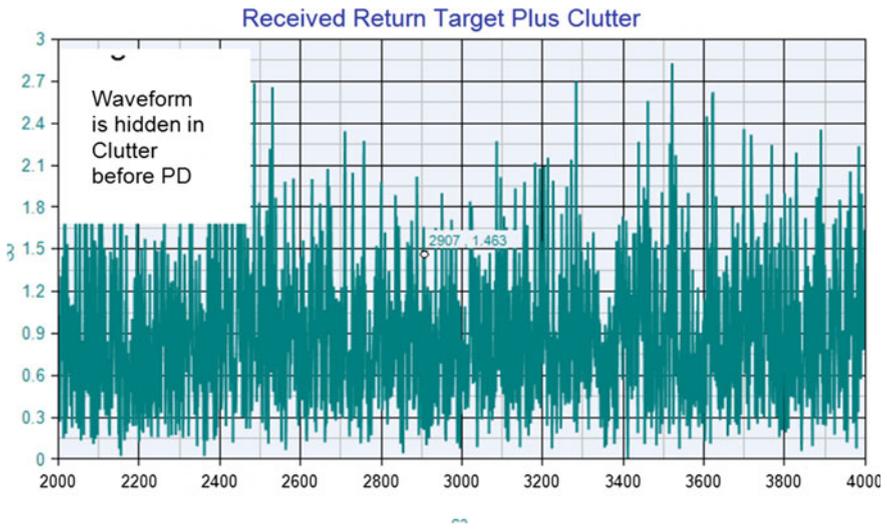


Fig. 6 Spectrum of received return target plus clutter [23]

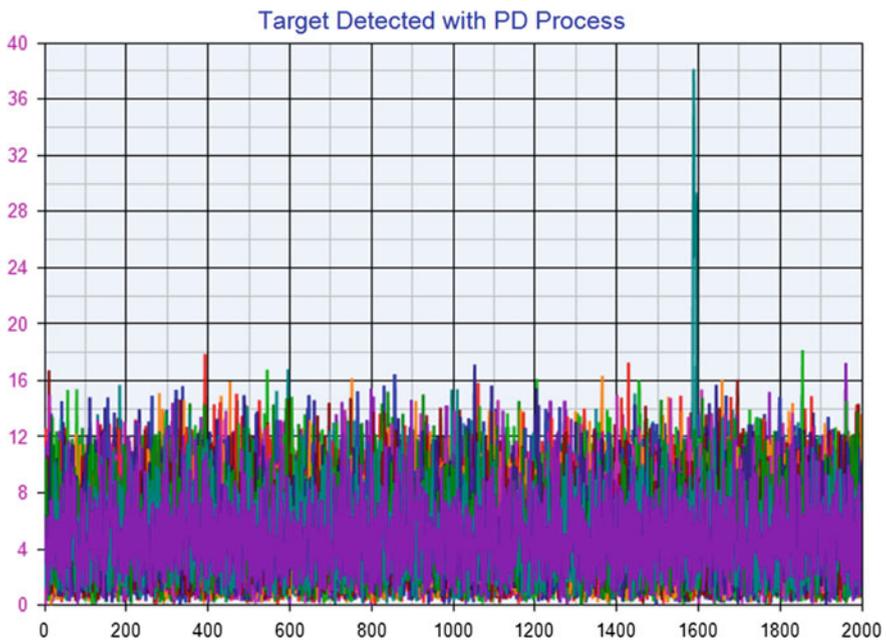


Fig. 7 Spectrum of target detected with PD process [23]



Fig. 8 Spectrum of return target shown CFAR processor [23]

7 Spread Spectrum Digital Beamforming Radar

A novel approach that combines the features of spread spectrum and digital beamforming is the Spread Spectrum Digital Beamforming technique, being used to overcome the limitation of conventional digital beamforming by use of “one transceiver per element”. This makes the SSDBF technique CSWAP (cost, size, weight and power) empowered. It uses low cost/low profile/low power DBF phased arrays. It can range to high frequencies (e.g., X, Ku, and Ka bands) and high bandwidth (e.g., hundreds of MHz to GHz). Its waveforms and parameters are agile and programmable, it can be made to support any radar or communication applications [25–27].

7.1 Block Diagram of Single Channel SSDBF

See Figs. 9 and 10.

I. Transmitter

The preliminary stage in the transmitter section is the generation of Spread Spectrum waveform, code such as polyphaser code can be used for spreading. This

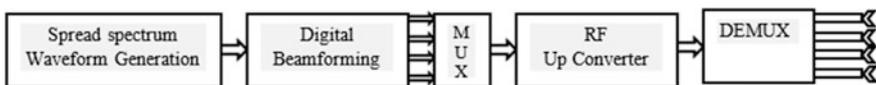


Fig. 9 Radar transmitter [28]



Fig. 10 Radar receiver [28]

process is followed by digital beamforming block. DBF process is done in the base band, the task of DBF like phase shifting and amplitude scaling are done using general purpose DSP’s or dedicated beamforming chips. Then, each data output from DBF block will be multiplexed using a multiplexer. Thus the multiplexed output will be passed through RF up-conversion block where the baseband signal will be converted to RF. After RF conversion de-multiplexing is a vital operation to provide feed for each antenna element [28, 29].

II. Receiver

In the receiver section, RF signals are received by multiple antenna elements. Multiplexer combines the output of each antenna element. The multiplexed RF signal is then down converted to BB by the RF down conversion block. This BB signal is then de-multiplexed, followed by DBF process. Then finally the signal is correlated and further passed for Radar signal processing.

7.2 Comparison of Pulse Doppler Radar and SSDBF Radar

See Table 1.

Table 1 Comparison of pulse doppler radar and SSDBF radar

S. No.	Pulse doppler	SSDBF
1	Single antenna system	Multiple antenna systems (MIMO)
2	It can measure target velocity and target distance accurately but it cannot detect target under severe channel condition	It has the capability to detect target under severe channel condition
3	No anti-jamming capability	It has anti-jamming capability
4	Spread spectrum and digital beamforming techniques not used	Use of spread spectrum and digital beamforming techniques for improved anti- jamming features and increased SNR at the receiver and is a CSWAP system
5	Does not use digital method for transmitting information	Use of digital method for transmitting information

8 Conclusion

The Pulse Doppler radar is widely used radar and it has the capability to measure target velocity and range precisely but it does not possess anti-jamming capabilities. Therefore, it is not a better choice in some applications such as missile seeker since it has zero anti-jamming capability. Therefore, in this paper we have presented an overview of different techniques used in RADAR and also projected the idea of integrating the Spread Spectrum and Digital Beamforming techniques to develop SSDBF Radar which has capability to resist jamming and is a CSWAP (cost, size, weight, and power) enabled system.

References

1. Skolnik: Radar Handbook, 2nd edn. pp. 22–24. McGraw-Hill, Inc. (1990)
2. Pickholtz, R., Schilling, D., Milstein, L.: Theory of spread-spectrum communications—a tutorial. *IEEE Trans. Commun.* **5**, 855–884 (1982)
3. Cook, C., Marsh, H.: An introduction to spread spectrum. *IEEE Commun. Mag.* **2**, 8–16 (1983)
4. Milstein, L.B., Davidovici, S., Schilling, D.L.: The effect of multiple-tone interfering signals on a direct sequence spread spectrum communication system. *IEEE Trans. Commun.* **30**, 436–446 (1982)
5. Ristenbatt, M., Daws, J.: Performance criteria for spread spectrum communications. *IEEE Trans. Commun.* **8**, 756–763 (1977)
6. Kelley, K.J., Weber C.L.: Principles of spread spectrum radar. In: Military Communications Conference, MILCOM IEEE, vol. 3, pp. 586–590 (1985)
7. Majumder, U.K., Bell, M.R., Rangaswamy, M.: A novel approach for designing diversity radar waveforms that are orthogonal on both transmit and receive, In: IEEE Radar Conference, vol. 57, pp. 394–398 (2009)
8. Friedlander, B.: On the relationship between MIMO and SIMO radars. *IEEE Trans. Signal Process.* **57**, 394–398 (2009)
9. Hanada, Y., Kohno, R.: Vehicular spread spectrum radar using multi-beam antenna. In: IEEE 47th Vehicular Technology Conference, vol. 1, pp. 116–120 (1997)
10. Zhang, Z., Nowak, M.J., Wicks, M., Wu, Z.: Bio-inspired RF steganography via linear chirp radar signals. *IEEE Commun. Mag.* **6**, 82–86 (2016)
11. Haynes, T.: A primer on digital beamforming. *Spectr. Sig. Process.* **11**, 1–15 (1998)
12. <http://www.keysight.com/en/pd2686184pnW1720EP/systemvue-phased-array-beamforming-kit>
13. Fredrick, J.D., Wang, Y., Itoh, T.: A smart antenna receiver array using a single RF channel and digital beamforming. *IEEE Trans. Microw. Theory Tech.* **12**, 3052–3058 (2002)
14. Jeon, S.S., Wang, Y., Qian, Y., Itoh, T.: A novel smart antenna system implementation for broad-band wireless communications. *IEEE Trans. Antennas Propag.* **5**, 600–606 (2002)
15. Cheng J., Kamiya, Y., Ohira, T.: Adaptive beamforming of ESPAR antenna using sequential perturbation. In: Microwave Symposium Digest, 2001 IEEE MTT-S International, vol. 1, pp. 133–136 (2001)
16. Herd, J.S., Conway, M.D.: The evolution to modern phased array architectures. *Proc. IEEE* **3**, 519–529 (2016)
17. Fenn, A.J., Temme, D.H., Delaney, W.P., Courtney, W.E.: The development of phased-array radar technology. *Lincoln Lab. J.* **2**, 321–340 (2000)

18. Rocca, P., Oliveri, G., Mailloux, R.J., Massa, A.: Unconventional phased array architectures and design methodologies—a review. *Proc. IEEE*. **3**, 544–560 (2016)
19. Boles, T., Varmazis, C., Carlson, D., Palacios, T., Turner, G.W., Molnar, R.J.: High voltage GaN-on-silicon HEMT. *Phys. Status Solidi*. **5**, 844–848 (2013)
20. Li, C., Lubecke, V.M., Boric-Lubecke, O., Lin, J.: A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring. *IEEE Trans. Microw. Theory Tech.* **61**, 2046–2060 (2013)
21. Cohen, D., Eldar, Y.C.: Reduced time-on-target in pulse Doppler radar slow time domain compressed sensing. In: 2016 IEEE Radar Conference, pp. 1–4 (2016)
22. Richards, M.A.: *Fundamentals of Radar Signal Processing*. Tata McGraw-Hill Education (2005)
23. Agilent SystemVue 2013.01
24. Mahafza, B.R.: *Radar Systems Analysis and Design Using MATLAB*. 3 edn., pp. 282–283. CRC Press (2013)
25. Bergamo M.A.: Spread spectrum digital beamforming, in phased array systems and technology (ARRAY). In: IEEE International Symposium, vol. 7, pp. 651–657 (2010)
26. Bergamo, M.A., Weedon, W.H.: Spread spectrum digital beamforming virtual array and scalability to millimeter wavelength frequencies MMW-SSDBF virtual array, In: IEEE International Symposium on Phased Array Systems and Technology, pp. 649–652 (2013)
27. Woo, T.K.: Orthogonal variable spreading codes for wideband CDMA. *IEEE Trans. Veh. Technol.* **51**, 700–709 (2002)
28. Bera, S., Sur, S.N., Bera, R.: Spread spectrum-digital beam forming radar with single rf channel for automotive application. In: 9th International Radar Symposium, pp. 1–5 (2013)
29. Lu, D., Yao, K.: Importance sampling simulation techniques applied to estimating false alarm probabilities. In: IEEE International Symposium on Circuits and Systems, vol. 446, pp. 598–601 (1989)

An Empirical Analysis of Three Moments on Sattriya Dance Single-Hand Gestures Dataset

Mampi Devi and Sarat Saharia

Abstract The single-hand gestures of Indian classical dance are termed as ‘Asamyukta Hastas,’ which is a combination of two Sanskrit words, *asamyukta* meaning ‘single’ and *hastas* meaning ‘hand gestures’. There are eight officially recognized classical dance forms in India. This paper focuses on the 29 single-hand gestures of Sattriya dance which is one of the Indian classical dance forms. It presents an analysis on recognition of single-hand gestures of Sattriya dance form images using different classifiers such as k -nearest neighbor (k -NN), naive Bayes, Bayesian network, decision tree, and Support Vector Machine (SVM). In this work, we have used Hu’s seven invariant moments, Zernike moments, and Legendre moments up to tenth order each. In this analysis, it indicates that Legendre moments show a better performance compared to other moments for all variation of dataset, and could achieve an accuracy of 96.03%.

Keywords Sattriya classical dance · Hand gestures recognition
Moments features · Machine learning classifiers

1 Introduction

Sattriya dance originated in the state of Assam. It is one of the popular dance forms among the eight Indian classical dance forms. This classical dance uses several hand gestures, most of them are similar to other classical dances which is performed by both male and female dancers. The Indian classical dances on which research work of gesture recognition have been done are Bharat Natyam [1–3], Odissi [4, 5]. However, no work has been reported in the literature on Sattriya dance. One of the

M. Devi (✉) · S. Saharia

Department of Computer Science and Engineering, Tezpur University, Tezpur 784028, Assam, India

e-mail: mampi.mca08@gmail.com

S. Saharia

e-mail: sarat@tezu.ernet.in

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443, https://doi.org/10.1007/978-981-10-4765-7_69

665

applications of this research is to create universal communication environment for a dance drama which is independent of the language of the associated song. It has also applications in the self-judgement and e-learning of dances. In the art of dance learning, hand gestures are first and the most important step to learn because of its flexibility and utility. The hand gestures used in classical dance form are known as mudras, where as in Sattriya dance they are known as hastas [6]. The single-hand gestures which are used in Sattriya dance are known as Asamyukta hastas. It is mentioned in the book ‘Sattriya Nrityar Rup Darshan’ by Borah [6], there are totally 76 hastas used in Sattriya dance. Among them 29 are single-hand gestures and remaining 47 are double hand gestures. These double hand gestures are divided into Samyukta hastas and Nritya hastas.

This paper focuses on single-hand gestures of Sattriya dance. The main challenges of hand gesture recognition are the feature extraction which is the most important step for any recognition system. Moments features are considered as the rotation, translation, and scaling invariant features, which has many applications in shape analysis and hand gesture recognition. There are two types of moments in the literature: non-orthogonal and orthogonal. Non-orthogonal moments are those moments where reconstruction of images is quite difficult. However, the orthogonal moments has minimal information redundancy which makes it easy in reconstruction of images.

In this paper, three types of moments namely M.K. Hu’s seven invariant moments, Zernike orthogonal moments, and Legendre moments are used on Sattriya dance single-hand gestures dataset. Also, the performance analysis of different machine learning classifiers on this feature set is evaluated. The experimental results show that the performance of Legendre moments better than other classifiers, which can give up to 96.03% accuracy. The rest of the paper is organized as follows: Sect. 2 provides the overview of proposed gesture analysis system. Experimental results are discussed in Sect. 3. Section 4 presents the related work in this domain. Finally, Sect. 5 concludes the paper and discusses the scope of future work.

2 Proposed Methods for Gesture Recognition

The objective of the proposed gesture recognition method is to recognize a set of 29 hastas of Sattriya dance. The schematic diagram for overall recognition system for asamyukta hastas is shown in Fig. 1. The proposed recognition system consists of the following steps: (i) Preprocessing, (ii) Moments Features Computation, and (iii) Classification. Each step of this diagram is explained in the next subsections.

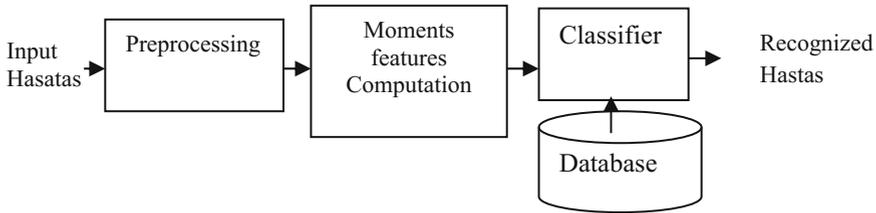


Fig. 1 Schematic diagram for the proposed recognition system

2.1 Preprocessing

This step is an important step to do any further processing. In this method, the preprocessing step is done in two substeps: background removal and Gaussian filtering. Background removal is done by using GMM [7] on RGB images. Then, a Gaussian filter approach has been used to make the images smooth and noise free. Thereafter, the RGB images are converted into gray images and binary images by using MATLAB function and continue to the next section. The binary images are represented by '1' and '0' where objects are represented by '1' and remaining part by 0. Size of all the images are normalized into 200 x 200 pixels. After size normalization, the Gaussian filter with window size of 25 x 25 and sigma value 15 is used to make all the images noise free.

2.2 Moments Features Computation

The main aim of feature extraction is to transform the input images into set of numeric values which is also known as feature vector. The extracted features are used to find out the meaning of gestures. In this paper, three set of moments features are extracted for all the 1450 images. They are Hu's seven invariant moment features [8], Zernike orthogonal moment [8] features up to tenth order, and Legendre moments [8] up to tenth order.

3 Experimental Results

The experimental description and the obtained outcome for the proposed method on Sattriya dance single-hand gestures dataset are discussed in the following subsections.

3.1 Dataset Creation

The dataset used in this experiment is our own dataset. This dataset is created by capturing 1450 images from six volunteer dancers for 29 hastas using a digital camera of 13 megapixels. The dataset sample of 29 hastas is shown in Fig. 2 and the name of each hasta is ardhasuchi, kartarimuka, pataka, tripataka, ardhachandra, tamrachuda, chatura, urnanava, sarpasira, shasaka, khatkhamukha, singhamukha, padmokosha, trishula, sandhangsha, suchmukha, krishnasarmukha, dhanu, sikhara, mukula, bhromora, mustika, granika, ankusha, kopittha, tantrimukha, alpagma, ban, hangsamukha, respectively. Out of six volunteers 290 images from each of four dancers and 145 images from each of the remaining two dancers are collected. So, $1450 (29 \text{ types} \times 10 \text{ variation} \times 4 \text{ individuals}) + (29 \text{ types} \times 5 \text{ variation} \times 2 \text{ individual})$ with different angle variations are used for this experiment. The samples of datasets are collected from different Sattras like Nikamul, Kamalabari, Auniati Sattras of Assam. The volunteer dancers belong to different age groups ranging from 16 to 26. All the captured images are cropped and resized to 200×200 resolutions for this experiment.

3.2 Results and Discussions

In this paper, we adopt Hu's invariant moments, Zernike moments, and Legendre moments methods on asamyukta hastas (single-hand gesture) of Sattriya dance dataset to extract the features value from the images. The overall classification accuracy achieved for these hastas of Sattriya dance for Hu's invariant feature set, Zernike moments feature set, and Legendre moments are shown in three consecutive tables. Here, five popular machine learning classifiers are selected k -NN, Bayesian network, naive Bayes, decision tree, and SVM to classify the asamyukta hastas. Moreover, the SVM classifier is experimented on three types of kernel: linear, polynomial and RBF kernel. The analysis of Hu's moment feature set with



Fig. 2 Sample of 29 asamyukta hastas (single-hand gestures) of Sattriya dance

different classifiers is shown in Table 1. All the experimented are done in binary image dataset. Tables 2 and 3 show the performance analysis of Zernike moments and Legendre moments features. Among all these classifiers, Legendre moments achieves good performance with 96.73% accuracy. The comparison evaluation result of five classifiers using three moments features set are shown as Fig. 3.

Table 1 Accuracy of Hu’s moments features set

Classifiers	Total no. of instances	Correctly classified	Average recognition rate (%)
<i>K-nn</i> ($n = 5$)	1015	727	71.68
Bayesian network	1015	637	72.66
Naïve Bayes	1015	720	70.93
Decision tree	1015	818	80.59
SVM-linear Kernel ($c = 1$, $E = 1$)	1450	986	68.00
SVM-polynomial	1450	960	66.20
SVM-RBF ($c = 9$, $\text{gamma} = 0.033$)	1450	1031	71.10
Average recognition rate for Hu’s moment features			71.59

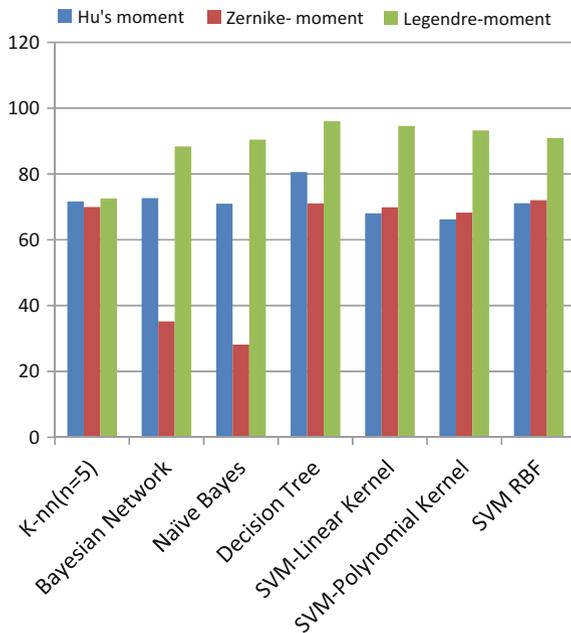
Table 2 Accuracy of Zernike moments features set

Classifiers	Total no. of instances	Correctly classified	Average recognition rate (%)
<i>K-nn</i> ($n = 5$)	1015	710	69.95
Bayesian network	1015	380	35.17
Naïve Bayes	1015	285	28.16
Decision tree	1015	714	71.03
SVM-linear kernel ($c = 1$, $E = 1$)	1450	1013	69.86
SVM-polynomial ($C = 1$, $E = 1$)	1450	990	68.27
SVM-RBF ($C = 8$, $\text{gamma} = 0.0321'$)	1450	1044	72.00
Average recognition rate for Zernike moments features			59.20

Table 3 Accuracy of Legendre moments features

Classifiers	Total no. of instances	Correctly classified	Average recognition rate (%)
<i>K</i> -nn (<i>n</i> = 5)	857	622	72.57
Bayesian network	857	750	88.37
Naïve Bayes	857	775	90.43
Decision tree	857	823	96.03
SVM-linear kernel (<i>c</i> = 1, <i>E</i> = 1)	857	807	94.61
SVM-polynomial (<i>C</i> = 1, <i>E</i> = 1)	857	799	93.23
SVM-RBF (<i>C</i> = 8, gamma = 0.0321')	857	779	90.90
Average recognition rate for Legendre moments features			89.45

Fig. 3 Comparison results for three moments features set with different classifiers on Sattriya dance single-hand gestures dataset



4 Related Work

Generally, the hand gestures are mainly used by deaf people. However, in day-to-day life, hand gestures are even used by any normal persons for the better understanding of expressions while talking or exchanging views in a conversation. In similar way, Irrespective of the language associated with the song, dance

Table 4 Related works based on moments and gesture recognition

Year	Research domain	Description
1961 [10]	Image analysis	Geometric moment for automatic character recognition
1962 [11]	Pattern recognition	Geometric moment for digital character recognition
1971 [12]	Ship identification	Automatic ship photo interpretation using moment features
1977 [13]	Artifact identification	Automatic recognition of aircraft types from optical images using moments
1977 [14]	Pattern matching	Pattern matching under affine transformations
1980 [15]	Introduced orthogonal moment	Provided the basic concept on Legendre and Zernike moment
1981 [16]	Geometric moment became extended to radial moments	Provided a general framework for deriving radial and angular invariants
1984 [17]	Introduced complex moments	
2000 [18]	Gestures recognition	Pose classification
2003 [19]	Image analysis using Krawtchouk moments	To extract the local features from any region-of-interest in an image
2003 [20]	Legendre moments	Representation of printed character with Legendre moments
2004 [21]	Chebyshev moments and Legendre moments	A comparative study on two orthogonal moments for representation of printed characters
2007 [22]	Hand gesture recognition	Visual-based human machine interface using hand gestures
2010 [23]	Hand gesture recognition	A study on static hand gestures recognition using moments
2016 [24]	Hand gesture recognition	Indian sign languages

gestures have capability to express universal expressions. Historically, Efron was the pioneer of the research work on gestures recognition. He started the work in the year 1941 [9]. And, in the area of moments invariant Hu's first started the work in the year 1961 [10]. Some of the related works based on moment's invariants and gestures recognition are presented in Table 4.

5 Conclusion and Future Direction

In this paper, a Sattriya dance single-hand (asamyukta hastas) dataset of 1450 images are used. We adopt Hu's seven invariant moments, Zernike moments and Legendre moments feature extraction method on this dataset. Five machine learning classifiers namely K -nearest neighbor, Bayesian network, naive Bayes, decision tree, and SVM with different kernels are used to analyze the Sattriya dance single-hand gestures recognition. In this experiment, the result indicates that Legendre moments classifier

shows a better performance compared to other classifiers for all variation of dataset, and could achieve an accuracy of 96.03%. The results show that the recognition accuracy for each individual class is good but it can be improved. The reason behind some of the hand gestures are misclassified may be because of the fact that most of the asamyukta hastas are very similar to each other, thus low accuracy. Thus, our future work focuses on the application of other orthogonal moments on these images so that a better accuracy can be achieved.

References

1. Hariharan, D., Acharya, T., Mitra, S.: Recognizing hand gestures of a dancer. In: Pattern Recognition and Machine Intelligence, pp. 186–192. Springer, Berlin (2011)
2. Saha, S., Ghosh, L., Konar, A., Janarthanan, R.: Fuzzy 1 membership function based hand gesture recognition for bharatanatyam dance. In: Computational Intelligence and Communication Networks (CICN), pp. 331–335. IEEE (2013)
3. Mozarkar, S., Warnekar, C.S.: Recognizing Bharatnatyam Mudra using principles of gesture recognition. *Int. J. Comput. Sci. Netw.* **2**(2), 46–52 (2013)
4. Konar, A., Atulya, K.N., Saha, S., Ghosh, S.: Gesture recognition from indian classical dance using kinect sensor. In: Fifth International Conference on Computational Intelligence, Communication Systems and Networks, pp. 3–8. IEEE (2013)
5. Saha, S., Ghosh, S., Konar, A., Nagar, A.K.R.: Identification of Odissi dance video using Kinect sensor. In: Advances in Computing, Communications and Informatics (ICACCI), pp. 1837–1842. IEEE (2013)
6. Borah, K.: Satriya Nrityar Rup darshan. Grantha-Sanskriti. Tarazan, Jorhat (2009)
7. Mukherjee, S., Das, K.: An adaptive GMM approach to background subtraction for application in real time surveillance. ArXiv preprint [arXiv:1307.5800](https://arxiv.org/abs/1307.5800), vol. 2.1, pp. 39–46 (2013)
8. Mukundan, R., Ramakrishnan, K.R.: Moment Functions in Image Analysis: Theory and Applications. World Scientific, Singapore (1998)
9. Efron, D.: Gesture and Environments: A Tentative Study of Some of the Spatio-Temporal and “Linguistic” Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York city, Living Under Similar As Well As Different Environmental Conditions. King’s Crown Press, Morningside Heights (1941)
10. Clowes, M.B., Parks, J.R.: A new technique in automatic character recognition. *Comput. J.* **4**(2), 121–128 (1961)
11. Alt, F.L.: Digital pattern recognition by moments. *JACM* **9**(2), 240–258 (1962)
12. Smith, F.W., Wright, M.H.: Automatic ship photo interpretation by the method of moments. *IEEE Trans. Comput.* **100**(9), 1089–1095 (1971)
13. Dudani, H., Breeding, K.J., McGhee, R.B.: Aircraft identification by moment invariants. *IEEE Trans. Comput.* **100**(1), 39–46 (1977)
14. Dirilten, H., Thomas, G.N.: Pattern matching under affine transformations. *IEEE Trans. Comput.* **100**(3), 314–317 (1977)
15. Teague, M.R.: Image analysis via the general theory of moments. *JOSA* **70**(8), 920–930 (1980)
16. Reddi, S.S.: Radial and angular moment invariants for image identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **3**(2), 240 (1981)
17. Abu-Mostafa, Y.S., Demetri, P.: Recognitive aspects of moment invariants. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 698–706 (1984)

18. Ng, C.W., Ranganath, S.: Gesture recognition via pose classification. In: Proceedings of 15th International Conference on Pattern Recognition, pp. 699–704. IEEE (2000)
19. Yap, P.T., Paramesran, R., Ong, S.H.: Image analysis by Krawtchouk moments. *IEEE Trans. Image Process.* **12**(11), 1367–1377 (2003)
20. Saharia, S., Bora, P.K., Saikia, D.K.: Representation of printed characters with Legendre moments. In: 5th International Conference on Advances in Pattern Recognition, pp. 280–283. ISI Kolkata (2003)
21. Saharia, S., Bora, P.K., Saikia, D.K.: A comparative study on discrete orthonormal Chebyshev moments and Legendre moments for representation of printed characters. In: *ICVGIP*, pp. 491–496 (2004)
22. Venkataramana, A.: Radial Krawtchouk moments for rotational invariant pattern recognition. In: 6th International Conference on Information, Communications Signal Processing, pp. 1–5. IEEE (2007)
23. Padam, S.P., Bora, P.K.: A study on static hand gesture recognition using moments. In: International Conference on Signal Processing and Communications (SPCOM), pp. 1–5. IEEE (2010)
24. Gupta, B., Shukla, P., Mittal, A.: K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion. In: International Conference on Computer Communication and Informatics (ICCCI), pp. 1–5. IEEE (2016)

Implementation of Doppler Radar at 24 GHz

Manisha Das, Tarini Singh and Soumyasree Bera

Abstract Doppler radar has applications in many fields. It can be used in weather forecasting. It can also be used to check speeding violations of vehicles and in air traffic control radar to detect incoming and outgoing aircrafts. Astronomers use it to discover new planets and stars. This effect is analyzed using transforms like Fast Fourier Transform (FFT), Short-Time Fourier Transform (STFT), and Wigner–Ville Distribution (WVD). WVD has been preferred more for its high resolution. In software, low Doppler will be detected while in hardware due to limitations, low Doppler will not be detected. In this report, the methods of extraction and mitigation of Doppler are explained with viable algorithms and results.

Keywords Fast Fourier Transform (FFT) · Short-Time Fourier Transform (STFT) · Wigner–Ville Distribution (WVD) · Doppler radar

1 Introduction

The Doppler radar uses the Doppler effect to measure the velocity of a moving object. The change in frequency of a wave caused by the relative motion between the source and the observer is called Doppler effect. A radar transmits a signal toward the target and after interaction of the signal with the target, it returns to the receiver. The returned signal is affected by the characteristics of the target. Consider an object that emits wave. When the object moves toward the observer, the frequency nearer to the

M. Das (✉) · T. Singh · S. Bera
Sikkim Manipal Institute of Technology, Sikkim Manipal University,
Majitar, Sikkim, India
e-mail: manisha.07july.md@gmail.com

T. Singh
e-mail: tarini_singh91@yahoo.com

S. Bera
e-mail: soumyasree.bera@gmail.com

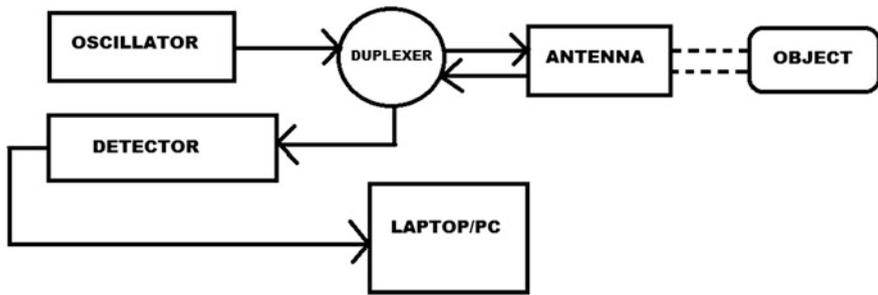


Fig. 1 Block diagram of doppler radar

observer increases, and when it moves away from the observer, the frequency decreases.

Microwave sensor radars are applicable to calculate the distance and velocity of a detected object. If the radar frequency is higher, the resolution will be better and the size of the sensor will be smaller [1]. Radars are being used for long-range forecasting of southwest monsoon rainfall over India using Ensemble Linear Multiple Regression (EMR) and Projection Pursuit Regression (PPR) [2]. Radar validation is used for surface rainfall intensity which is more accurate because radar measurement is totally based upon reflectivity factor and the surface instrument observations [3]. Doppler radar application can also be implemented using FFT in FPGA. Heart rate detection accuracy can be generated using stepped frequency continuous wave in FPGA [4, 5] (Fig. 1).

2 Mathematical Modelling

There are different methods that can be used for implementation of Doppler radar. They are as follows.

2.1 Fast Fourier Transform

FFT is basically the fastest method for finding the Discrete Fourier transform (DFT). A complicated signal can be broken down into simple waves. In Fourier transforms, the broken waves are analysed in terms of frequencies for the analysed signal [6].

Mathematical Expression:

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\left(\frac{2\pi}{N}\right)nk} \quad (1)$$

Here, $k = 0, 1 \dots N - 1$. An FFT rapidly computes such transformations by factorizing the DFT matrix into a product of sparse (mostly zero) factors.

2.2 Short-Time Fourier Transform

In Short-Time Fourier Transform, the function $f(t)$ is divided into sections by applying a window $w(t)$ and then the Fourier transform of each section is taken out [7].

A Short-Term Fourier transform that correlates f with all time–frequency shifts (τ, ω) of the window $w(t)$ can be defined as

$$S_W(\tau, \omega) = \int_{-\infty}^{\infty} f(t)W^*(t - \tau)e^{-j\omega t} dt \quad (2)$$

The window $w(t)$ is symmetric and real, so that

$$w * (t - \tau) = w(t - \tau) = w(\tau - t). \quad (3)$$

2.3 Wigner Distribution Function

The WVD does not suffer from leakage effects as the STFT does. Hence, the WVD gives the best spectral resolution. The autocorrelation function in terms of average time and time lag is calculated and then the lag is Fourier transformed. This gives the Wigner function.

$$W_x(t, f) = \int_{-\infty}^{\infty} x\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right)e^{-2\pi i \tau f} d\tau \quad (4)$$

Limitations: If FFT and STFT is to be compared, WVD has better time and frequency resolution. But the limitation of the WVD is the cross-term interference. This phenomenon shows the frequency components that do not exist in reality which affect the illustration of the time–frequency plane. Practical applications of WVD are affected by cross-term interference [8–11] (Fig. 2).

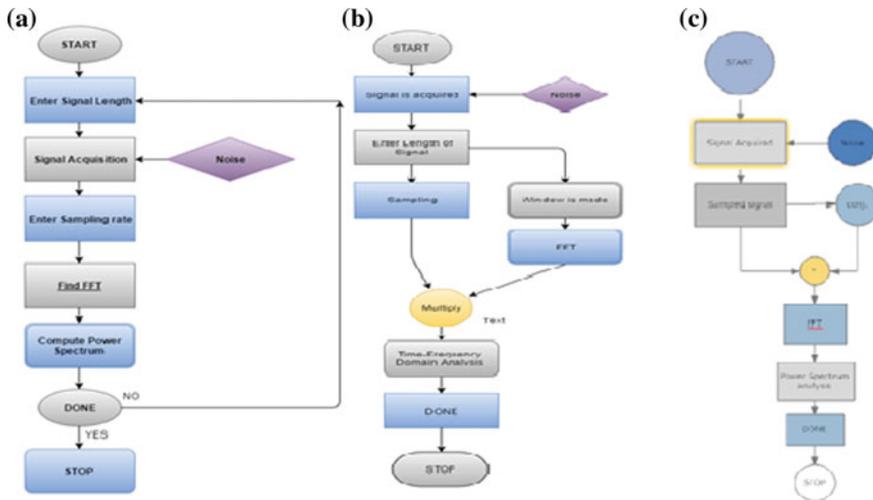


Fig. 2 Flowchart for a FFT, b STFT and c WVD

3 Result and Discussion

The methods used in Doppler extraction are first studied, then designed and further tested in LabVIEW. Here, the work has been divided into two parts: Software and Hardware.

First, the simulations of the methods were simulated using LabVIEW. After verifying the results in simulation, all the three methods have been tested considering actual Doppler-shifted signal.

3.1 Software Simulation Using LabVIEW

The platform allows the virtual implementation of the hardware components on a software platform. First, signal is generated, the generated signal is applied to each of the three methods and following are the results of the same.

3.1.1 For FFT

The signals used are a sine wave at 450 Hz and other acts as an added Doppler signal at 50 Hz which are added and then FFT is implied. This provides the desired result at 50 Hz (Fig. 3).

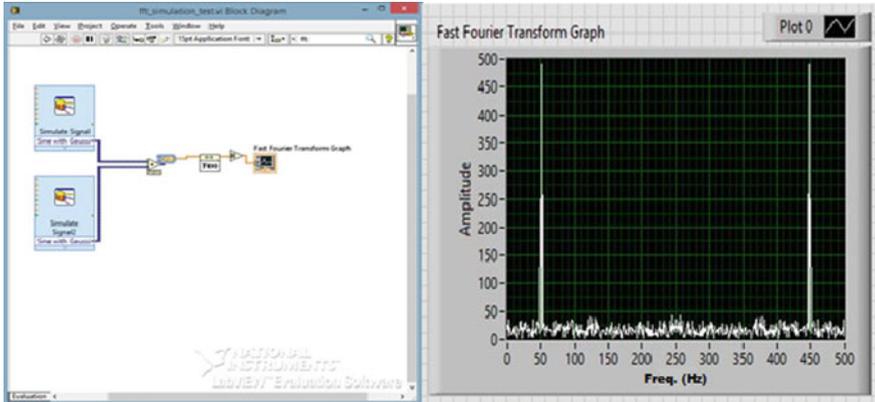


Fig. 3 LabVIEW front panel and block diagram for FFT of test signal

3.1.2 For STFT

Here, the signal under analysis is chirp signal that is signal with variable frequency at variable time. Figure 4 depicts the generated chirp signal also the time–frequency

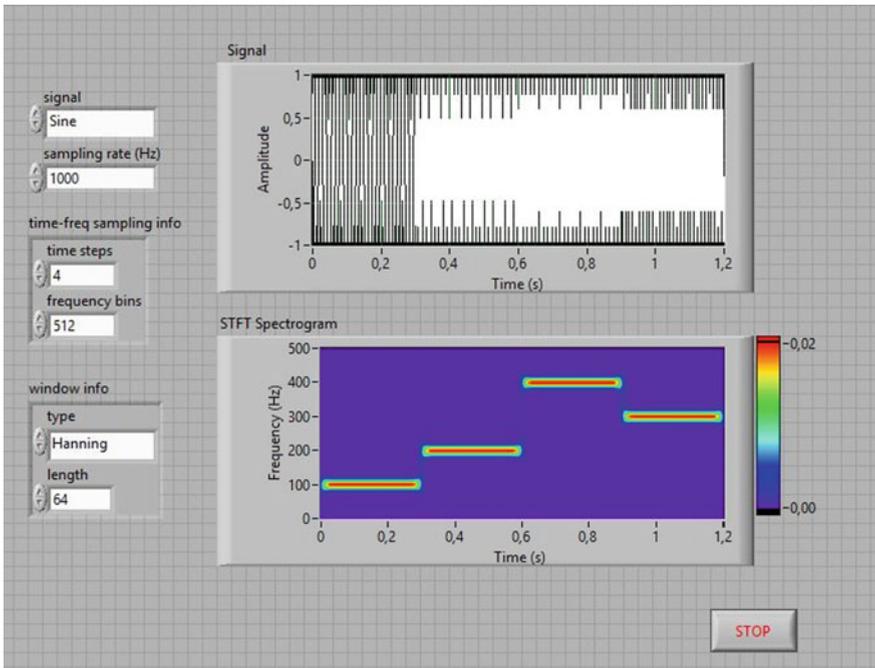


Fig. 4 Front panel of STFT O/P

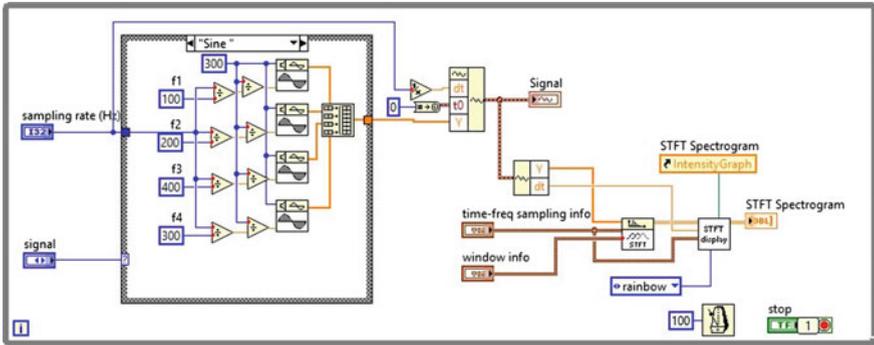


Fig. 5 Block diagram of STFT for chirp signal

analysis of the same signal. Since the signal has variable frequency at different time instances, hence the time–frequency spectrogram of the signal shows different frequencies at different times.

Figure 5 shows the actual block diagram of the system.

3.1.3 WVD

The signals used are a sine wave at 450 Hz and other acts as an added Doppler signal at 50 Hz which are added and then FFT is implied. This provides the desired result at 50 Hz (Fig. 6).

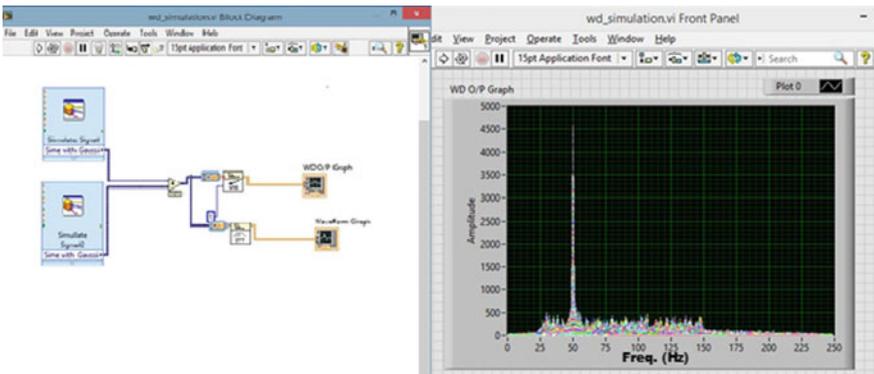


Fig. 6 The WVD output for a sine with Gaussian noise signal

3.2 Hardware Result

The results of the software simulation are considered and then analysed by replacing the test signals with real-time signals from the surroundings and the values are taken (Table 1).

Hardware Setup:

1. First 24 GHz continuous waves are transmitted whose frequency is modified using a rotating fan.
2. Then the Doppler radar signal is received via the antenna and the Doppler is extracted from the same signal.
3. This is done using the Data Acquisition Card in LabVIEW.
4. The actual RPM of the rotating fan is measured using stroboscope (Fig. 7).

Table 1 System parameters

Parameters	Specification	Remarks
RF	24 GHz	Transmitted waves
Doppler range	100–300 Hz	Doppler created using a small rotating fan
Antenna with transceiver	–	Combination of transmitter and receiver in the same package
Data acquisition card	LabVIEW	Converts the incoming analog signals to digital form so that it can be interpreted by the computer
Stroboscope	–	Finds the rotation per minute (RPM) of a signal

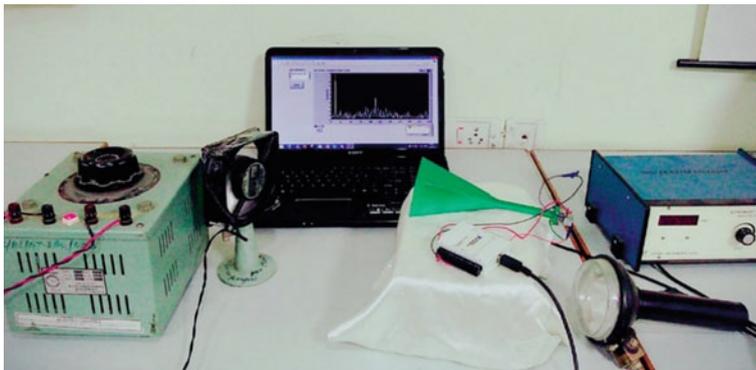


Fig. 7 Hardware setup

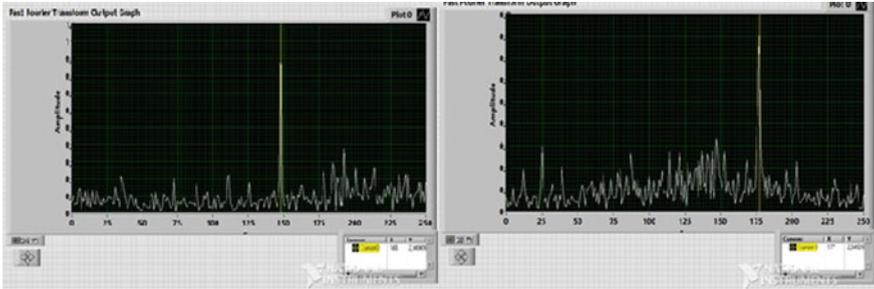


Fig. 8 FFT of real-time signal and faster signal

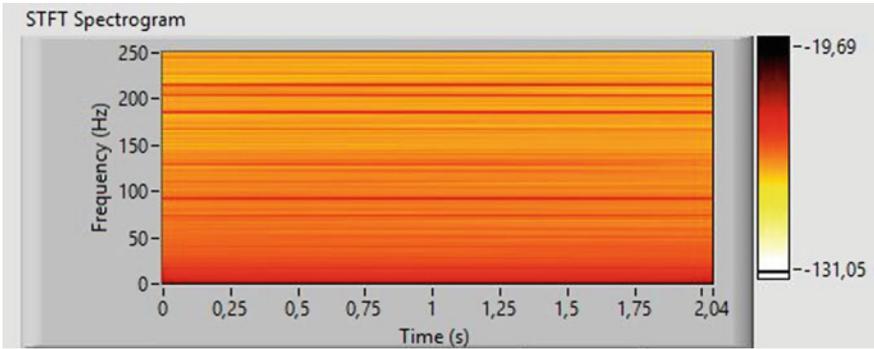


Fig. 9 STFT output in Hanning Window

3.3 FFT Results

As discussed earlier, FFT does not show the change in frequency with respect to its time (Fig. 8). Therefore, STFT and WVD is used to remove this limitation.

3.4 STFT Results

Here the result (Fig. 9) shows that there is a linear change in the frequency of the signal.

3.5 WVD Results

Since WVD has better resolution than STFT and FFT, therefore the graph Fig. 10 is more accurate.

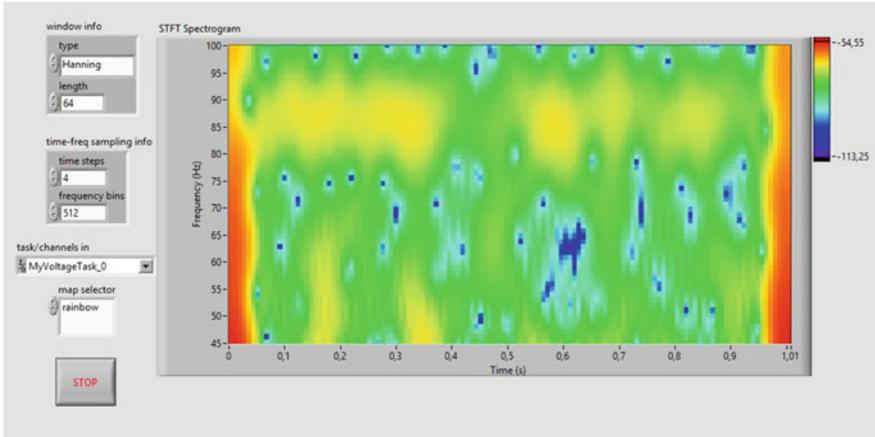


Fig. 10 WVD output in Hanning Window

Table 2 Measured RPM with respective doppler shifts in carrier frequency

Rotations per minute (RPM)	Velocity (m/s)	Doppler shift at 24 GHz C.F (Hz)
591	3.52	283.56
521	3.10	249.72
499	2.977	239.25
482	2.87	231.19
451	2.69	216.69
335	1.99	160.31
284	1.69	136.14

4 Conclusion

This paper presents FFT, STFT and WVD approach for the extraction of Doppler features of radar returned signal from the moving targets. From the results of FFT approach (Table 2), it is observed that as the velocity of the waves increase, the Doppler shift also increases with it. But WVD is preferred as it has more resolution capability than the other two methods. In software, low-range frequency was detected but in hardware due to limitations, low Doppler has not been able to detect but works have been done.

References

1. Heide, P.: Commercial microwave sensor technology—an emerging business. *Microw. J.* **42** (5), 348–352 (1999)
2. Rajeevan, M., Pai, D.S., Anil Kumar, R., Lal, B.: New statistical models for long range forecasting of southwest monsoon rainfall over India. *Clim. Dyn.* **28**(7), 813–828 (2007)
3. Kumari, P., Kumar, A..A.: Doppler weather radar surface rainfall intensity validation study. *Int. J. Environ. Eng. Manage.* **4**(6), 561–566 (2013). ISSN 2231–1319
4. Sirisha, N., Rani, A.D.S., Barua, P., Anuradha, D., Kuloor, R.: FPGA implementation of doppler processing for human heart-beat detection and ranging through a barrier using UWB SFCW approach. In: 9th International Radar Symposium India (2013)
5. Cohen, L.: *Time-Frequency Analysis*. Prentice-Hall, New York, USA (1995)
6. Immoreev, I.I., Fedotov, P.G.S.D.V.: Ultra wideband radar systems: advantages and disadvantages. *Ultra Wideband Syst. Technol.* 201–205 (2002)
7. Okumura, S.: *The Short Time Fourier Transform and Local Signals*. Carnegie Mellon University (2011)
8. Suresh, P., Thayaparan, T., SivaSankaraSai, S., Sridharan, K.S., Venkataramaniah, K: Gabor-wigner transform for micro-doppler analysis P. In: 9th International Radar Symposium India (2013)
9. Chen. V.C., Ling, H.: *Time-Frequency Transform for Radar Imaging and Signal Analysis*. Artech House, Boston (2002)
10. Preis, D., Georgopoulos, V.C.: Wigner distribution representation and analysis of audio signals. *J. Audio Eng. Soc.* **47**(12) (1999)
11. Jones, D., Park, T.: A resolution comparison of several time-frequency representations. *IEEE Trans. Signal Process.* **40**, 413–420 (1992)

Efficient Use of Location of Unlicensed Users in Improving the Utilization of Licensed Spectrum

G. Verma and O.P. Sahu

Abstract The success of the CR systems which are based on energy detection-based scheme depends primarily on the selection of threshold (λ). This λ is selected either by using the CDR principle in which the probability of detection P_d is kept constant or by using the CFAR principle in which the probability of false alarm P_{fa} is kept constant. The approach used in this paper is based on the power received from primary transmitter. To properly utilize the licensed spectrum, schemes used are namely: CDR under opportunistic spectrum access scheme (OSA) and spectrum sharing with sensing scheme. Compared to the approach where the CDR principle is used blindly the throughput under the proposed approach experiences a significant gain.

Keywords Cognitive radio · Energy detection · Distance · Spectrum sensing
Transmission power

1 Introduction

The problem of spectrum scarcity is increasing due to increase in wireless services [1]. But the studies of FCC shows that the spectrum is not utilized properly [2]. Most of the spectrum allocated to PUs remains vacant. This underutilized spectrum can be used by SUs. CR is one such technology that helps in achieving this. It searches for vacant spectrum and allocates it to SUs in a way that the services of PUs are not hindered [3]. Spectrum sensing plays a major role in the proper working of CR. Spectrum sensing is the process by which CR detects the presence of PUs. For spectrum sensing, energy detection scheme is widely popular. Under energy detection, probability of false alarm P_{fa} and probability of detection P_d are

G. Verma (✉) · O.P. Sahu
Department of Electronics and Communication Engineering,
National Institute of Technology Kurukshetra (NITKKR),
Kurukshetra 136119, Haryana, India
e-mail: gauravnitk13@gmail.com

the two major parameters [4, 5]. P_{fa} is probability to falsely detect the signal and it should be low. P_d is the probability to correctly detect the signal and it should be high to ensure a certain quality of service for PU. To access licensed spectrum, SU can use namely two schemes: spectrum sharing (SS) and opportunistic spectrum access (OSA) [4–8]. In case of SS the spectrum is accessed independent of its status, but it should ensure that primary users are not interrupted, while in case of OSA it is accessed only if the status is idle.

The success of the energy detection scheme depends on the value of the threshold selected. It is very important to properly select the threshold to find the status of target band [4, 9, 10]. With a low value of λ , the value of P_{fa} increases which in turn affects the utilization of licensed spectrum. On the other hand, if the value is high then P_d decreases which in turn affects the services of PU. This value of λ is calculated either by CDR principle or by CFAR principle. In [5, 6, 10] authors used CFAR principle while in [6–8] CDR principle was taken under consideration. CDR principle is adopted when the priority is the protection of PUs but sometimes it becomes overprotective in nature and hinders the proper utilization of spectrum. In this paper, the proposed approach utilizes the opportunities missed by the blind use of CDR principle. To accomplish this spectrum sharing is adopted when the distance between transmitter and receiver is large.

The distance between SU and PU has been found by using the signal-processing techniques. Works in [11, 12] also aim to find the distance between transmitter and receiver. In this proposed approach, the worst case scenario is considered, in which the receiver is located on the cell boundary of primary cell. The critical distance d_c is calculated such that the signal-to-interference-plus-noise ratio (SINR) at receiver equals to the value of $SINR_{th}$. $SINR_{th}$ is the minimum value of SINR. After the value of d_c is calculated, at distance $d_{min} = d_c + R$ minimum power P_{min} is calculated. Here, R is the service range of transmitter. If CR detects the power from primary transmitter lesser than P_{min} then distance is assumed to be more than the value of d_{min} . For the case in which $P < P_{min}$ spectrum-sharing scheme is used, otherwise OSA scheme is preferred.

In comparison to the cases where CDR principle was used blindly [4–8], or where both CDR and CFAR principles were adopted the throughput achieved under this approach is significantly improved.

2 Throughput of CR Under the CDR-OSA Scheme

The values of P_{fa} and P_d under the consideration that primary user signal is phase-shift keying and noise signal is circularly symmetric complex Gaussian signal (CSCG) are given by [5–7, 9, 10]:

$$P_{fa} = Q\left(\left(\frac{\lambda}{\sigma_u^2} - 1\right)\sqrt{N}\right) \tag{1}$$

$$P_d = Q\left(\left(\frac{\lambda}{\sigma_u^2} - \text{SNR}_p - 1\right)\sqrt{\frac{N}{2\text{SNR}_p + 1}}\right) \tag{2}$$

Under CDR principle, $P_d(=\overline{P}_d)$ is constant, the values of threshold $\lambda(=\lambda_d)$ and probability of false alarm $P_{fa}(=P_{fa,\text{CDR}})$ are given as

$$\lambda_d = \sigma_u^2\left(\left(Q^{-1}(\overline{P}_d) \cdot \sqrt{\frac{2\text{SNR}_p + 1}{N}}\right) + \text{SNR}_p + 1\right) \tag{3}$$

$$P_{fa,\text{CDR}} = Q\left(\left(Q^{-1}(\overline{P}_d) \cdot \sqrt{\frac{2\text{SNR}_p + 1}{N}} + \text{SNR}_p\right)\sqrt{N}\right) \tag{4}$$

Transmission rates C_{00} and C_{01} for a CR system using frame structure of Fig. 1 are given as [4, 9, 10]:

$$C_{00} = \log_2\left(1 + \frac{P_{rs}}{\sigma_u^2}\right) = \log_2(1 + \text{SNR}_s) \tag{5}$$

$$C_{01} = \log_2\left(1 + \frac{P_{rs}}{\sigma_u^2 + P_{pr}}\right) = \log_2\left(1 + \frac{\text{SNR}_s}{1 + \text{SNR}_p}\right) \tag{6}$$

where C_{00} and C_{01} are transmission rates of CR when target channel is detected as idle under H_0 and H_1 respectively. P_{rs} is power of SU received at its receiving end, and SNR_s is corresponding SNR. In case of CDR-based OSA scheme P_{rs} which uses the fix value of transmission power P_{pre} is given by

$$P_{rs} = c \cdot P_{pre} \cdot \frac{1}{d_{ss}^r} \tag{7}$$



Fig. 1 General frame structure of CR where SS spectrum sensing and DT data transmission

where d_{ss} is the distance between secondary transmitter and receiver, r is path loss exponent and c is a constant term. Total throughput of CR system under CDR-OSA approach is given as:

$$C_{\text{CDR-OSA}} = \left(\frac{T - \tau}{T}\right)P(H_0)(1 - P_{\text{fa,CDR}})C_{00} + \left(\frac{T - \tau}{T}\right)P(H_1)(1 - \overline{P_d})C_{01}. \quad (8)$$

3 The Proposed Approach

The PU network is considered in this paper has one fixed primary transmitter, which is serving as a base station with service range is R (Fig. 2) [6, 8, 10]. The proposed approach is given in Fig. 3 in which SU first computes its distance d_{ps} from primary transmitter by using any of the works as given in [10–12]. d_{ps} is used to calculate the distance d of primary receiver from CR transmitter as:

$$d = d_{ps} - R \quad (9)$$

After the value of distance has been computed, it is compared with critical distance d_c . The critical distance d_c is given in Eq. (12), and is computed as follows:

Under the case of missed detection, CR starts transmitting and corresponding power P_{rsp} received at primary receivers given by

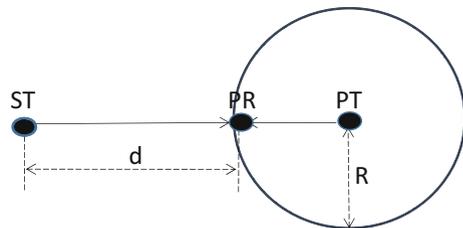
$$P_{\text{rsp}} = \alpha \cdot P_{\text{pre}} \cdot \frac{1}{d^r} \quad (10)$$

where P_{pre} is preset value of SU transmission power, d is closest distance between primary receiver and secondary transmitter (as shown in Fig. 2), α is a constant term and r is path loss exponent.

SINR value in the case when primary receiver is situated at boundary of the cell is given by

$$\text{SINR} = \frac{\beta \cdot P_p \cdot \frac{1}{R^r}}{\alpha \cdot P_{\text{pre}} \cdot \frac{1}{d^r} + \text{noise}} \quad (11)$$

Fig. 2 System model considered for the proposed approach



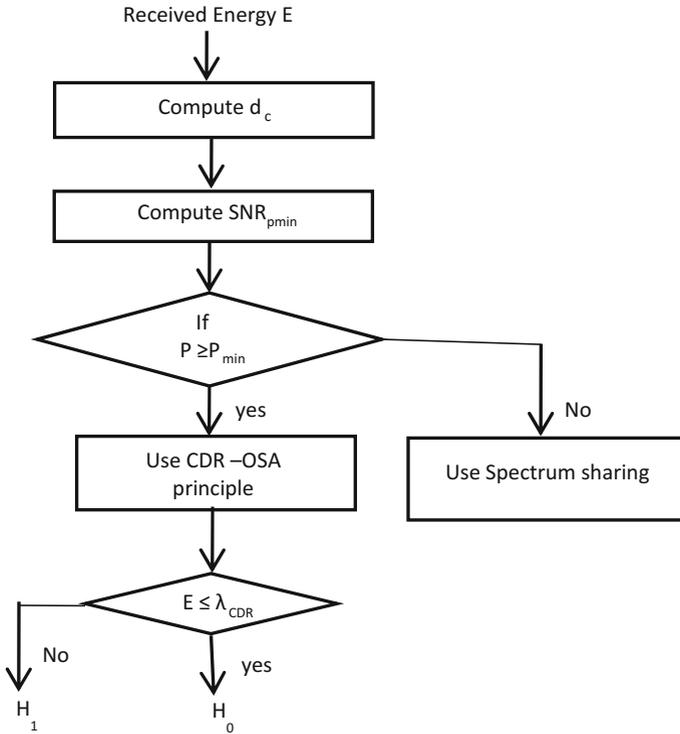


Fig. 3 Flowchart representation of the proposed approach

where the numerator term in Eq. (11) shows the transmitted power P_p of primary transmitter, received at its receiver section.

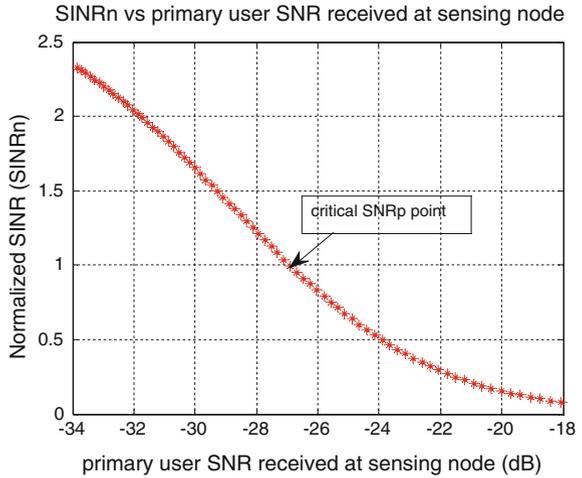
In the case of simultaneous transmissions by PU and SU, primary receiver can decode its own information if SINR in Eq. (11) is greater than or equal to a threshold value ($=SINR_{th}$) or in other words $SINR_n = \frac{SINR}{SINR_{th}} > 1$ only, where $SINR_n$ is the normalized SINR. So, the critical distance d_c corresponding to the case $SINR = SINR_{th}$ is given by Eq. (12) and is also shown through in Fig. 4.

$$d_c = \exp\left(\frac{1}{r} \cdot \ln\left(\frac{\alpha \cdot P_{pre} \cdot SINR_{th}}{\beta \cdot P_p \cdot \frac{1}{R^r} - noise \cdot SINR_{th}}\right)\right) \tag{12}$$

Using critical distance d_c , we find the distance d_{min} and the corresponding power P_{min} as follows:

$$d_{min} = d_c + R \tag{13}$$

Fig. 4 The graph showing the critical primary user SNR_{pmin} where $\text{SINR}_n = 1$



and

$$P_{\min} = c \cdot P_p \cdot \frac{1}{d_{\min}^r} \tag{14}$$

the SNR corresponding to P_{\min} (i.e., SNR_{pmin}) can be written as

$$\text{SNR}_{\text{pmin}} = \frac{P_{\min}}{\sigma_u^2} \tag{15}$$

The proposed approach can be explained much better with the help of flowchart in Fig. 3, as follows: During sensing, when the condition $P \geq P_{\min}$ exists, the licensed spectrum is accessed using the CDR-OSA scheme of spectrum access. In the case $P < P_{\min}$ where $d > d_c$ and PU communication remains protected, the CR communication is performed using the scheme of spectrum sharing (with sensing). The CR throughput under the scheme of spectrum-sharing with sensing is given by

$$C_{\text{CFAR}} = \left(\frac{T - \tau}{T}\right)P(H_0)C_{00} + \left(\frac{T - \tau}{T}\right)P(H_1)C_{01} \tag{16}$$

The spectrum-sharing scheme is used after ensuring that P is less than P_{\min} . The advantages of proposed approach as compared to conventional CDR-based OSA scheme are as follows:

- (1) Proposed approach does not require spectrum sensing during each frame of CR when $d > d_c$, because at this distance SINR received is greater than SINR_{th} and protection of PU remains no more a primary concern. Therefore, full frame duration (T) is used for transmission due to which throughput of CR increases.

- (2) When the distance d between receiver and transmitter is large, CR can communicate with higher power without causing any interference to PU. In the proposed approach, CR adopts its transmission power corresponding to distance d and transmits with high power when distance d is more which is not possible in CDR-based OSA scheme.
- (3) In the condition when $P < P_{\min}$, the CR remains free from the problem of frequent channel hand-offs.

4 Results and Analysis

For simulations the parameters are assumed as follows. Total frame duration $T = 100$ ms, time required for sensing $\tau = 14.2$ ms, $P(H_1) = 0.2$, path loss exponent $r = 4.2$, frequency of the target channel is 60 MHz, PUs transmission power P_p is 1 W, SU's transmission power $P_{ts} = 0.1$ W, gain of transmitting and receiving antenna is unity, $\text{SINR}_{\text{th}} = 8$ dB which is the minimum value of SINR required to demodulate the binary phase-shift keying (BPSK) signals having bit error rate of $2e - 4$ and $\text{SNR}_s = 20$ dB. Figure 5 plots the variations in SUs throughput with respect to SNR of PU received at the secondary transmitter node. For $\text{SNR}_p \leq \text{SNR}_c$, the throughput of SU is same for the proposed approach and also to the conventional approach. But for $\text{SNR}_p > \text{SNR}_c$, the proposed approach which opportunistically adopts spectrum-sharing scheme gives better throughput than the conventional approach blindly following the CDR principle.

Fig. 5 The graph showing comparison of the throughput of the proposed approach with that of the approach under the blind use of CDR principle

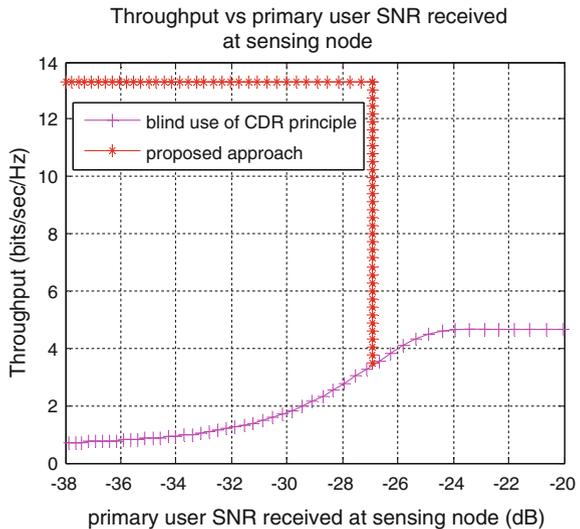


Fig. 6 Throughput of the proposed CR under the various values of path loss exponent r

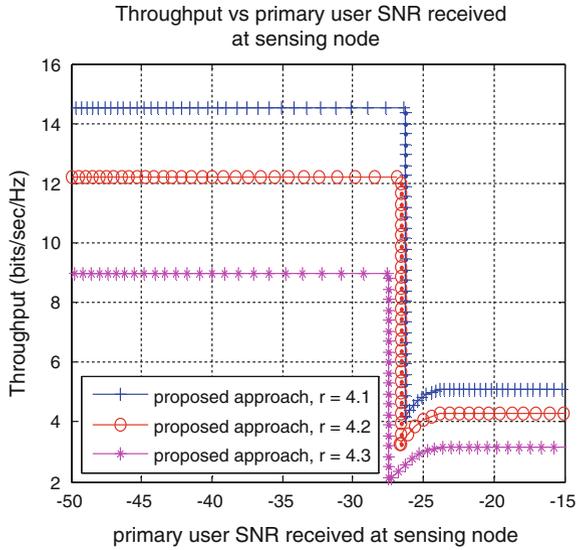


Figure 6 plots the throughput of SU under the proposed approach with respect to SNR_p , for different values of r . As the value of r increases, path loss between PU receiver and CR transmitter increases. Due to increase in path loss, interference to the primary receiver decreases and creates more transmission opportunities for SU. Therefore as r increases, throughput of SU also increases (as shown in Fig. 6).

5 Conclusion

In the proposed approach, advantages of both OSA scheme and spectrum-sharing scheme are taken. The proposed approach is more efficient than the techniques in which CDR principle is used blindly or in which both CDR and CFAR principles are used. This approach proves efficient results when working in low SNR region also.

References

1. Mitola, J.: Cognitive radio: an integrated agent architecture for software defined radio. Ph.D. Dissertation, Royal Inst. Technology, Stockholm, Sweden (2000)
2. Verma, G., Sahu, O.P.: Intelligent selection of threshold in cognitive radio system. *Telecommun. Syst.* 1–10 (2016)
3. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Sel. Areas Commun.* **23**, 201–220 (2005)

4. Liang, Y.C., Edward Peh C.Y.: Sensing throughput tradeoff for cognitive radio networks. *IEEE Trans. Wirel. Commun.* **7**(4), 1326–1337 (2008)
5. Wang, N., Zhang, X., Gao, Y.: Adaptive spectrum sensing algorithm under different primary user utilizations. *IEEE Commun. Lett.* **17**(9), 1838–1841 (2013)
6. Choi, H.-H., Jang, K., Cheong, Y.: Adaptive sensing threshold control based on transmission power in cognitive radio systems. In: *IEEE Conference of Cognitive Radio Oriented Wireless Networks and Communication (CrownCom)*, pp. 1–6 (2008)
7. Verma, G., Sahu, O.P.: Interference aware optimization of throughput in cognitive radio system. *Def. Sci. J.* 1–6 (2015)
8. Lin, Y.-E., Liu, K.-H., Hsieh, H.-Y.: On using interference-aware spectrum sensing for dynamic spectrum access in cognitive radio networks. *IEEE Trans. Mob. Comput.* **12**(3), 461–474 (2013)
9. Verma, G., Sahu, O.P.: Opportunistic selection of threshold in cognitive radio system. *J. Wirel. Pers. Commun.* 1–16 (2016)
10. Verma, G., Sahu, O.P.: Intelligent selection of threshold in cognitive radio system. *J. Telecommun. Syst.* 1–10 (2016)
11. Celebi, H., Arslan, H.: Enabling location and environment awareness in cognitive radios. *J. Comput. Commun.* **31**, 1114–1125 (2008)
12. Radhi, N., Kahtan, A., Sofian, H.: Estimate primary user localization using cognitive radio networks. In: *IEEE International Conference on Innovations in Information Technology (IIT)*, pp. 381–385 (2011)

Parking Places Discovery and Reservation Using Vehicular Ad Hoc Networks

Alak Roy, Jayati Paul, Ratnadeep Baidya and Mampi Devi

Abstract In today's digital era, to discover a parking places in urban city is a very challenging task due to rapid increases in the number of vehicles. In this situation, Vehicular Ad Hoc Networks (VANETs) can help a vehicle to communicate with each other and with road infrastructure. This paper mainly focuses on the on-board unit (OBU)–road side unit (RSU) communication in VANETs. In this paper, we proposed an approach for parking places discovery and reservation for the parking in an urban city. Also, we have implemented a small part of our proposed approach using Network Simulators-2 and present the performance of different routing protocols such as AODV, DSDV, and DSR based on different parameters. Here, we used IEEE 802.11 MAC protocol for simulation results.

Keywords Parking places discovery and reservation · VANETs Architecture · OBU–RSU communication

1 Introduction

In early days, to avoid traffic congestion, drivers use their voice, horns, observation of each other's trajectory, and gestures. But as the vehicles are increasing day by day, it is not possible to control the traffic by the traditional methods. Thus a new kind of

A. Roy (✉) · J. Paul · R. Baidya
Department of Information Technology, Tripura University,
Suryamaninagar, Agartala, India
e-mail: alakroy@tripurauniv.in

J. Paul
e-mail: jayati.260@gmail.com

R. Baidya
e-mail: sayan4mail@gmail.com

M. Devi
Department of Computer Science and Engineering,
Tezpur University, Tezpur 784028, Assam, India
e-mail: mampi.mca08@gmail.com

MANET's wireless ad hoc network is introduced that allows vehicles to communicate with each other with roadside infrastructure known as Vehicular Ad Hoc Networks (VANETs) [1–3]. This network provides benefits to billions of users around the world. It has enormous amount of applications to improve vehicle safety and road safety. It not only gives comfort to the passenger and driver but also provides a communication tool between the user and the road infrastructure [2]. Various communications of VANETs include vehicle-to-vehicle, vehicle-to-roadside, and vehicle-to-infrastructure communication. VANETs address all the issues related to the wireless communications that supports more customized and complete information to be exchanged [4, 5]. As compared to other types of MANETs, VANETs have unique characteristics, such as high node mobility, fast topology change, variable network density, high computational ability, and geographic position availability.

In this paper, we proposed an approach about parking place discovery and reservation problem. As the number of vehicles is increasing drastically, so, to avoid accidents, jams and to improve road safety, the drivers need a space where they can park their car after reaching their destination. If they have the up to date information about the traffic situation and the available parking spots around destination then this will decrease traffic jams. Otherwise, searching for available parking places require much waiting time and wastage of diesel, gasoline and incur great economical losses. For example, in Munich, Germany, searching for a free parking spot causes the following problems: economical damage of total 20 million Euros, 170,000 h of waiting time, every second a vehicle is searching for an available parking spot amounts 45% of the entire traffic area, gasoline, and diesels are wasted for the searching of free parking space. That is why the discovery of parking places is a most important issue now a day.

The rest of this paper is organized as follows. Section 2 provides an overview of the proposed approach for parking place discovery and reservation. The simulation studies has been carried out in Sect. 3 to implement a part of the proposed approach and to evaluate the performance of AODV, DSDV, and DSR routing protocols considering packet delivery ratio and end-to-end delay. Finally, Sect. 4 concludes the paper with possible research directions.

2 Proposed Approach

In this section of this paper, we discuss about the discovery and reservation of a parking place in municipal areas. In a municipal traffic situation, searching for a free parking place is a great challenge for millions of city dwellers every day. For example, let someone is driving into a crowded city area. He is going to cinema hall to see a movie. He wants to park his car near his destination. If there is an efficient parking guidance service for the drivers then the searching time and cost of the parking spaces would be greatly reduced and as a result to some extent the traffic congestion would also be relieved. So, the discovery of parking places is necessary part of our daily life.

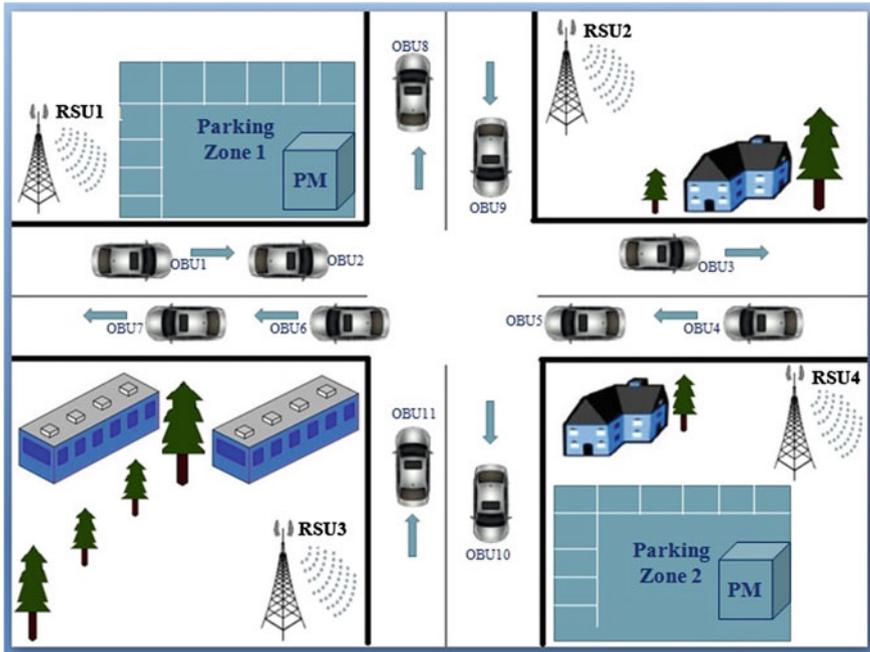


Fig. 1 A scenario for parking place discovery and reservation

In the recent past, several existing approaches are available [6–10]. However our approach is differ from the existing approaches like we take care about discovery of reservation at a time and discuss about how all the blocks work communicate with each other. Furthermore, this approach also overcomes some problem related to parking such as congestion, competition between the vehicles and duplication of messages. The most important applications of parking places are in the high traffic areas for example municipality area. In municipal areas there are huge numbers of vehicles moving every second. So the people have to suffer from road jams everyday. Therefore, to decrease road accidents, jams, and to improve road safety, user safety, parking place discovery of parking place is a necessary task in these areas. The steps for discovery of parking place and reservation for this place is explained in the next subsection.

2.1 Parking Place Discovery

In this section, we propose an approach for discovering parking places. Here, we consider a network that consists of parking meters (PM), road side units (RSU), and vehicles (OBU) as shown in Fig. 1. Considering the RSUs are deployed in an area

and each of the RSU covers a smaller area. There are many parking meters are scattered in this area. The RSUs are gathering the parking information from the PM by communicating with the PMs in its range from the nearby intersecting RSUs, and stores the information in its parking profile. Since there is a broadcast interval, each RSUs will send the current parking status of the area to all the OBUs in its communication range. The vehicle that lies within that range will get the information. At the same time they can query for suitable place where they want park. In this way, parking necessity people can directly communicate with the PM via RSU and can reserve a parking slot. The flowchart for discovery of parking place is shown in Fig. 2.

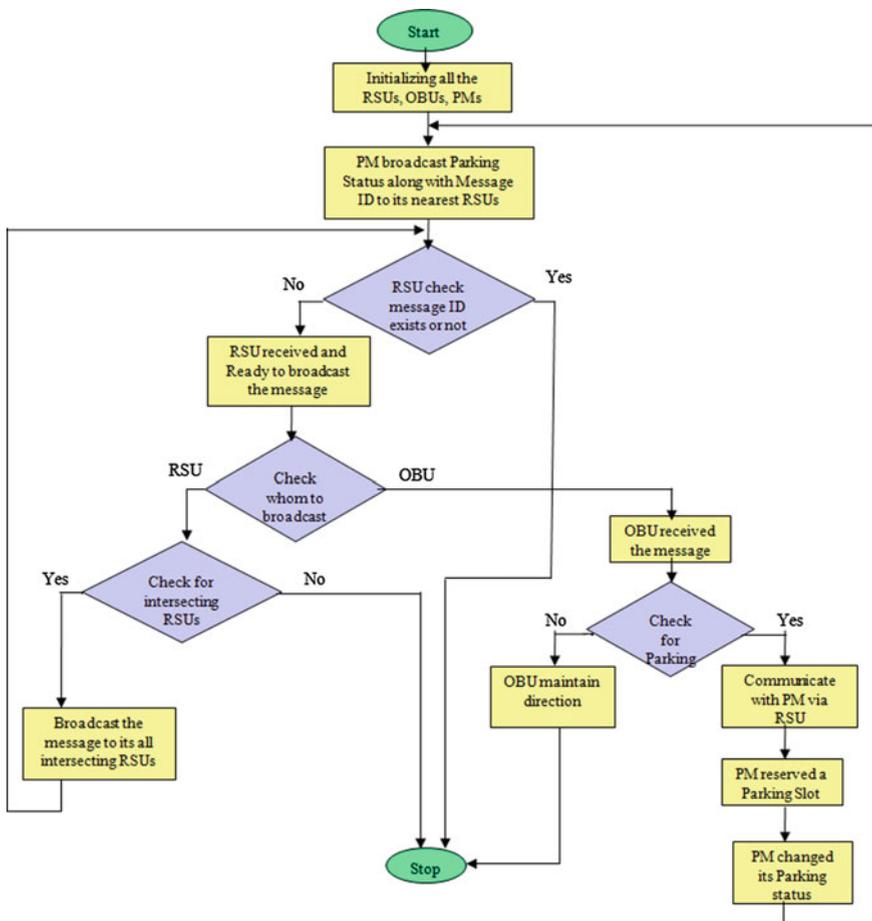


Fig. 2 Communication flowchart for parking place discovery

2.2 Parking Place Reservation

In this section, we discuss about reservation of parking spots along with flowchart shown in Fig. 3 and algorithm. The limitation of this process is that vehicles should place within certain limited range. Otherwise, it may be of hazardous situations for parking of other vehicles. The main aim of this approach is to give the systematic way to quickly identify the interested vehicles which are very close to the parking zone for reservation. Thus, all the RSU periodically broadcasts regarding

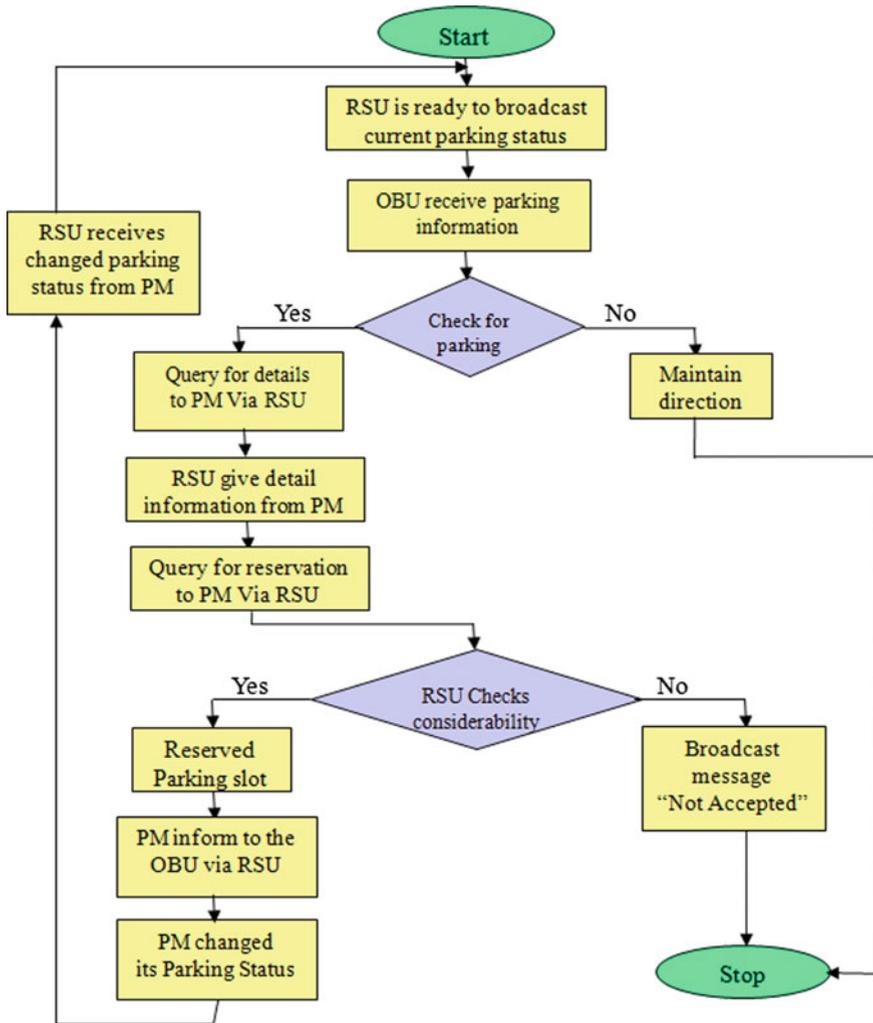


Fig. 3 Flowchart for parking place reservation

availability of the PM so that the nearest OBUs could get the information. This allows each OBU to communicate nearby PM via RSU for reserving a free parking space. After receiving the confirmation message from OBU, the OBU is ask to give the detail about parking information to the PM via nearest RSU. If several OBU can communicate with the same RSU at the same time then the RSU gets the advantages to sort out the number of vehicles for reservation in the way of FCFS. The RSU matches the OBU Block and the parking profile, if a match then RSU communicates with the PM to reserve the spot. Then PM reserved a space for the particular vehicle. Otherwise the RSU broadcast a message “Not Accepted”. So, after reserving the parking slot for that vehicle, PM changes its parking space record. The again PM broadcasts its updated parking status to its nearby RSUs and RSU rebroadcast new parking status to all the RSUs, OBUs in its range. In case no vehicle replies, the RSU repeatedly broadcast the message.

3 Experimental Evaluation

This section presents a simulation study of the proposed approach using NS-2 [11] with network topology depicted in Fig. 4. Moreover, it also presents a simulation study of AODV, DSR, and DSDV routing protocols [4, 5] with respect to IEEE 802.11 MAC protocol [3]. The main of this task to select the best routing protocol that performs best and is applicable for implementation of our approach. For the simulation study, we use IEEE 802.11 as the MAC protocol because of its most popular and widely used basic standard among the various MAC protocols available for VANETs [3]. Here, we consider that the vehicular traffic model is based on the amount of traffic planning and patterns in the time from 7:00 am to 11:00 am. In the simulation study, we will determine how many vehicles are searching for a free parking spot over the time. The simulation starts after setting all parameters.

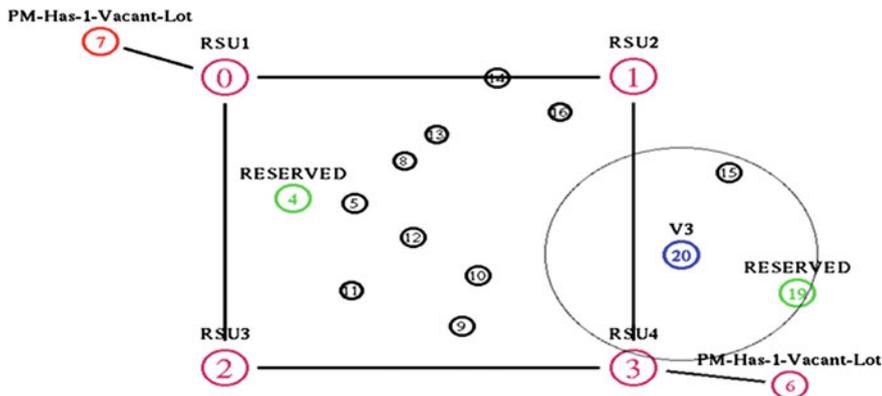


Fig. 4 A snapshot of the parking discovery and reservation simulation using NS-2

Then, it is possible to see the graphical simulated output along with movements of all nodes. In the simulation process, it also generate trace file to analyze output further.

For simulation study, a network topology of size approximately $1800 \times 900 \text{ m}^2$ as shown in Fig. 4 is considered. The average dimensions are 900 m^2 in north-south and 1800 m^2 in east-west. In this topology, considering two parking meters (PMs) and four road side units (RSUs) as fixed nodes, and thirteen vehicles known as on board units (OBUs) moving into the area. The four RSUs are RSU1 (node 0), RSU2 (node 1), RSU3 (node 2), and RSU4 (node 3) as shown in the Fig. 4 with maroon color. The PMs (node 6 and node 7) are shown as red color, and nodes 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 are moving vehicles and RSUs, OBUs, and PMs are represent by circle as depicted in Fig. 4. The radio transmission range of 300 m is considered among these nodes. Keeping this topology as fixed, apply three different routing protocols at different data rates. Here, TCP connections are used for RSU to RSU communication as they are fixed nodes and UDP connection for RSU to OBU communication as OBUs are mobile nodes. At time 0.0 the simulation starts and stops at 7 ms. From 0.0 to 0.5 ms the PM node at node 7 sends the current parking status to the RSU1. Figure 4 shows the snapshots of the simulation which represent the PM and RSU communication and reservation. After receiving the parking info from PM the RSU1 send this status to other intersecting RSUs which are RSU2 and RSU3. It sends the parking info to RSU2 from 0.6 to 0.8 ms. Then it sends to RSU3 from 1.1 to 1.5 ms. At 1.6 ms the RSU2 send again the same parking info to RSU4 up to 2.0 ms. At 2.0 ms, RSU1 send the parking info to an OBU (node 4) which lies in its range. After receiving the parking info, the OBU 4 will communicate with the PM via RSU1 for reservation. The RSU4 then informs PM to reserve a spot. At 3.0 ms the PM has became reserved for the OBU.

On the other way, to analyze the performance of the AODV, DSR, and DSDV routing protocols in terms of packet delivery ratio (PDR) and end-to-end delay (E2E), we use the simulation parameters as available in [12]. Here, the term packet delivery ratio (PDR) is defined as the ratio of total number of delivered data packets to the destination, and end-to-end delay is the average time for a data packet to

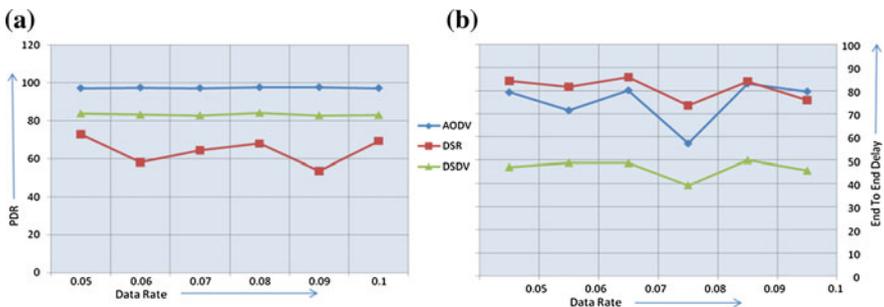


Fig. 5 a PDR and b delay analysis of routing protocols to be used for reservation

arrive in the destination. In case of PDR, we found that AODV gives highest PDR among AODV, DSDV, and DSR at different data rates as shown in Fig. 5a. In case of E2E delay, AODV gives average E2E delay and DSDV gives lowest E2E delay at different data rates as shown in Fig. 5b. Hence, after analyzing the performance it can be concluded that AODV is more applicable for implementing the proposed approach since it has the capacity of fast delivery of packets among other routing protocols. In this paper, it also shows a part of simulation study of the proposed approach, where we can say AODV performs better among the others.

4 Conclusion

In this paper, an approach for parking place discovery and reservation is proposed to reduce road jams and accidents. When the proposed approach is coupled with GPS receiver, it became helpful for drivers to find a parking place. In this approach, IEEE 802.11 as well as IEEE 802.11p MAC protocol can be used. This paper also analyzed the PDR and delay of three routing protocols, viz, AODV, DSR and DSDV at different data rates through simulation studies. The simulation results showed that AODV performs better than DSR and DSDV for parking place discovery and reservation. Implementation of the proposed approach using AODV and IEEE 802.11 MAC is a challenging task which is left as our future work.

References

1. Cunha, F., Villas, L., Boukerche, A., Maia, G., Viana, A., Mini, R.A., Loureiro, A.A.: Data communication in VANETs: protocols, applications and challenges. *Ad Hoc Netw.* **44**, 90–103 (2016)
2. Schoch, E., Kargl, F., Weber, M.: Communication patterns in VANET. *IEEE Commun. Mech. IEEE* **11**(46), 119–125 (2008)
3. Menouar, H., Filali, F.: A survey of qualitative analysis of MAC protocols for vehicular Ad Hoc networks. *IEEE Wirel. Commun. IEEE* **13**(15), 30–35 (2006)
4. Jaap, S., Bechlar, M., Wolf, L.: Evaluation of routing protocols for vehicular Ad Hoc networks in typical road traffic scenarios. In: *Proceedings of the 11th EUNICE Open European Summer School on Networked Applications*, vol. 23, pp. 584–602. Braunschweig, Germany (2005)
5. Kim, W., Gerla, M.: Routing protocols for cognitive vehicular Ad Hoc networks. *Cogn. Veh. Netw.* **78** (2016)
6. Majidi, A., Polat, H.: Finding a best parking place using exponential smoothing and cloud system in a metropolitan area. In: *Smart Grid Congress and Fair (ICSG), 2016 4th International Istanbul*, pp. 1–5. IEEE (2016)
7. Alhammad, A., Siewe, F., Al-Bayatti, A.H.: An infostation-based context-aware on street parking system. *IEEE* **978**(1), 4673–5157 (2013)
8. Caliskan, M., Graupner, D., Mauve, M.: Decentralized discovery of free parking places. In: *Proceedings of the 3rd international workshop on Vehicular ad hoc networks*, pp. 30–39. ACM (2006)

9. Caliskan, M., Barthels, A., Scheuermann, B., Mauve, M.: Predicting parking lot occupancy in vehicular Ad Hoc networks. In: Vehicular Technology Conference, VTC2007, pp. 277–281. IEEE (2007)
10. Chang, G.Y., Sheu, J.P., Chung, C.Y.: Zooming: a zoom-based approach for parking space availability in VANET. In: Vehicular Technology Conference, VTC2010, pp. 1–5. IEEE (2010)
11. The Network Simulator: ns-2. www.isi.edu/nsnam/ns/. Accessed 01 Oct 2016
12. Mamatha, B.L., Raju, C.S., Kantharaju, T.: A cooperative opportunistic routing in VANET using PSR. *Int. J. Res. Eng. Adv. Technol.* **2**(2) (2014)

Design of 5.8 GHz Rectenna for Space-Based Solar Power

Deepak Kumar and Kalpana Chaudhary

Abstract In this work, a simple and inexpensive microstrip rectenna is designed and tested. The rectenna circuit consists of microstrip patch antenna array resonate at 5.8 GHz and Rectifier Schottky diodes. An HSMS-8202 diode has been investigated for the antenna and rectifier circuit impedance matching and rectification efficiency. Further Dickson charge pump voltage multiplier has employed in the rectifier circuit to increase output voltage for small microwave input power condition. The charge pump methodology with different stages is tested and analyzed performance wise.

Keywords Microwave · Power · Rectenna · Voltage multiplier
Space-based solar power

1 Introduction

There is a populace increment around the world. Present day progress is more rely on upon the innovation hardware/software apparatus that requires higher electrical vitality generation [1]. The vast majority of the power plants on the planet are fossils fuel based. It has issues like a worldwide temperature alteration and their restricted assets for merely a few of more decades. In this manner, there is a necessity of environmentally friendly power vitality assets. Therefore, the broad association of renewable vitality assets required for higher electrical power generation to fulfill global energy demand [1, 2]. Sun-powered power is richly accessible on the earth as heat and light. Sunlight-based photovoltaic are to changes in sun-oriented irradiance into electrical vitality. In any case, Sun-based irradiance

D. Kumar (✉) · K. Chaudhary
Department of Electrical Engineering, Indian Institute of Technology BHU,
Varanasi 221005, India
e-mail: Deepak.rs.eee14@iitbhu.ac.in

K. Chaudhary
e-mail: Kchaudhary.eee@iitbhu.ac.in

shifts self-assertively in the day time and it goes terminated during the night. In this manner, Sun-based photovoltaic are not reasonable for base load power [2]. P. Glaser proposed space-based solar power (SBSP) strategy in 1968 [3]. In a space-based solar power station, space satellite gathers sunlight-based vitality and photovoltaic proselytes it into electrical vitality. The electrical power further transforms into the microwave and radiating that microwave power remotely to rectenna on the Earth [2–4]. The antennas-associated rectifiers change over microwave power back to electrical power. Like this, power is accessible to supply grid power after reasonable handling. In this way, it is reasonable to supplant customary wellsprings of power generation sources. SBSP has a few advantages over earthbound sunlight-based power because of unrestricted and unvarnished Sun-powered irradiance accessible in space. The SBSP has more than three-overlay increments in power production limit over the earth-based solar power system. Though, SBSP has some innovation difficulties and there are major economic issues for practical implementation [4].

2 Space-Based Solar Power Global Research

Numerous space exploration associations have studied SBSP and researchers are taking a shot at the improvement of its acknowledgment strategies. Remote power exchange for such substantial separation is a sturdy assessment that necessities exploratory execution. In the early stage, National Aeronautics and Space Administration (NASA) and Department of Energy (DOE) mutually recommended an SBSP reference system. Although in 1997, the Fresh-Look-SBSP presented an enhanced space-based solar power model [4]. The suggested “Sun Tower” SBSP demonstrate has various spearheading thoughts that diminish SBSP comprehension operation and also the implementation cost [4]. The Japan Aerospace Exploration Agency (JAXA) is concentrating on the SBSP calculated and specialized possibility at various technology levels. For the space solar energy, it is conceivable to beam space power down to the Earth utilizing either microwave or laser technique. The microwave strategy is gaining particularly fast growth, yet the optical strategies perpetually have climate related issues. In 2001, a 5.8 GHz 1 GW SBSP JAXA model utilized microwave transmission [5]. A variety of designs not quite the same as the NASA-DOE model have been technically amended, practically assessed, and recommended. It has also proposed that sunlight-based gathering and microwave transmission ought to be performed on the same Sun facing surface. This way the rear is freely available for heat dissipation. Radiation movement (Sun-oriented power gathering and microwave transmission) happens on one side, and undesirable heat would be dissipated from the opposite side. So the powered Sun cells and transmission antenna are all on the same surface, that is, adjusted one next to the other [4, 5].

3 System Architecture

In the SBSP space fragment, mostly microwave innovation is included. The ground section is rectenna and power processing unit. For the successful SBSP as illustrated in Fig. 1, a high directivity and high gain antenna are imperative. On realizing appropriate antenna design, one can outline profoundly order directivity. The microwave beam density on the rectenna has a constraint. As per United States Occupational Safety and Health Act (OSHA), work environment presentation limits for the microwave is 10 mW/cm^2 [6]. The microwave beam density restriction chooses the rectenna estimate for the required electrical power generation. The rectenna proficiency is an essential consideration toward fruitful remote power exchange [7]. In this way, SBSP requires profoundly productive rectenna arrays. In small power rectenna, losses are due to threshold and breakdown voltage and device parasitic. Diode nonlinearity produces harmonics generation. These harmonics can reradiate antenna bringing about abrasion. Proper impedance matching requires among antenna and rectifier circuit for the most extreme power exchange [7]. For microwave frequency in GHz, determination of diode is an extreme errand.

This work is a design and performance analysis of a 5.8 GHz rectenna framework as shown in Fig. 2, considering space-based solar power system (SBSP). The simulation and rectenna outline prepare was performed in Agilent ADS programming (ADS) [8, 9]. The design procedure utilizes ADS tool compartments, microstrip line transform tool, Smith graph for impedance coordinating, and filter design. The antenna array design utilizes S11 simulation and EM full wave simulation in ADS. Rectenna execution investigation utilizes Harmonic Balance method. The rectifier execution investigation and Dickson charge pump topologies of various stages are attempted and examined [8, 9].

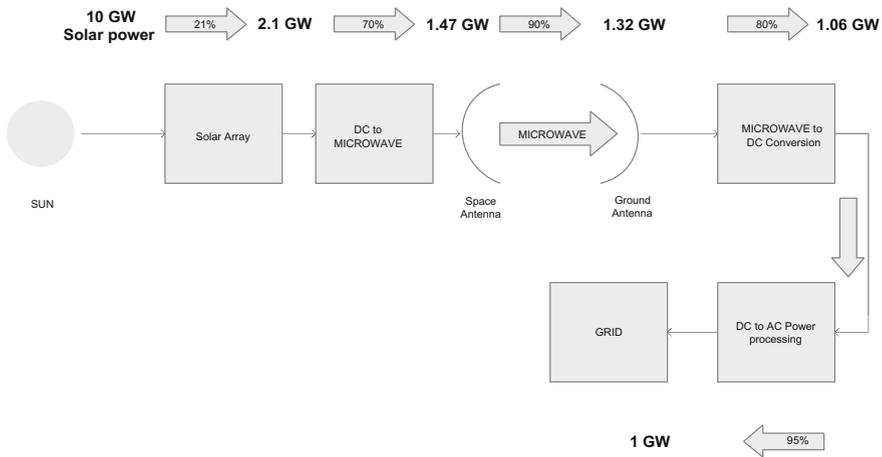


Fig. 1 Space-based solar power diagram [2]

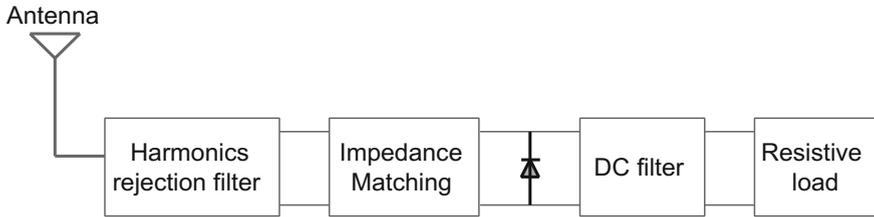
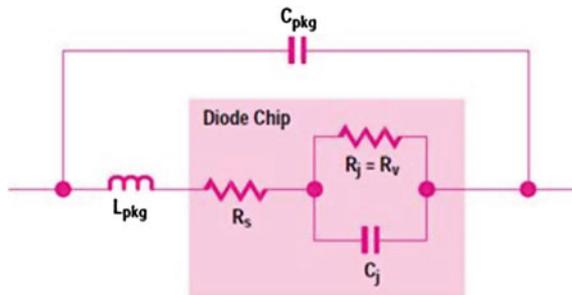


Fig. 2 Rectenna circuit components [3]

Fig. 3 Schottky diode linear model [5]



4 Rectenna Design

4.1 Rectifiers

For higher frequency operation, i.e., 5.8 GHz application Schottky diode is used for converting AC to DC. Schottky diodes have confined by plating a metal contact on n or p-sort semiconductor. This metal semiconductor convergence goes about as a diode with lower forward voltage drop. Little crossing point capacitances of a Schottky diode make it speedier trading among coordinating and non-driving mode [8]. The Schottky diode linear model is presented in Fig. 3. In this work, an HSMS-8202 surface mount microwave Schottky Mixer Diodes made by Avago Technologies is used.

4.2 Harmonics Rejection Filter

An input filter is a low-pass filter, which is required in the rectenna system to prevent harmonics. These harmonics reflection can reradiate antenna and cause losses. Harmonics rejection filters must be capable of transmitting radiation frequency microwave power toward the rectifier circuit [8]. Any filter design can adopt with ease of fabrication and minimal cost. Maximally, flat filter designs are frequently used due to low-order design and thus easy for microstrip conversion [10].

Table 1 Rectifier input impedance values for fundamental and harmonics

Frequency (GHz)	Input impedance
5.8	152.372/116.071
11.6	50.000/-2.096E-14
17.4	50.000/-2.188E-14
23.2	50.000/2.769E-15
29	50.000/3.677E-14

4.3 Impedance Matching

For outlining rectenna framework, it is crucial to decide input impedance of rectifier circuit. The information impedance of the rectifier is characterized in ADS software using Harmonic Balance simulation for order 5. A 5.8 GHz power source with 50 Ω interior impedance has utilized. The input impedance of rectifier has found $-67.41 + j131.12$ as illustrated in Table 1. Input impedances of harmonics are not to mind as it will sift through utilizing low-pass filter among antenna and rectifier circuit.

The discovered estimation of rectifier's information impedance must match to 50 Ω considering 50 Ω microstrip lines at antenna yield. For impedance matching Smith chart tool is utilized in ADS filter design guide toolbox. Microstrip open stub and line length has utilized for the coordinating reason for the simplicity of manufacture. Standardized esteem as Normalized value has chosen in Smith chart for unproblematic impedance matching [11].

5 Voltage Multiplier Rectifier Circuit

For low-power rectenna design, where required yield voltage is high, voltage multiplier is the undeniable decision. Voltage multipliers are charge pump, for rectifier's non-directing period it charges input capacitor, though in leading time the first-charged capacitor supplies voltage to the output load. This work utilizes multi-organized Dickson charge pump rectifier topologies. Dickson n stages charge pump approach as exhibited in Fig. 4 [9, 10].

Here we utilize Harmonic balance simulation tool in ADS for rectenna circuit assessment and performance analysis. A single tone frequency power source with impedance 50 Ω has utilized [10].

The Schottky diode HSMS 8202 ADS simulation model is accessible in ADS library under high-frequency diodes. The rectenna circuit is nonlinear in nature, and therefore either time domain or frequency domain analysis is possible. In the time domain, the recognized answer for the solving nonlinear circuit is large to small signal analysis methods. In the frequency domain, harmonic balance solving is the prescribed solution. The simulation setup like AC simulation and DC simulation is

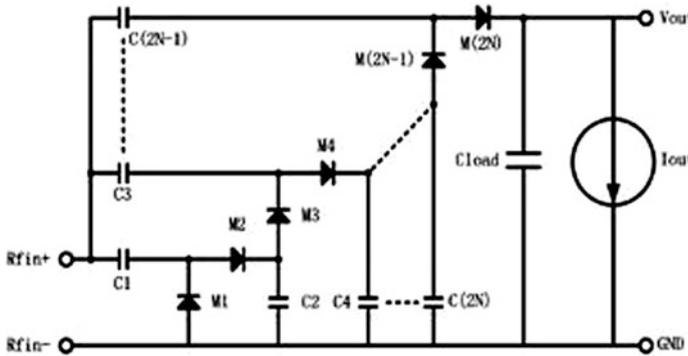


Fig. 4 Voltage multiplier circuit [9]

not appropriate for high-frequency operation of the nonlinear circuit. Therefore, rectenna circuit has simulated in harmonic balance simulation environment in ADS [8–12].

6 Results and Discussion

The circuit is simulated and tested for different capacitor values in the range 100 nf to μf using sweep parameter setup in ADS. The variation in capacitor values has minimal impact on circuit performance and result. The parameter sweep has applied to vary load from few ohm to kilo ohms. The rectifier's voltage profile and efficiency improve at high load resistance. At a fixed load of 2-k Ω , output voltage profile for varying input power as presented in Fig. 5. The input power ranged from -20 to 40 dBm. At low input power less than 0 dBm, the output voltage is very low, and charge pump does not work. As the input power increases, the rectifier's output voltage profile improves. With the growth of charge pump stages, the output voltage reaches a higher value. Within 20 dBm, out DC voltage follow square law slightly masked by noises and offset. At higher dBm value, the voltage profile slope depends on operating frequency, diode junction capacitance, and load resistance. The rectenna efficiency is more than 70% achieved. The rectenna efficiency profile has shown in Fig. 6. At low dBm the efficiency is low. The efficiency plot has a maximum at 10 dBm around. It is expected that the rectenna circuit matched for this power level. Further increase in the dBm power efficiency decreases. The diode forward voltage drop guides the reason of increasing efficiency plot. The reason efficiency decreases with dBm power is led by reverse breakdown voltage limit of the diode. The higher efficiency up to 70% has achieved in the present design.

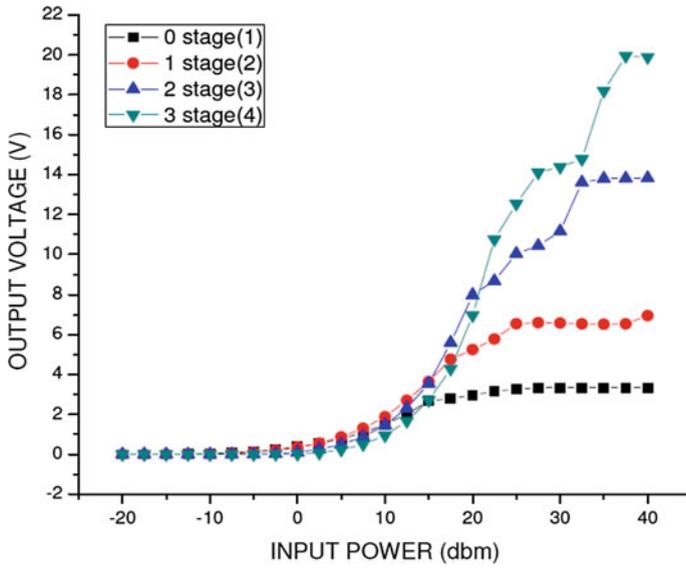


Fig. 5 Rectifier's output voltage versus input power (dbm)

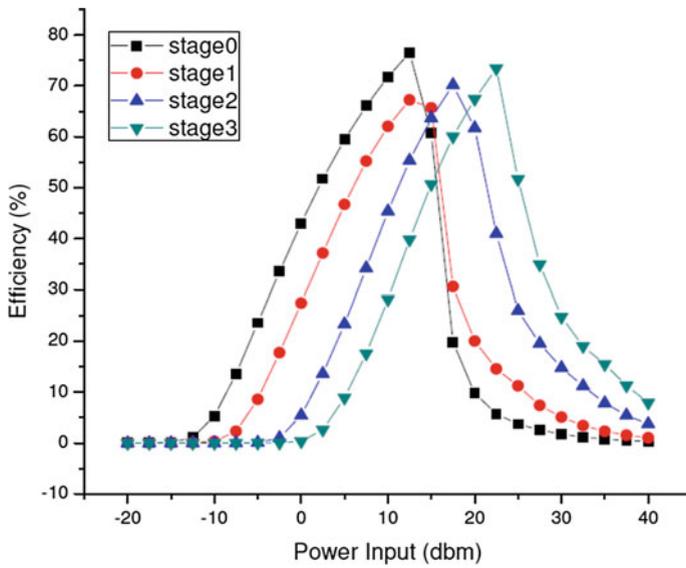


Fig. 6 Rectifier efficiency versus input power (dbm)

7 Conclusion

A 5.8 GHz rectenna framework has been analyzed, discussed, and designed in this work to investigate actualizing space-based Sun power. Utilizing Dickson charge pump rectifier topology, the simulation results have identical rectification properties according to the theoretical perspective. As per with the increase of charge pump stages, the yield voltage achieves a higher esteem. In the impedance matching segment, the stub-lengths were satisfactory for the HSMS-8202 diode and the match was accomplished with good results. Impedance matching performed considering 10 dBm power. The high rectenna circuit efficiency of 70% has attained. With the development of charge pump arranges, the rectenna efficiency has a most extreme at 10 dBm around. The planned rectenna framework can be effectively executed in SBSP ground-based microwave to DC control transformation.

References

1. Hayami, H., Nakamura, M., Yoshioka, K.: The life cycle CO₂ emission performance of the DOE/NASA solar power satellite system: a comparison of alternative power generation systems in Japan. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **35**(3), 391–400 (2005)
2. Brown, W.C.: Satellite power stations, a new source of energy. *IEEE Spectr.* **10**(3), 38–47 (1973)
3. Brown, W.C.: The history of power transmission by radio waves. *IEEE Trans. Microw. Theory Tech. MTT* **32**, 1230–1242 (1984)
4. US Department of Energy and NASA: Satellite Power System. Concept Development and Evaluation Program, Reference System Report, Oct 1978. Published Jan 1979
5. Johnson, G., Hunt, M.E., Determan, W.R., Hosang, A., Ivanenok, J., Schuller, M.: Design and integration of a solar AMTEC power system with an advanced global positioning satellite. In: *Proceedings of 31th IECEC*, vol. 1, p. 623. IEEE, Washington, D.C. (1996)
6. Lin, J.C.: Space solar-power stations, wireless power transmissions, and biological implications. *IEEE Microw. Mag.* 36–42 (2002)
7. Lin, J.C.: The new IEEE standard for human exposure to radio-frequency radiation and the current ICNIRP guidelines. *Radio Sci. Bull.* **317**, 61–63 (2006)
8. Akkermans, J.A.G., van Beurden, M.C., Doodeman, G.J.N., Visser, H.J.: Analytical models for low-power rectenna design. *IEEE Antennas Wirel. Propag. Lett.* **4**, 187–190 (2005)
9. Fully Integrated Dickson Charge Pumps with Optimized Power Efficiency. Doureloigne, Jan. San Francisco, USA: In: *Proceedings of the World Congress on Engineering and Computer Science*, vol. 2, 20–22 Oct 2010 (2010)
10. McSpadden, J.O., Fan, L., Chang, K.: Design and experiments of a high-conversion-efficiency 5.8-GHz rectenna. *IEEE Trans. Microw. Theory Tech. MTT* **46**(12), 2053–2060 (1998)
11. Takhedmit, H., Cirio, L., Merabet, B., Allard, B., Costa, F., Vollaïre, C., Picon, O.: Efficient 5.8 GHz rectenna design including harmonic rejecting rectifier device. *Electron. Lett.* **46**, 811–812 (2010)
12. Suh, Y.H., Chang, K.: A high-efficiency dual-frequency rectenna for 5.8- and 5.8-GHz wireless power transmission. *IEEE Trans. Microw. Theory Tech.* **50**(7), 1784–1789 (2002)

Dynamic 2-D Pattern-Based Passive Communication in Swarm Agents

V. Navya Deepthi, Chayan Banerjee and Shubhra Saxena

Abstract In a swarm of autonomous agents, agent to agent communication plays a vital role. A sustained communication over a large deployment area always contributes handsomely to the power budget. Indirect or passive communication methods can be a way out, where an agent does not directly communicate with another agent, but they do so by changing one's own appearance in some way. In the proposed framework, a robot will display its status or any important information in a 2-D barcode like, binary pattern, using low power led-based displays. Adjacent/neighbor robots with their onboard camera will capture the displayed pattern and decode it to know about the information. We have implemented and tested it for readability over different ranges and lighting conditions. The framework is found to work efficiently within a small to medium sized neighborhood (3–4 ft.).

Keywords Extended hamming code · Image thresholding · Led matrix Normalization · 2-D barcode

1 Introduction

Autonomous robotic swarms both aerial and terrestrial have become a much researched topic and its application in a number of fields has proven quite effective, especially in the domains of surveillance, exploration of unknown terrains,

V. Navya Deepthi (✉)

Sree Vidyaniethan Engineering College, Tirupati, Andhra Pradesh, India
e-mail: navyadeepthi45@gmail.com

C. Banerjee

BITS, Pilani, Rajasthan, India
e-mail: cbanerjee929@gmail.com

S. Saxena

Career Point Technical Campus, Rajsamand, Rajasthan, India
e-mail: shubhrasaxena25@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_74

localization, and mapping tasks. Despite the general availability of immediate communication devices, present onboard autonomous robotic agents, a range of factors that can either be of external or internal nature can compromise their reliability. Physical damage of a communication device conventionally leads to a permanent loss of competency to communicate, while a failure to establish a communication link between agents' results in temporal communication impairment [1–3]. Though the above approaches are applied to robot human interaction cases, involving gesture-based communication. But we try to involve two robots in pattern display, and machine vision-based communication. To specifically state, a major part of the energy of the robots is spent by the wireless communication subsystem that connects all the agents in the group into a network. Though there has been a lot of research in this field to reduce the communication energy budget, using sleep–awake mechanisms, etc., but it is necessary to rethink about the necessity to engage all the robots in a single network, or take other options like multi-hop communication, etc. Besides we also need to consider the amount of information that is needed to be transferred between the robots and the size of the neighborhood that a robot is willing to address and these parameters should dictate the type of communication that the agents should follow.

Our objective was to develop a framework which will help to communicate between the robots, using a type of indirect or passive communication strategy. In challenging circumstances like a post-disaster scenario or a natural calamity, direct communication can become severely constrained. SLAM techniques may work but again it may not be very efficient in mobile swarm robotic systems constrained in power and processing capabilities due to their sheer size. In the proposed framework, a robot will display its status or any important information in a 2-D barcode like, binary pattern, using low power led-based displays. Adjacent/neighbor robots with their onboard camera will capture the displayed pattern and decode it to know about the information. Section 2 presents a brief survey of approaches that have used pattern-based techniques for robot–robot or environment–robot communication, information gathering and localization. Sections 3 and 4 explain in detail the hardware and implementation of the proposed approach. Section 5 provides conclusion with clear points of improvement and future work.

2 2-D Barcode Data Representation

Recently, there has been a lot of application of barcode (both 1-D and 2-D)-based tags for identification and marking issues in robotics.

Inherently, the robots contain data like their position, orientation, and their identifier, which can be exceptionally valuable for alternate people to know. Therefore, the capability to recognize other robots relative outlook and location is very serene for implementation of large amount complex swarm algorithms (e.g., [7, 8]). In [6] for detection of landmark and other robots, four main features are introduced: QR code, Robot LED, and Robot body pattern (see Fig. 1).

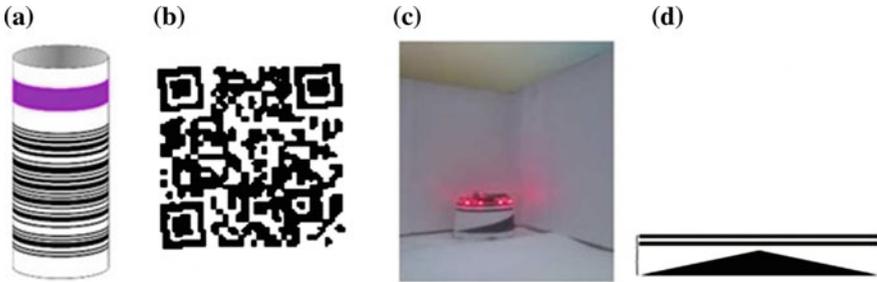


Fig. 1 The detectable features [6], **a** landmark barcode, **b** the QR code, **c** the LEDs on robot, **d** robot orientation pattern

The principle approach for identification of these components is to first utilize fundamental filters for highlighting the required data and after that zooming into the informative part of the picture for understanding it in more details. QR code is used here for dense data communication. In [9], similar feature-based framework is used for insect-inspired robot coordination. To empower visual robot–robot discovery, a robot (Turtle bot) was equipped with six unique markers, which were oriented in a way that at least one marker is visible from any angle. To track and decode these markers a toolkit called ALVAR was used. Kalman filtering was applied to get better and more stable readings, and therefore a more precise estimate of the detected robots position, heading and speed. These parameters are used again for collision avoidance.

In order to simplify the computer vision required by the robots to see the blocks, localizable 2-D barcodes called April Tags [10] are added to the faces of the block. Apart from representing a certain landmark or a certain encoded robot ID, the barcode can represent a set of much other information. What we are proposing to use a Led Dot Matrix display to make a 2-D barcode like feature can be accessed by any one of the neighbors of the robot and acted upon on the information. Now there can be different types of information that a certain swarm robot may wish to broadcast or narrowcast. The following variety of behavioral messages can be exchanged:

- Information about location, i.e., “I’m here”
- About Intention, i.e., “I am committed to this action”
- Ability, i.e., “I’m able to do this”
- Opportunity for Action, “These are the conditions for this action”
- About action accomplishment, i.e., “These are the conditions for this action”
- Informing about goal, “I have this goal”
- Informing about result, “This is the result”

There are different types of barcodes available in the market, namely QR code, Data Matrix, Aztec Code, etc. They vary in data density and complexity of implementation. There are two basic issues that all the popular barcodes tries to address; first, the proper identification of the tag’s orientation by the code reader

software and second, erasure correction, i.e., in the case the tag is not completely readable, the code will be able to correct and reclaim the unread part.

In our approach, since we are using a solid-state led-based display, we devised a simple Hamming-based format, since it is not quite possible to display QR and codes on small (8×8) LED-based displays. Besides in goods and other products where these tags are used, the robots are very rarely toppled and thus we found the inclusion of orientation data in the code, quite unnecessary.

3 Proposed Technique of Passive Communication

We propose the use of a LED matrix-based display to broadcast the current state of a robot. As shown in Fig. 2, every robot in this framework would have a low power 8×8 led-based displays, on each side and would also have an onboard camera to capture the pattern displayed by the neighborhood robots. The processor with the robots will be able to decode the pattern and let the other robot know about the fact that the pattern is displayed for.

4 Hardware Implementation and Analysis

4.1 Implementation Basics

We have used an extended Hamming code for the matrix. The data is read using a camera and across a distance, there may be a chance of erroneous readings on the status of the led's, in the LED matrix. So here we have used an extended version of Hamming error detection and correction code. Through the industry standard for

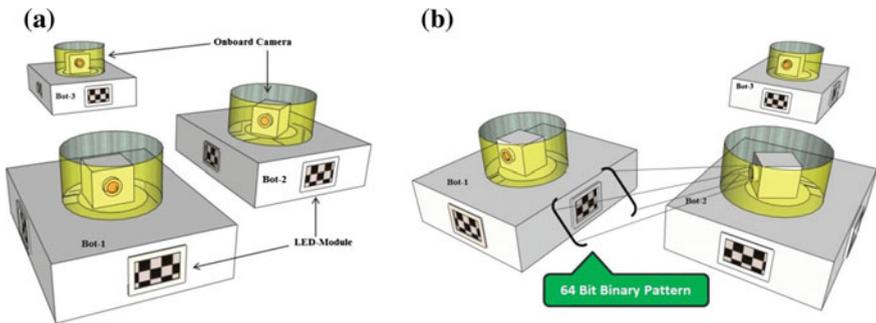


Fig. 2 Passive communication: **a**, **b** accessing of the 64 bit pattern on the LED display, by onboard camera of the neighbor robot



Fig. 3 Code word pattern for Hamming code

error detection and correction is Reed Solomon encoding, but for simplicity we have chosen the Hamming method.

Hamming belongs to the group of linear block codes. A systematic Hamming code has the following form (see Fig. 3), with properly separated Information Bits and Parity Bits.

May be represented as a (n, k) block code.

n : Block length of code = $2^r - 1$,

k : Information length of code = $2^r - r - 1$

r : Parity bits, where $r = n - k$

A block code can correct up to t errors, where $t = \frac{1}{2} (d_{\min} - 1)$ and d_{\min} is the minimum distance. And can also detect up to l number of error patterns. ‘ t ’ and ‘ l ’ are called the error correction and error detection limits. For Hamming code, $d_{\min} = 3$, thus it can detect up to two error patterns and can correct up to one error.

We have implemented an extended Hamming code $(n + 1, k)$, which involves incrementing the block length and the parity bit by 1.

The led matrix which we are using contains 64 LEDs, thus using the extended setting, total block length (code-word length) = 64.

Using unextended Hamming $r = 6$, since $2r - 1 = 63$.

Hence using extended setting, $r = 7$ and thus $n = 63 + 1 = 64$. The Hamming code can now be represented as $(64, 57, \text{ and } 7)$ or $(64, 57)$. Though there are 57 information bits available, we only use 56 ($=7 \times 8$) bits as we have considered 8 bits per message symbol. This bit remains always in High state, marking the left-most corner of the matrix, and helps in image acquisition and pattern recognition by image processing programs. The extended Hamming is able to correct a single error and detect double errors.

As it is clear from Fig. 4, the topmost row of the matrix contains the parity bits as will be generated by the Hamming coder. The rest of the rows will contain the rest of the code-word (or the information section of the block code).

We consider the symbols as an ASCII character, and first convert it to the equivalent 8 bit binary form. The serially appended binary sequence serves as the information bits for the Hamming coder. The Hamming coder generates a code-word, with seven parity bits and this bit sequence is finally displayed on the led matrix.

A neighbor robot may access this pattern using its onboard camera and using proper image processing algorithms will be able to decode the information embedded in the pattern.

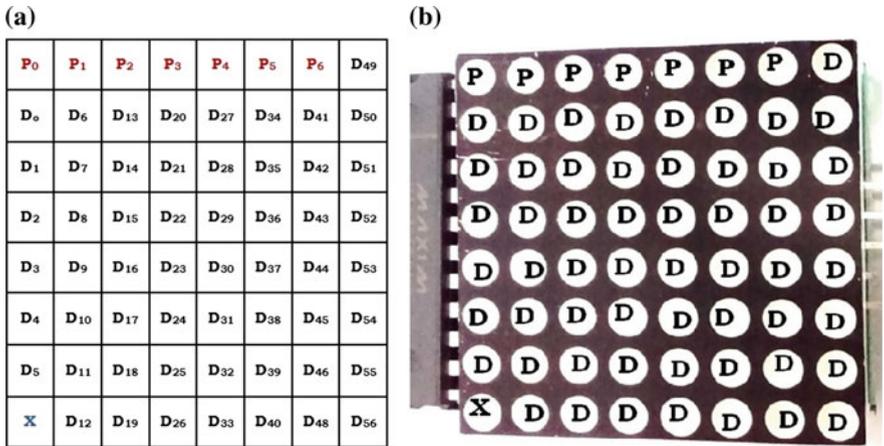


Fig. 4 Hamming encoded data representation: a, b on Led Matrix

The most important thing about this method of pattern generation and display is that the pattern is dynamic and two robots may also use this framework to communicate with each other.

5 Hardware Implementation

For the hardware implementation, we have used an ATmega-328 based Arduino board and a MAX 7219 driver based, LED matrix display. Table 1, shows the complete encoding of a message, LOC22 using extended Hamming code.

Table 1 Complete encoding of a message, LOC22 using extended Hamming code

Information	Individual symbols (ASCII)	Binary equivalent
LOC4022 (Above message conveys the information about the current location 'LOC' of the robot, in 'x = 40' and 'y = 22' coordinates)	L	01001100
	O	01001111
	C	01000011
	4	00110100
	0	00110000
	2	00110010
	2	00110010
Appended information binary sequence [56 + 1(Unused Bit, X)]		
010011010010011110100001100110100001100000011001000110010		
Hamming code-word (57 information + 7 parity)		
0100110100100111101000011001101000011000000110010001100101100000		

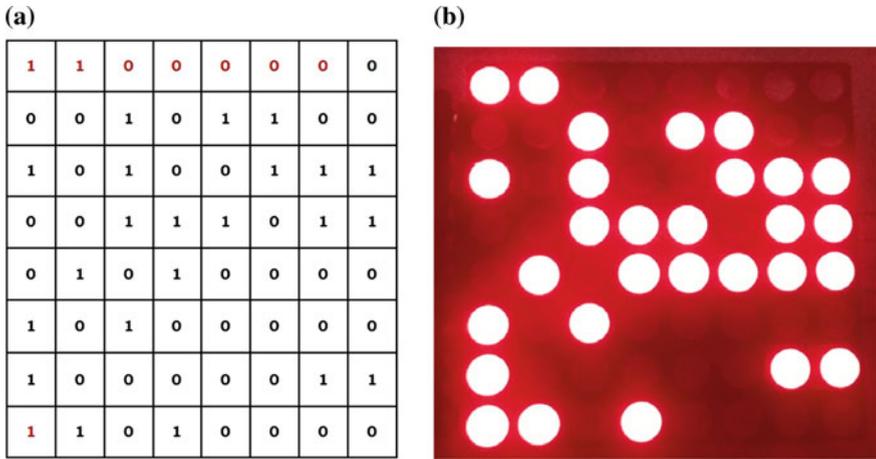


Fig. 5 Representation of coded data in led matrix: **a** data arrangement, **b** pattern as displayed on led matrix

Figure 5a, shows the representation of the encoded data on the led matrix and (b) shows the pattern displayed using the Led matrix.

6 Image Capture and Processing

The photographs were taken using a low resolution camera under different lighting conditions (indoor). The two dimension-scaled (without optical zoom) and grayscale images from two different distances (see Fig. 6) show that the pattern of the led matrix is clearly legible within a range of 2–3 ft., the image starts to get blurred if the capture distance goes beyond 4 ft. Still due to the availability of cheap and portable webcams with zoom and autoimage adjustment features (e.g., Logitech C170, C270, C310 etc.) the pattern can be easily captured from a greater distance and under relative motion of the vehicles and successfully decoded with no or minimum errors.

Parikh and Jancke [11] explains in detail the process of image localization and segmentation. But unlike our case they work with a 2-D color barcode. But still the initial steps of barcode localization apply well to our cause. The image is first captured, converted to gray scale, and normalized to remove changes brought in a single image by lighting variations. The image is then globally thresholded to remove the background. The step is followed by corner localization and barcode localization processes, like row localization and symbol localization. This paper does not cover the details of image processing and decoding techniques for our led matrix-based pattern.

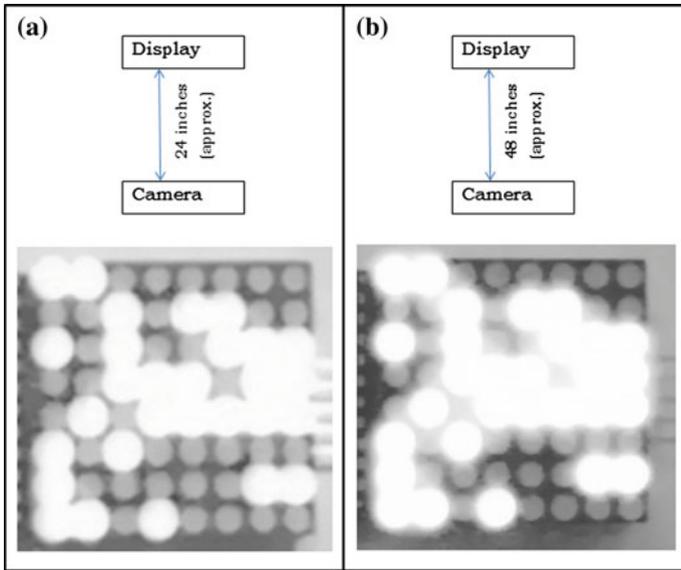


Fig. 6 Greyscale images captured from a distance of: a 24 in. b 48 in.

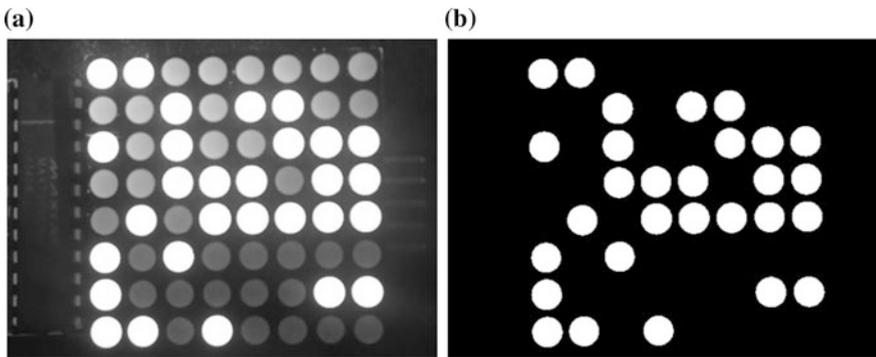


Fig. 7 Image capture distance < 2 ft.: a image in greyscale, b after global threshold, set threshold point = 0.90

However, we present a comparison of normalized threshold images, see Figs. 7 and 8. The image captured from 2 to 3 ft. (see Fig. 7) provides a clean pattern when thresholded. Image captured from a distance of around 4–5 ft. (see Fig. 8) is blurred, but when thresholded it provides a legible pattern which can be easily decoded without or with a little enhancement.

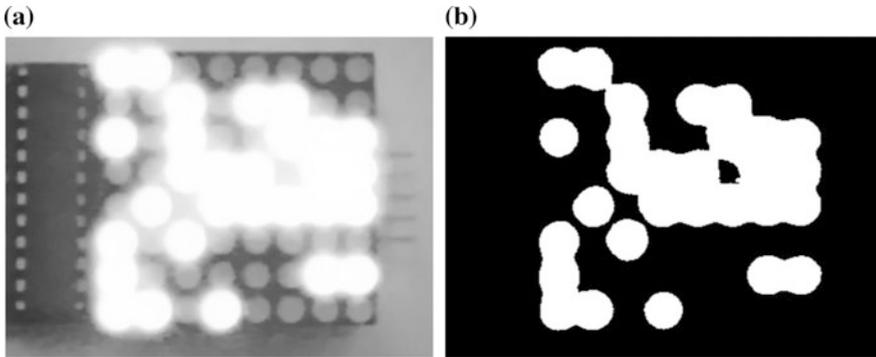


Fig. 8 Image capture distance >4 ft.: **a** Image in greyscale, **b** after global threshold, set threshold point = 0.95

7 Conclusion and Future Work

We have proposed a new approach of indirect or passive communication in autonomous agents, while working in a group. In this approach the agents with low power, led matrix-based display will advertise or narrowcast information to its immediate neighborhood, inform of an encoded 2-D binary pattern. To exemplify, we took arbitrary location information, encoded it in extended hamming code and displayed it on the matrix. We showed that the pattern image thus captured by a neighbor robot in a neighborhood radius of below 4 ft. and a clear line of sight can be evidently decoded for the embedded information.

Robots are mobile beings and so at a given instant a neighbor cannot guarantee that it will have a clear line of sight of the display of the intended robot. Another robot or an obstruction may block the view. Replacing the Hamming code with Reed Solomon may resolve the issue. RS code works well for erasure correction and so in our case it may work out well for partially visible/captured pattern image also. There is also a need to work on the image processing and enhancement of the captured image, such that the range of the neighborhood of a robot can be extended. An experimental study should also be made on the effect of robot agent velocities when they are displaying and capturing the information/pattern.

References

1. Nattanmai Parasuraman, R., Kershaw, K., Ferre Perez, M.: Experimental investigation of radio signal propagation in scientific facilities for telerobotic applications. *Int. J. Adv. Rob. Syst.* **10** (2013)
2. Mocialov, B.: Cooperation via indirect communication. Dissertation. Heriot-Watt University, (2015)

3. Mocialov, B., Vargas, P.A., Couceiro, M.S.: Towards the evolution of indirect communication for social robots. <http://www2.macs.hw.ac.uk/~bm4/papers/ieee.pdf>
4. Alers, S., Ranjbar-Sahrae B., May, K., Tuyls, S., Weiss, G.: Evaluation of an experimental framework for exploiting vision in swarm robotics. In: *Advances in Artificial Life, ECAL*, vol. 12, pp. 775–782, Sept 2013
5. Alers, S., Bloembergen, D., Hennes, D., De Jong, S., Kaisers, M., Lemmens, N., Tuyls, K., Weiss, G.: Bee-inspired foraging in an embodied swarm (demonstration). In: *Proceedings of the Tenth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2011)*, pp. 1311–1312 (2011)
6. Lemmens, N., Alers, S., Tuyls, K.: Bee-inspired foraging in a real-life autonomous robot collective. In: *Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC 2011)*, pp. 459–460 (2011)
7. Alers, S., Tuyls, K., Ranjbar-Sahraei, B., Claes, D., Weiss, G.: Insect-inspired robot coordination: foraging and coverage. In: *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, vol. 14, pp. 761–768 (2014)
8. Olson, E.: AprilTag: a robust and flexible visual fiducial system. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3400–3407. IEEE, May 2011
9. Parikh, D., Jancke, G.: Localization and segmentation of a 2D high capacity color barcode. In: *IEEE Workshop on Applications of Computer Vision, 2008. WACV 2008. IEEE (2008)*

Cloud-Based Cognitive Premise Security System Using IBM Watson and IBM Internet of Things (IoT)

Shubhradeep Nandi

Abstract This paper is a humble effort to demonstrate how we can capitalize on democratized artificial intelligence, commoditized internet of things, and convert our regular consumable hardware into a cost-efficient cognitive premise security system. Today, most of the commercial organizations either choose human security guards or deploy costly biometric systems to enable premise security. This approach is a stepping stone to replace such costly hardware with a convenient, reliable, scalable, and smart alternative that is built using IBM Watson and IBM internet of things.

Keywords Artificial intelligence · Cloud computing · Cognitive computing
IBM Watson · Internet of things (IoT) · Visual recognition

1 Introduction

We did a random survey in Bangalore, India and found that organizations, institutions, and residential complexes still handle their gate security either manually using human guards or using very expensive biometric systems.

While the former is prone to breaches due to human error, the latter is based on closed protocols where upgradation to efficient open-source protocols or incorporating any add-on features is an uphill task.

Two key technological advances of this decade are:

- (1) IBM opening up its powerful cognitive artificial intelligent system Watson for public consumption via APIs
- (2) Rapid innovation and fast commoditization of internet of things platforms.

As a data science solution architect and research scholar gave me an opportunity to innovate and build a cost-effective cognitive premise security system.

S. Nandi (✉)
IBM, Bangalore, India
e-mail: Shnandi1@in.ibm.com; shunandi@gmail.com

My approach was primarily to combine a hardware device that gets used in our day-to-day activities and build a system around it combining artificial intelligence.

2 IBM Watson and Its Relevance in Solving the Problem

IBM Watson is an artificial intelligence platform that combines the power of learning algorithms, understanding of natural language, and helps discover knowledge from volumes of unstructured data formats [1].

It can analyze unstructured data, understand complex questions, and present answers in form of recommendations using confidence scores. It learns a new subject from both structured and unstructured information provided to it and then when poised with a challenge it can find billion of data pointers to meaningfully extract an answer [2, 3]. Watson can collate, score, and identify the quality of the data pointers. It then uses its rank algorithm to weigh all possible replies taking into cognizance the weights of its supporting data pointers [4].

As of now Watson has around 25+ APIs exposed for public consumption via its platform as service called Bluemix [5]. Some of the key APIs include Alchemy API for Text Analytics, Sentiment Analysis, Visual Recognition, Natural Language Classifier, Conversation, and Tradeoff Analytics.

For building our innovative solution, the API that played the key role is visual recognition.

2.1 *Watson Visual Recognition*

Watson's vision helps a user to understand what composes an image or a video frame [6]. The user can request to the service to explore the content of an image of relevance, and it analyzes the image with a deep learning algorithm, post analysis it comes back with a quantized value of relevance for the identified class [6] representing the entities within.

The key value that Watson vision brings are as follows:

- Does this image hold any value for the use case in hand?
- Value proposition of collating and sequencing the content within the images?
- The taxonomy, richness, entity classification that can add value to the overall use case.

Customization of vision based on self-prepared images is possible as shown in Fig. 1 making it robust for personalized and customized domain-specific use cases.

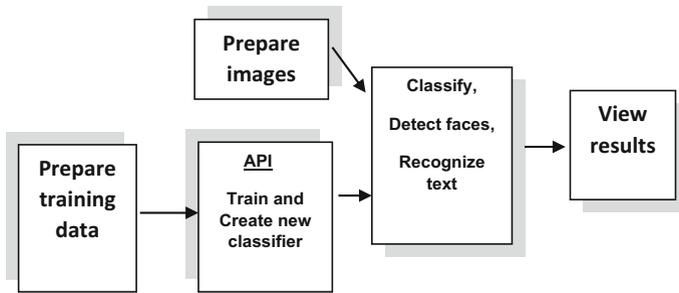


Fig. 1 Using the Watson visual recognition as defined by IBM

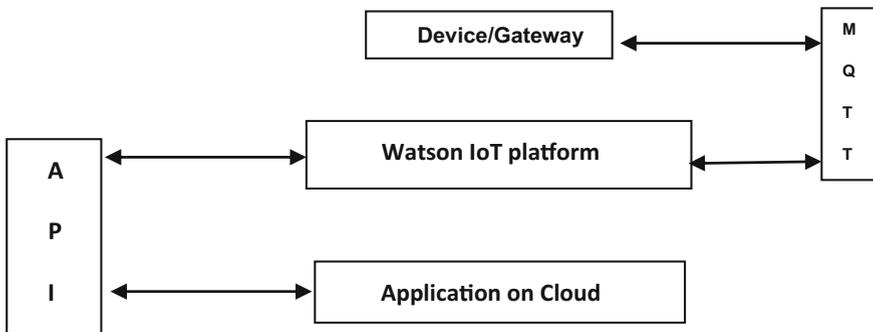


Fig. 2 Using the IBM internet of things platform as defined by IBM

3 Internet of things and Its Relevance

The internet of things (IoT) is to connect physical as well as smart devices and enable them to collect and exchange data [7]. Once the data pipeline is made available, it enables real-time analysis of that data [7] and then aid toward building an integrated or segregated holistic network of meaningful devices that can aid in building solution of business critical problems.

3.1 IBM Internet of things Foundation

The IBM internet of things service helps your application connect and, consume data transmitted by your connected devices, sensors, and gateways. IBM has built a set of organized recipes that make it easy to get devices connected to internet of things cloud platform. Also it is relatively easier to control and command devices distributed across geography via the combination of Cloud and very light MQTT protocol as shown in Fig. 2 [8].

4 The Edison Module for Compute from Intel

This compute module [9] is a double-core central processing unit and it has one core mc that can perform crunching of complex problems yet consuming less power. We have used this device to interpret logical commands that are transmitted from IBM IoT platform in accordance with recommendations from IBM Watson.

5 Building the Cognitive Premise Security System

To build this system, four key entities were brought into existence as shown in Fig. 3.

1. The Passport—A hardware that holds the key to authentication and identification and is the one that is both personal and handy and used in our day-to-day usage.
2. The brain—A customizable analytical engine that has visual recognition capabilities and is capable of classifying images in various classes.
3. The Master Mediator—An algorithm based logical gateway that has three key capability.
 - a. Authenticate and register hardware.
 - b. Consume coordinate and location intelligence.
 - c. Capability to logically bridge the communications from passport to brain and brain to executor.
4. The Executor—A device that can act as a microcontroller ensuring opening and closing of the premise gateway (cognitive entrance).

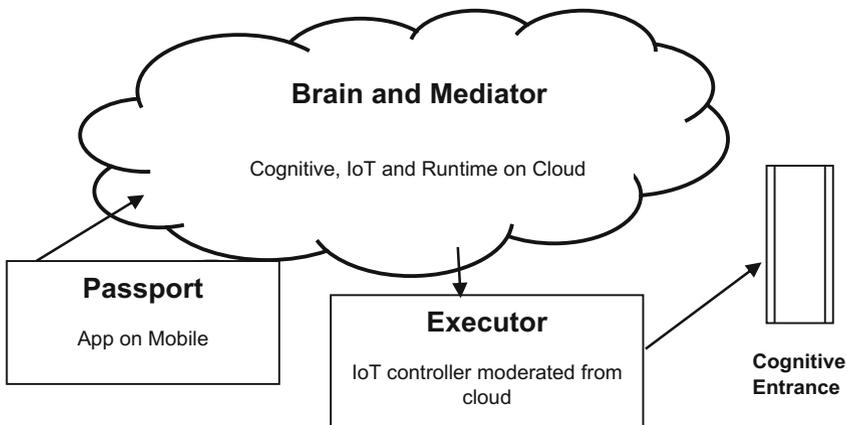


Fig. 3 Cognitively premise security system

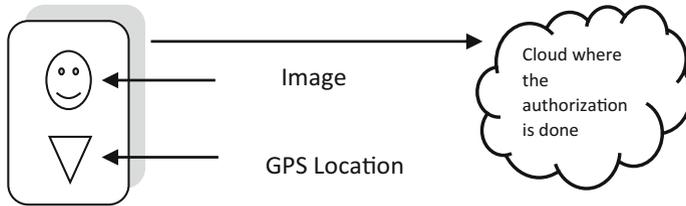


Fig. 4 Passport

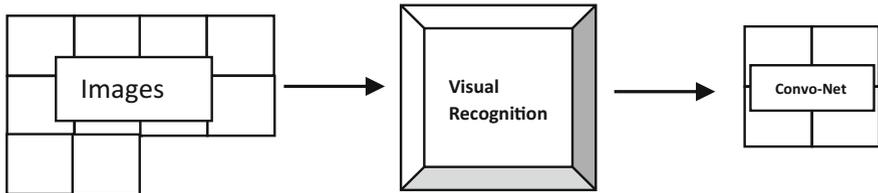


Fig. 5 Training the brain

5.1 The Passport

The passport includes an essential hardware that we use in our day-to-day life and a smart application installed on it.

The essential hardware that we have chosen is mobile as that has slowly become a non-detachable entity in our day-to-day life. From the smart application perspective, we have written a native android application that captures the real-time image of an individual from the camera, intercepts the location using the global positioning system, and transmits the same to cloud as shown in Fig. 4.

5.2 The Brain

The brain consumes the information transmitted by the passport and takes the decision whether to allow or dis-allow an individual to enter the premise.

The brain in our case is IBM Watson an artificial intelligent system with cognitive capabilities that amplifies human cognition and is democratized via APIs in IBM Bluemix platform as a service [5]. We trained the system as shown in Fig. 5. With a set of specialized images from the individuals who are allowed to enter the authorized premise.

The Watson AI system internally uses its deep learning convolution network algorithms to classify the images into authorized and rejected buckets.

Once a request for entry to a secured premises is received, it is combined with a real-time image from the individual taken at the entry location. This image

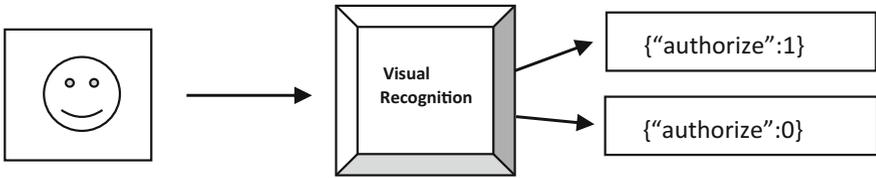


Fig. 6 Testing the brain

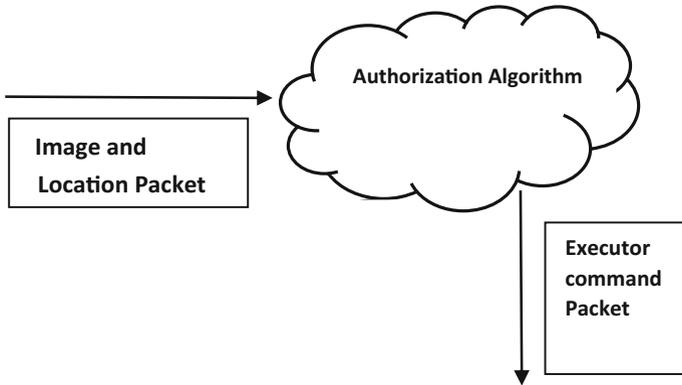


Fig. 7 Master mediator engine

subsequently forms the basis for visual recognition and authentication by the brain as shown in Fig. 6.

This information is further transmitted to the master mediator engine which conveys it further down to the executor.

5.3 The Master Mediator

The master mediator is the one which consumes the request from passport, initiates the brain to recommend, and prescribes the executor to take the appropriate action as shown in Fig. 7.

The master mediator in our case is IBM internet of things Platform which includes internet of things service and Node based master application runtime platform available on IBM Bluemix [5].

Each time there is a request for accessing a premise the following steps are followed:

- (1) The internet of things service identifies whether the device raising the request is among the set of registered devices using unique id and holds the generated authentication token.

- (2) The master runtime consumes the location coordinates and uses it as the second step of authentication. In the second step of authentication, it consumes the device coordinates and uses the location intelligence algorithm to validate the device proximity.
- (3) Once the proximity is validated, the image is send to the brain that is the custom trained Watson visual recognition service to identify the individual and his real-time surrounding.
- (4) Once the image is validated, the master runtime commands the executor to take the necessary action.

5.4 The Executor

The executor is one which consumes the digital command from the master mediator and help trigger an analog event on the underlying hardware in our case it is the automated gate or security alarm as shown in Fig. 8.

In our case, the executor is Intel Edison microcontroller device which based on the digital command from master mediator either opens up the automated gate or triggers a security alarm for human intervention or escalation.

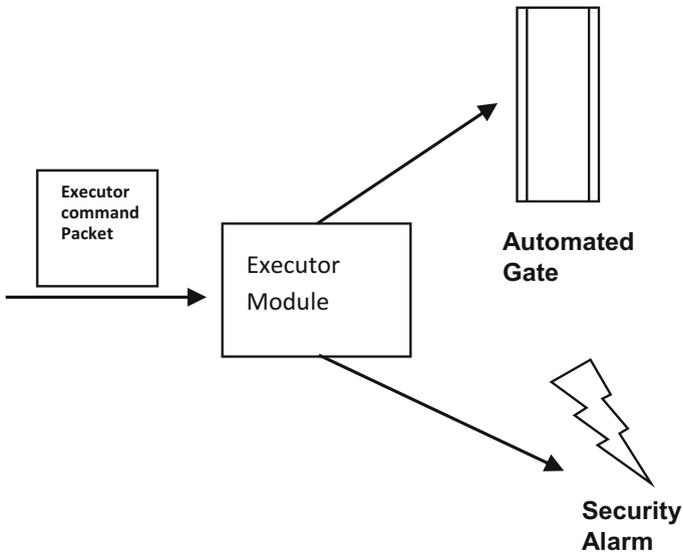


Fig. 8 Executor

6 Benefits of Deploying this System

There are multiple key benefits of deploying this cloud-based internet of things enabled artificially intelligent premise security system.

1. As it is a cloud-based pay as you go system you pay only for the resources that you use. Also the IBM platform as a service cloud Bluemix has a freemium model which brings down the overall cost [5].
2. World class deep learning algorithms for visual recognition that provide precise recommendations using spatial pixel data analysis.
3. Algorithms run on a highly scalable infrastructure that leaves no room for latency or delay.
4. Training the Watson algorithms is quick owing to the advanced convolution network based on visual recognition algorithm.
5. Data security is the responsibility of IBM and it has been a pioneer in enterprise delivery over decades.
6. IBM IoT platform uses MQTT a light messaging protocol that helps quicker transmission of messages across devices.
7. Device security, authentication is taken care by robust IBM IoT platform.
8. Human errors are no more a concern as machines work tirelessly and with precision.
9. Mobile is an essential smart hardware that all of us carry so there is no need to invest on any separate hardware.
10. Application can be customized for generic or specific clusters.
11. Global positioning system [10] does not need an active internet connection hence can still transmit the location data offline in a localized network deployment. The localized network deployment can open a secured channel to Internet and talk to a public cloud for decision making.

7 Conclusion and Future Scope

The premise and perimeter security today is primarily handled using manual workforce and this accounts to 93% of all the security breaches.

Biometric systems although effectively involve special hardware which comes with separate cost. It also involves additional cost of integration, algorithms, platform, and hardware to execute them. These systems are also prone to hacking as they are built using closed local algorithms and are installed and run on the premise.

This project is a humble effort to secure the premise with a cognitive, scalable, customizable, legitimate, and reliable system yet without involving any costly or dedicated hardware.

This project is now at its nascent state, but I have received significant accuracy during our pilot deployment and the prototype has also been well accepted in our demonstration workshops.

Acknowledgements I would like to thank International Business Machines Corporation (IBM) for providing me all the software tools and devices for carrying out this experiment. I would also like to thank Mr. Rajesh Jeypaul my colleague at IBM for his contribution in programming the Intel Edison device.

References

1. William Murdock, J.: Decision Making in IBM Watson™ Question Answering. Web presentation: Ontology Summit (2015)
2. Lally, A., Prager, J.M., McCord, M.C., Boguraev, B.K., Patwardhan, S., Fan, J., Fodor, P., Chu-Carroll, J.: Question analysis: how Watson reads a clue. *IBM J. Res. Dev.* **56**(3.4), 2–1 (2012) (IBM)
3. Lally, A., Bachi, S., Barborak, M.A., Buchanan, D.W., Chu-Carroll, J., Ferrucci, D.A., Glass, M.R., Kalyanpur, A., Mueller, E.T., Murdock, J.W., Patwardhan, S.: WatsonPaths: Scenario-based Question Answering and Inference over Unstructured Information. IBM Research Report RC25489, IBM (2014)
4. Gondek, D.C., Lally, A., Kalyanpur, A., Murdock, J.W., Duboue, P., Zhang, L., Pan, Y., Qiu, Z.M., Welty, C.: A framework for merging and ranking of answers in DeepQA. *IBM J. Res. Dev.* **56**(3/4), 14 (2012)
5. Nandi, S.: Methodical Evaluation and Exploration of Bluemix—A Versatile Platform as a Service (Xpaas) that is built to Disrupt. Indian Technology Congress (2015)
6. <http://www.ibm.com/watson/developercloud/doc/visual-recognition/customizing.shtml>. Accessed 14 June 2016
7. <https://console.ng.bluemix.net/docs/services/IoT/analytics.html>. Accessed 17 Aug 2016
8. <http://internetofthingsagenda.techtarget.com/definition/MQTT-MQ-Telemetry-Transport>. Accessed 3 July 2016
9. <http://www.intel.com/content/www/us/en/do-it-yourself/edison.html>. Accessed 19 July 2016
10. Langley, R.B.: The global positioning system: past, present, and future. In: *The Future of Geodesy and Geomatics Research and Education* Fredericton, N.B. (2001)

A MATLAB-Based Simulator for Amorphous Silicon and Polycrystalline Silicon Thin Film Transistor

Suman Das, Subham Datta and Somenath Chatterjee

Abstract High-performance active matrix organic light emitting diode (AMOLED) display is driven by a matrix of thin film transistors (TFTs) as a back plane technology. Nowadays, the majority of large area AMOLED display uses a-Si:H TFT and Poly-Si TFT to control the driving current in each and every pixels of AMOLED display. AMATLAB-based simulator for a-Si:H and poly-Si based thin film transistor (TFT) is proposed to address the pixel of AMOLED display. The purpose to design the simulator is to provide a ready reference of the output parameters of conventional (mentioned above) TFTs for the optimization of fabrication processes. This simulator is able to explain transfer characteristics, output characteristics and field effect mobility for both the TFTs.

Keywords a-Si:H TFT · Poly-Si TFT · MATLAB based graphical user interface · Transfer characteristics · Output characteristics and field effect mobility

1 Introduction

Amorphous silicon TFT [1] and polycrystalline silicon TFT [2] are the two back-plane technologies used as a switching device in AMOLED display [3]. Earlier a-Si:H TFT dominated in back plane technology market for large area electronic display but due to its low mobility slowly a-Si:H TFT was replaced by poly-Si TFT. But due to higher cost poly-Si TFT is mainly used in small display technology.

S. Das (✉) · S. Datta · S. Chatterjee (✉)

Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, India
e-mail: das.suman194@gmail.com

S. Chatterjee
e-mail: somenath@gmail.com

S. Datta
e-mail: Subham1995datta@gmail.com

In this context, a semi-empirical MATLAB-based modelling was introduced in previous work [4].

In this work, a MATLAB-based graphical user interface (GUI) simulator (TFT) is reported for a-Si:H TFT and poly-Si TFT which has not been reported yet in the other literature. In this paper, we are introducing our TFT simulator in MATLAB software environment with its own GUI to obtain the transfer characteristics, output characteristics and mobility which has a good settlement with the experimental data reported in the other published article.

The advantage of user interface is referred to a ready information regarding the performance of TFTs (a-Si:H TFT and poly-Si TFT) to optimize the design as well as selection of materials based on the properties. The objective of this work is to make an easy understanding of TFT using a semi-empirical model-based simple simulator. This simulator has the ability to take the input from user and generates the characteristic curve.

2 Software Structure

A MATLAB GUI-based TFT simulator is proposed here. It will simulate its m-file program when the software will get inputs from the user. The GUI interface is simple and user friendly which will provide the following curves

- Transfer characteristics for a-Si:H TFT (Drain-to-Source Current I_{DS} vs. Gate-to-Source voltage V_{GS}).
- Transfer characteristics for Poly-Si TFT (Drain-to-Source Current I_{DS} vs. Gate-to-Source voltage V_{GS}).
- Output characteristics for a-Si:H TFT (Drain-to-Source Current I_{DS} vs. drain-to-Source voltage V_{DS}).
- Output characteristics for Poly-Si TFT (Drain-to-Source Current I_{DS} vs. drain-to-Source voltage V_{DS}).
- Field effect mobility for a-Si:H TFT (Field effect mobility μ_{FET} vs. Gate-to-Source voltage V_{GS}).
- Field effect mobility for Poly-Si TFT (Field effect mobility μ_{FET} vs. Gate-to-Source voltage V_{GS}).

Figure 1 shows the flowchart of the TFT simulator.

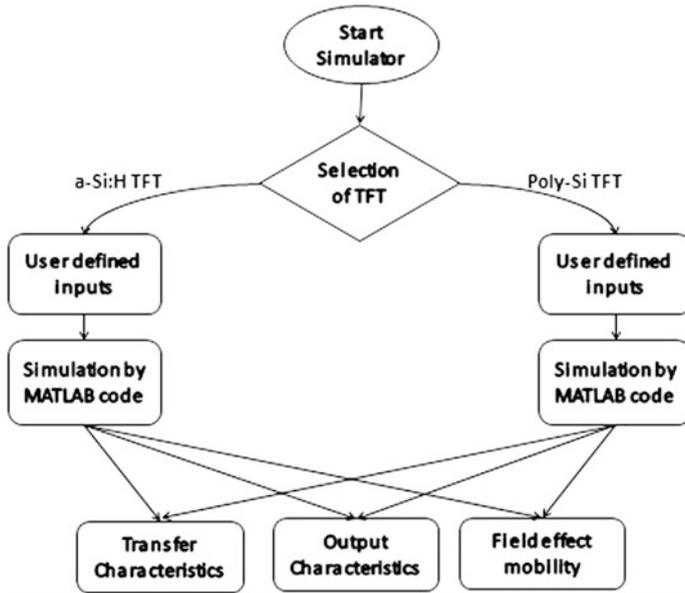


Fig. 1 Flowchart of TFT simulator used to design the GUI

3 Modelling and Validation of a-Si:H TFT Using TFT Simulator

3.1 Basic Equations Used in Modelling

Transfer characteristics of a-Si:H TFT is divided into three regions of operations namely leakage region, subthreshold region and above threshold region [5–7]. Hence, the current in leakage region is modelled as [5]

$$I_{leak} = I_{OL} \left[\exp\left(\frac{V_{DS}}{V_{DSL}}\right) - 1 \right] \exp\left(\frac{-V_{GS}}{V_{GSL}}\right) \exp\left[\frac{E_{AL}}{q} \left(\frac{1}{V_{tho}} - \frac{1}{V_{th}}\right)\right] + \sigma_0 V_{DS} \quad (1)$$

where I_{OL} is zero bias leakage current, V_{DSL} is hole leakage current drain voltage parameter, V_{GSL} is hole leakage current gate voltage parameter, E_{AL} is activation energy for hole leakage current and σ_0 is minimum leakage current parameter. The subthreshold current is modelled as [6]

$$I_{sub} = q \cdot \mu_{Band} \cdot n_{so} \left(\frac{t_m}{T_{OX}} \frac{V_{gfbce}}{V_0} \varepsilon \right)^{\left(\frac{2V_0}{V_{ce}}\right)} \cdot \frac{W}{L} \cdot \frac{V_{ds}}{\sqrt[M]{1 + \frac{V_{ds}}{\alpha_{sat} V_{gfc}}}} \quad (2)$$

where μ_{BAND} is the conduction band mobility, W is the width of the channel layer and L is the length of the channel layer; where n_{sb} is the sheet electron concentration in subthreshold region which depends upon n_{so} the dark sheet carrier density, T_{OX} the thickness of the insulating region, ϵ the relative dielectric constant of the gate insulator region, V_0 , the characteristics voltage of the deep state and t_m the charge channel thickness and M is the knee shape parameter.

$$V_{\text{gte}} = \frac{V_{\text{MIN}}}{2} \left[1 + \frac{V_{\text{gt}}}{V_{\text{MIN}}} + \sqrt{\Delta^2 + \left(\frac{V_{\text{gt}}}{V_{\text{MIN}}} - 1 \right)^2} \right] \quad (3)$$

where α_{SAT} is the saturation modulation parameter, V_{MIN} is the convergence parameter and Δ is the transition width parameter.

In above threshold region, the current equation becomes [8]

$$I_a = \mu_{\text{BAND}} \left(\frac{V_{\text{gs}} - V_T}{V_{\text{AA}}} \right)^\gamma \cdot \frac{W}{L} \cdot \frac{\epsilon}{q T_{\text{OX}}} \cdot \frac{V_{\text{gte}}}{V_{\text{aat}}} \left(\frac{V_{\text{gte}}}{V_{\text{aat}}} \right)^\gamma \cdot V_{\text{dse}} \cdot [1 + (\lambda \cdot V_{\text{ds}})] \quad (4)$$

where λ is the output conductance parameter, μ_{BAND} is the conduction band mobility, V_{AA} is the characteristic voltage for field effect mobility (determined by tail states) and γ is the power law mobility parameter.

$$V_{\text{aat}} = V_{\text{AA}} \cdot \exp \left[\frac{\text{EMU}}{q \cdot \gamma} \left(\frac{1}{V_{\text{th}}} - \frac{1}{V_{\text{tho}}} \right) \right] \quad (5)$$

where EMU is the field effect mobility activation energy.

4 Modelling of Transfer Characteristics Using a-Si:H TFT Simulator

Based on Eqs. (1), (2) and (4), an efficient modelling of transfer characteristics of a-Si:H TFT is presented here in MATLAB GUI environment. This a-Si:H TFT simulator is based on inverted staggered bottom-gated structure. It is taking the user inputs like drain to source voltage, insulator thickness, channel layer (active layer) width and channel layer length. After giving all inputs, plot command simulates the background MATLAB code and generates the outputs. The transfer characteristics of a-Si:H TFT is presented using TFT simulator as shown in Fig. 2.

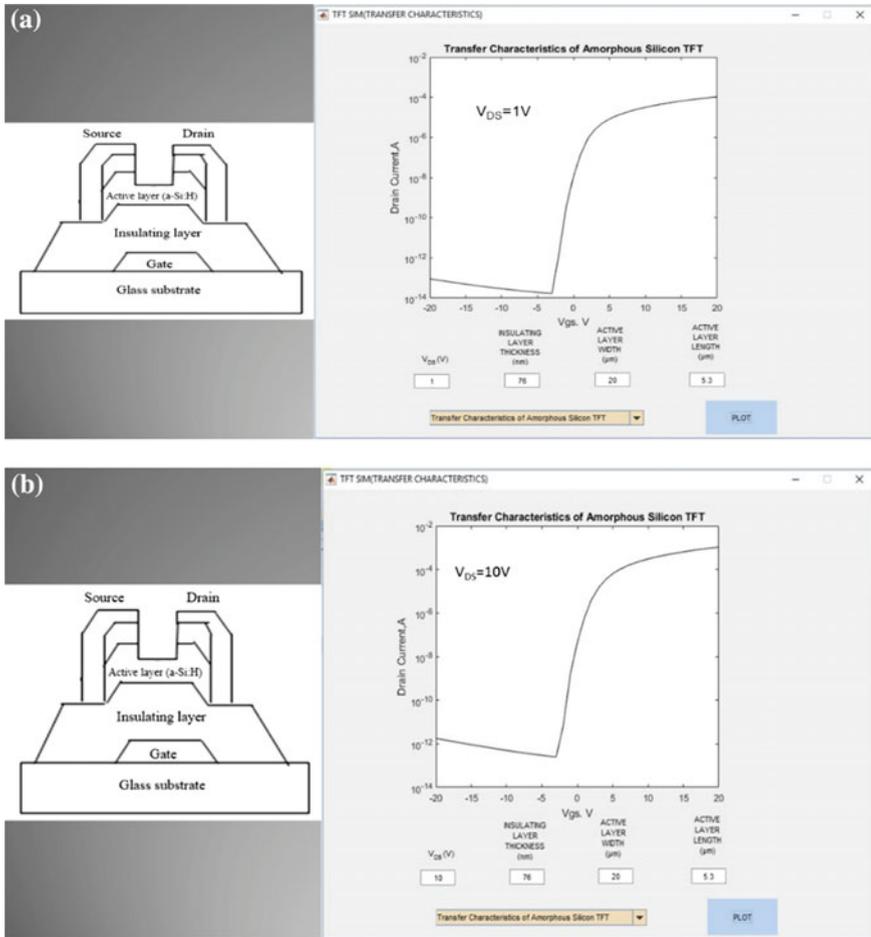


Fig. 2 Transfer characteristics of a-Si:H TFT for a $V_{DS} = 1 V$, b $V_{DS} = 10 V$ taking $W = 20 \mu m$, $L = 5.3 \mu m$ and $T_{OX} = 76 nm$

5 Modelling of Output Characteristics Using a-Si:H TFT Simulator

The output characteristics are obtained using Eq. (4) and it is depicted using our TFT simulator as shown in Fig. 3. It is taking the user inputs like gate to source voltage, insulator thickness, channel layer (active layer) width and channel layer length. Output characteristics are generated by plot command shown in the simulator.

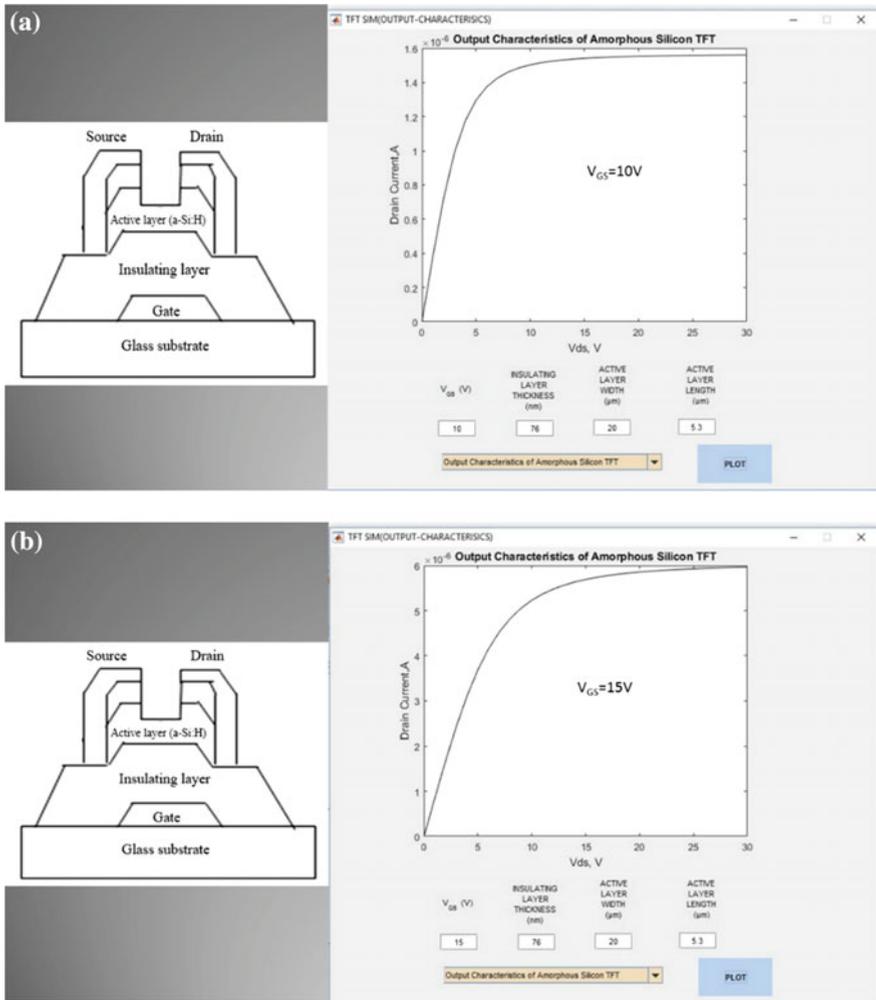


Fig. 3 Output characteristics of a-Si:H TFT for a $V_{GS} = 10$ V, b $V_{GS} = 15$ V

6 Modelling of Field Effect Mobility Using a-Si:H TFT Simulator

The simulated field effect mobility versus gate to source voltage for a-Si:H TFT is shown in Fig. 4 using a-Si:H TFT simulator. Here the observed mobility for $V_{GS} = 1$ V is almost $1 \text{ cm}^2/\text{Vs}$.

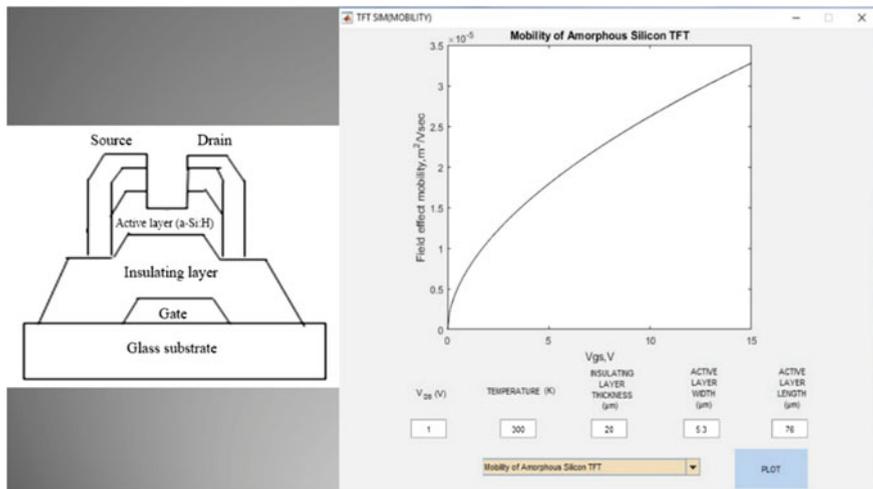


Fig. 4 Field effect mobility versus V_{GS} for a-Si:H TFT using a-Si:H TFT simulator

7 Modelling and Validation of Poly-Si TFT Using TFT Simulator

7.1 Basic Equations for Modelling

The poly silicon TFT suffers from relative high off current and this is the limitation of poly silicon TFT in large area electronics [9]. The leakage is expressed as

$$I_{leak} = q \cdot V_{OL} \cdot N_t \cdot \frac{e'_n \cdot e'_p - C_n C_p n p}{e'_n + e'_p + n C_n + p C_p} \tag{6}$$

where V_{OL} is the effective volume if the region where traps contribute to the leakage current and N_t is the trap density.

$$e'_n = \sigma_n v_n N_C \exp \left[\frac{-(E_c - E_t)}{KT} \right] \left[\exp \left(\frac{\varepsilon}{\alpha(T)^\beta} \right)^\gamma \right] \tag{7}$$

where e'_n is the expression for field emission rate in the presence of electric field, e_n is the electron emission rate of a trap in the absence of electric field and v_n is the thermal velocity of electron. N_C is effective density of states in the conduction band. σ_n is the electron capture cross section and E_t is the trap energy, where T is the temperature and α, β are the fitting parameters. Similarly,

$$e'_p = \sigma_n v_p N_V \exp\left(\frac{E_V - E_t}{KT}\right) \left[\exp\left(\frac{\varepsilon}{\alpha(T)^\beta}\right)^\gamma \right] \quad (8)$$

$$\varepsilon = \frac{V_{ds}}{d_1} - \frac{V_{GS} - V_{FB}}{d_2} \quad (9)$$

d_1 and d_2 gate oxide thickness.

N_V is the effective density of states in valance band. v_p is the thermal velocity of hole. σ_p is the hole capture cross section.

In subthreshold region, the drain current, I_{sub} can be modelled as

$$I_{sub} = \mu_s C_{OX} \frac{W}{L} (\eta V_{th})^2 \exp\left(\frac{V_{GS} - V_T}{\eta V_{th}}\right) \left[1 - \exp\left(\frac{-V_{DS}}{\eta V_{th}}\right) \right] \quad (10)$$

where μ_s is the subthreshold mobility, C_{OX} is the insulating oxide layer capacitance, W is the width of the active region and L is the length of the active region.

$$V_T = [(V_{ON} - D_{VT}) - D_{VTO}(T - T_{NOM})] - \frac{A_T \cdot V_{DS}^2 + B_T}{L} \quad (11)$$

where A_T and B_T are DIBL parameter, V_{ON} is the ON voltage, D_{VT} is the difference between the V_{ON} and threshold voltage and D_{VTO} is the temperature co efficient of V_{ON} .

In the above threshold region the drain current, I_a can be modelled as

$$I_a = \mu_{FET} C_{OX} \frac{W}{L} \left(V_{GT} V_{DS} - \frac{V_{DS}^2}{2\alpha_{sat}} \right) \text{ For } V_{DS} < \alpha_{sat} V_{GT} \quad (12)$$

$$I_a = \mu_{FET} C_{OX} \frac{W}{L} \frac{V_{GT}^2 \alpha_{sat}}{2} \text{ For } V_{DS} \geq \alpha_{sat} V_{GT} \quad (13)$$

where field effect mobility is denoted by

$$\frac{1}{\mu_{FET}} = \frac{1}{\mu_0} + \frac{1}{\mu_1 \left(\frac{2V_{GT}}{V_{sh}} \right)^M} \quad (14)$$

$$\mu_1 = MU1 - D_{MU1}(T - T_{NOM}) \quad (15)$$

$$\alpha_{sat} = A_{SAT} - \frac{L A_{SAT}}{L} - D A_{SAT}(T - T_{NOM}) \quad (16)$$

where μ_0 is the high field mobility, μ_1 is the low field mobility, M is the low field mobility exponent, A_{SAT} is the proportionality constant of V_{SAT} , $D A_{SAT}$ is the temperature co efficient of A_{SAT} and D_{MU1} is the temperature co efficient of μ_1 .

8 Modelling of Transfer Characteristics Using Poly-Si TFT Simulator

Based on Eqs. (6), (10), (12) and (13), an efficient modelling of transfer characteristics of Poly-Si TFT is presented here in MATLAB GUI environment. The transfer characteristics of Poly-Si TFT is presented using TFT simulator shown in Fig. 5.

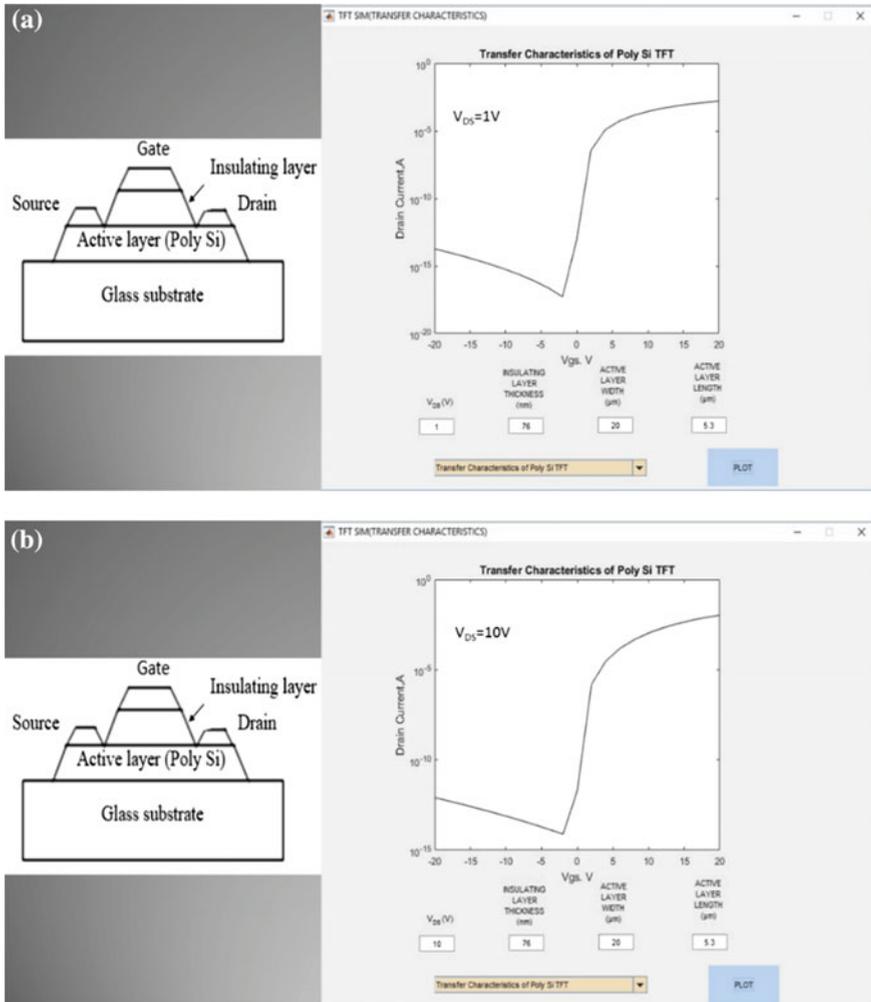


Fig. 5 Transfer characteristics of Poly-Si TFT for **a** $V_{DS} = 1V$, **b** $V_{DS} = 10V$ taking $W = 20 \mu\text{m}$, $L = 5.3 \mu\text{m}$ and $T_{OX} = 76 \text{ nm}$

From Figs. 2 and 5, it is well cleared that for same drain to source voltage poly-Si TFT simulator is showing less off current and more on current than a-Si:H TFT simulator. Relative ON/OFF ratio can be obtained from the simulator extracted data.

9 Modelling of Output Characteristics Using Poly-Si TFT Simulator

The output characteristics is obtained using Eqs. (12) and (13) and is depicted using our TFT simulator which is shown in Fig. 6.

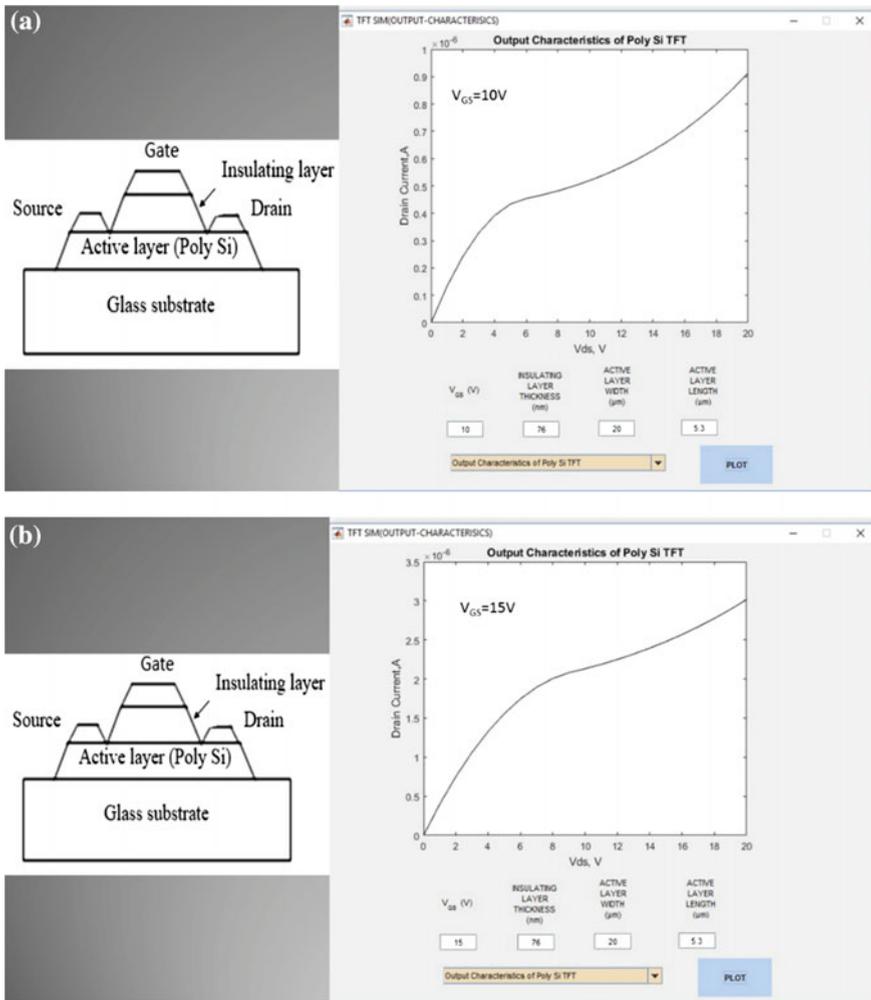


Fig. 6 Output characteristics of Poly-Si TFT for a $V_{GS} = 10\text{ V}$, b $V_{GS} = 15\text{ V}$

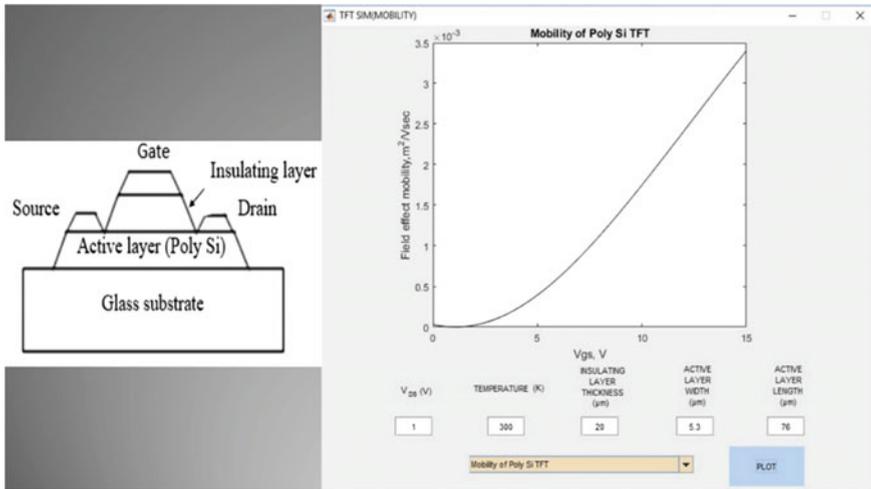


Fig. 7 Field effect mobility versus V_{GS} for Poly-Si TFT using Poly-Si TFT simulator

10 Modelling of Field Effect Mobility Using Poly-Si TFT Simulator

The simulated field effect mobility versus gate to source voltage for Poly-Si TFT is shown in Fig. 7 using Poly-Si TFT simulator. Here the observed mobility for V_{GS} = 1 V is almost 100 cm²/Vs which is 100 times greater than a-Si:H TFT simulator.

11 Conclusion

A simulator for a-Si:H TFT and Poly-Si TFT is developed to produce and explain the transfer characteristics, output characteristics and mobility in MATLAB GUI environment. All the curves are well visualized and it is well explained. From the above simulated outputs, it is clear that Poly-Si TFT exhibit better electrical performance than a-Si:HTFT. The main advantage of this simulator is it can take user inputs and can produce the outputs without any time delay. Thus, it is very user friendly and less complex.

References

1. Kuo, Y. (ed.): Thin Film Transistors Volume 1: Amorphous Silicon. Kluwer, Boston (2004)
2. Kuo, Y. (ed.): Thin Film Transistors Volume 2: Polycrystalline Silicon. Kluwer, Boston (2004)
3. Kuo, Y.: Thin Film Transistor Technology-Past, Present and Future. The Electrochemical Society (2013)
4. Das, S., Chatterjee, S.: An efficient modelling of amorphous silicon and polycrystalline silicon thin film transistor for AMOLED display using MATLAB. In: IEEE International Conference on Inventive Computation Technologies (ICICT-2016), (Aug 2016)
5. Servati, P., Nathan, A., Sazonov, A.: A physically-based SPICE model for the leakage current in a-Si:H TFTs accounting for its dependencies on process, geometrical, and bias conditions. Mater. Res Soc. Symp. Proc. **609**, A28.3.1–A28.3.6 (2000)
6. Servati, P., Nathan, A.: Modeling of the reverse characteristics of a-Si:H TFTs. IEEE Trans. ED-**49**(5), 812–819, (2002)
7. Nathan, A., Servati, P., Karim, K.S., Strikhilev, D., Sazonov, A.: Thin film transistor integration on glass and plastic substrates in amorphous silicon technology. IEE Proc. Circuits Devices Syst. **150**(4), 329–338 (2003)
8. Zhu, L.: Modeling of a-Si:H TFT I-V Characteristics in the Forward Subthreshold Operation, A thesis presented to the University of Waterloo, Waterloo, Ontario, Canada (2005)
9. Colalongo, L., Valdinoci, M., Baccarani, G., Migliorato, P., Tallarida, G., Reita, C.: Numerical analysis of poly-Si TFTs under off conditions. Solid-State Electron. **41**(4), 627–633 (1997)

Relay Node Selection Using Analytical Hierarchy Process (AHP) for Secondary Transmission in Multi-user Cooperative Cognitive Radio Systems

Jyoti Sekhar Banerjee, Arpita Chakraborty and Abir Chattopadhyay

Abstract In this paper, we have proposed a very new relay selection scheme based on the decision-making technique of analytical hierarchy process (AHP). Unlike many existing works where signal-to-interference-plus-noise ratio (SINR) is considered as the only parameter for relay selection, here the underlying decision criterion considers SINR at secondary destination (SD) as well as reliability and relative link quality (RLQ) of the relay node from destination.

Keywords Cognitive radio networks · Cognitive relay · Relay selection
Secondary transmission · Decision making · Analytic hierarchy process

1 Introduction

Cognitive radio (CR), the most contemporary buzz word [1–4] and one of the new long-term developments in wireless communication is empowered with the solution to alleviate the spectrum insufficiency problem by employing dynamic spectrum access where secondary user (SU) is allowed to access the licensed spectrum along with the primary user (PU) maintaining certain threshold value, i.e., interference temperature. First, cooperative diversity, then cooperative diversity [1, 5] based on relay selection has been studied in [6] for both the cognitive and non-cognitive systems, where transmission occurs in two channels. In this correspondence, we have specially concentrated on the formulation of the best relay selection criteria in cooperative CR networks [5, 7, 8]. Our major contributions can be summarized as

J.S. Banerjee (✉) · A. Chakraborty
Department of ECE, Bengal Institute of Technology, Kolkata 700150, India
e-mail: tojyoti2001@yahoo.co.in

A. Chakraborty
e-mail: chakraborty_arpita2006@yahoo.com

A. Chattopadhyay
Department of ECE, University of Engineering & Management, Kolkata 700156, India
e-mail: abir_chattopadhyay@yahoo.co.in

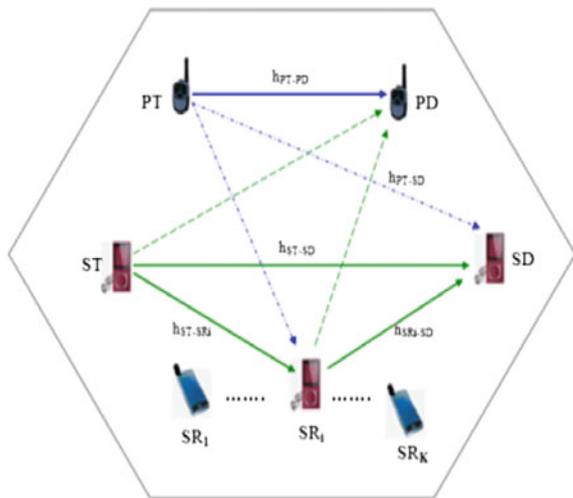
follows. First, we have proposed an AHP-based relay selection scheme [5, 9] which considers receiving SINR at secondary destination (SD) of the relay node, reliability, and relative link quality as the decision factors. Second, simulation shows a significant low latency in CR networks to select relay node using AHP.

2 System Model

Here an underlay cognitive relay [5, 6] network has been considered where both primary transmitter (PT) and secondary transmitter (ST) transmit their data message parallelly to the primary destination (PD) and secondary destination (SD), respectively, occupying the same frequency band (see Fig. 1). In this correspondence, we have depicted a secondary network model consisting of a single Source (ST) as well as a single destination node (SD) along with K relaying nodes (shown in Fig. 1). Here, the concept of aggregate channel model has been adopted that treats both path loss and slow Rayleigh fading [6] to simulate the wireless environment. In the underlay approach, time is divided into number of slots (t_f | f is any integer > 0) and each one is further divided into two sub-slots ($\langle t_f(i) | i \in (1, 2) \rangle$) viz *observation interval* and *relaying interval* that refers to two transmission phases. During the observation interval, all CR users try to listen to the primary user signal, and during the relaying interval each CR user behaves as a relay node. The fading coefficient is assumed to be constant during one time slot and varies independently with each time slot (see Fig. 2).

Observation interval/First sub-slot $\langle t_f(i = 1) \rangle$: During observation interval/the first sub-slot $\langle t_f(i = 1) \rangle$ ST transmits its data ($x_S \langle t_f(i = 1) \rangle$) of data

Fig. 1 Basic relay model



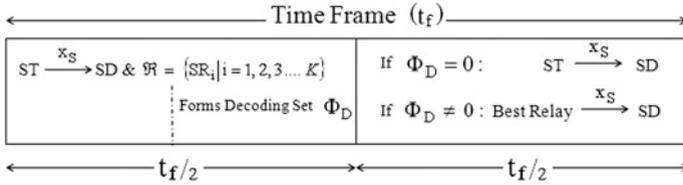


Fig. 2 Time slots of secondary transmission

rate R_{ST} to both secondary Relays (SRs) and secondary destination (SD) consuming power P_{ST} simultaneously with the PT, who transmits its data ($x_P \langle t_f(i=1) \rangle$) of data rate R_{PT} with power P_{PT} . Thus the signal received at the secondary relays (SRs) and SD becomes:

$$Y_{SR_i \in \mathcal{R}} \langle t_f(i=1) \rangle = \sqrt{P_{ST}} h_{ST,SR_i} \langle t_f(i=1) \rangle x_S \langle t_f(i=1) \rangle + \sqrt{P_{PT}} h_{PT,SR_i} \langle t_f(i=1) \rangle x_P \langle t_f(i=1) \rangle + n_{SD} \langle t_f(i=1) \rangle \tag{1}$$

$$Y_{SD} \langle t_f(i=1) \rangle = \sqrt{P_{ST}} h_{ST,SD} \langle t_f(i=1) \rangle x_S \langle t_f(i=1) \rangle + \sqrt{P_{PT}} h_{PT,SD} \langle t_f(i=1) \rangle x_P \langle t_f(i=1) \rangle + n_{SD} \langle t_f(i=1) \rangle \tag{2}$$

where $h_{ST,SD}$ and $h_{PT,SD}$ are coefficients of fading of the channel from ST to SD and from PT to SD, respectively, and $n_{SD} \langle t_f(i=1) \rangle$ indicates an additive white Gaussian noise. On receiving the signal from ST, all the SRs try to decode the received message. Only the SRs who are able to successfully decode the received message constitute a group named as decoding set and is represented as below:

$$\Phi_D \langle t_f(i=1) \rangle = \{ \varphi | \varphi \in \phi \cup \varphi_k, k = 1, 2, \dots, 2^K - 1 \}$$

The Relaying Interval/Second sub-slot $\langle t_f(i=2) \rangle$: If the Best Relay (SR_{BEST}) gets selected from the non empty Decoding set (φ_k) then during this phase SR_{BEST} transmits its correctly decoded message to the SD. Hence the received signal at SD is as follows:

$$Y_{SD} \langle t_f(i=2) \rangle |_{\varphi=\varphi_k} = \sqrt{P_{ST}} h_{SR_{BEST},SD} \langle t_f(i=2) \rangle x_S \langle t_f(i=2) \rangle + \sqrt{P_{PT}} h_{PT,SD} \langle t_f(i=2) \rangle x_P \langle t_f(i=2) \rangle + n_{SD} \langle t_f(i=2) \rangle \tag{3}$$

On the other hand, if $\varphi = \phi$, i.e., when not a single SR is able to decode the message correctly, consequently ST has to retransmit the original message to the SD over the direct link. Finally, SD executes MRC (Maximum ratio Combining) on the two copies of the received signal.

3 Best Relay Selection Parameters

So far, the existing studies reveal different best relay selection schemes which are based on received SINR at the SD [5, 6]. In this paper, for the first time, we have shown AHP-based relay node selection technique from the decoding set (φ_k) .

3.1 Signal-to-Interference-Plus-Noise Ratio (SINR_{SD})

It is a widely used metric for relay selection and is calculated at the destination by performing MRC on the two copies of the received signal transmitted by ST and SR_{*i*} nodes, respectively. Hence, the SINR obtained at SD is as follows:

$$\begin{aligned} \text{SINR}_{\text{SD}}(\text{SR}_i) \Big|_{i \in \varphi_k} &= \frac{P_{\text{ST}} |h_{\text{ST,SD}} \langle t_f(t=1) \rangle|^2}{P_{\text{PT}} |h_{\text{PT,SD}} \langle t_f(t=1) \rangle|^2 + N_0} \\ &+ \frac{P_{\text{ST}} |h_{\text{SR}_{\text{BEST,SD}}} \langle t_f(t=2) \rangle|^2}{P_{\text{PT}} |h_{\text{PT,SD}} \langle t_f(t=2) \rangle|^2 + N_0} \end{aligned} \quad (4)$$

3.2 Relative Link Quality (RLQ)

It determines the selected Relay node's channel quality relative to the other relay nodes belonging to the decoding set (φ_k) .

$$\gamma_{\text{SR}_i, \text{SD}} \Big|_{i \in \varphi_k} = \frac{H_{\text{SR}_i, \text{SD}}}{\frac{1}{K} \sum_{i=1}^K H_{\text{SR}_i, \text{SD}}} = \frac{|h_{\text{SR}_i, \text{SD}}|^2}{\frac{1}{K} \sum_{i=1}^K |h_{\text{SR}_i, \text{SD}}|^2} \quad (5)$$

where $h_{\text{SR}_i, \text{SD}}$ is the channel impulse response between nodes SR_{*i*} and SD.

3.3 Reliability

By the term Reliability, a relay node's performance to transmit a signal to the destination successfully, is expressed. As the correct copy of the signal reaches to the destination, an ACK (Acknowledgement) is delivered to both the ST and the corresponding SR_{*i*} indicating successful transmission, and likewise NACK is delivered to indicate failure transmission. Let during the time interval Δt (Where $\Delta t = t_1 - t_2$) total numbers of ACK and NACK received by the

relay node SR_i are $N_{SR_i|\Delta t}^{ACK}$ and $N_{SR_i|\Delta t}^{NACK}$, respectively. Rate of successful transmission ($\lambda_S(SR_i, \Delta t)$) of SR_i node may be defined as

$$\lambda_S(SR_i, \Delta t) \Big|_{i \in \phi_K} = \frac{N_{SR_i|\Delta t}^{ACK}}{N_{SR_i|\Delta t}^{ACK} + N_{SR_i|\Delta t}^{NACK}} \quad (6)$$

Now let us consider that during Δt time interval, probable number of transmissions carried by the SR_i node be denoted by: $\zeta_{(SR_i, \Delta t) \in \phi_K}$. \therefore Average number of successful transmissions carried by the SR_i relay node thus becomes

$$\bar{T}_S(SR_i, \Delta t) \Big|_{i \in \phi_K} = \lambda_S(SR_i, \Delta t) \Big|_{i \in \phi_K} \times \zeta_{(SR_i, \Delta t) \in \phi_K} \quad (7)$$

Thus, we may calculate the reliability (β_{SR_i}) of the particular node SR_i which is nothing but the average number of successful transmissions carried by the SR_i relay node over a longer period of time say T (where $T = t \times \Delta t$ | t is any integer and $\neq 0$) as considered in this paper.

$$\beta_{SR_i} = \sum_{j=1}^T \bar{T}_S(SR_i, \Delta t_j) = \sum_{j=1}^T \lambda_S(SR_i, \Delta t_j) \Big|_{i \in \phi_K} \times \zeta_{(SR_i, \Delta t) \in \phi_K} \quad (8)$$

4 AHP-Based Relay Node Selection

The analytic hierarchy process (AHP) is a method which is used in decision making where multiple criteria are involved. AHP is carried out in four main steps: (1) the problem needs to be decomposed into sub-problems or decision factors in decision hierarchy, (2) pairwise comparison of each decision factor, (3) evaluation of consistency, and (4) synthesize the entire results to obtain a final ranking of each alternative. In this paper, AHP is used to determine the relay node selection for secondary transmission in cooperative cognitive radio networks. Figure 3 shows the AHP hierarchy for proposed relay node selection scheme. At the top, the goal “selection of relay node” is mentioned. In the middle level, three criteria are considered as the decision parameter: (1) relative link quality (RLQ), (2) reliability, and (3) $SINR_{SD}$. The ‘n’ relay nodes of the decoding set stay at the bottom level of the hierarchy.

The weight of each criterion to the goal and weight of each alternative under each criterion are considered as local weights. Based on local weights, global weights, or final ranking of each alternative is calculated. After constructing the hierarchy, a matrix is developed using a pairwise comparison of each criterion with respect to the final goal. A pair of elements when compared, its result is a ratio of relative significance and it is expressed in fundamental 1–9 scale (see Table 1). Pairwise comparison results are presented in a square matrix A,

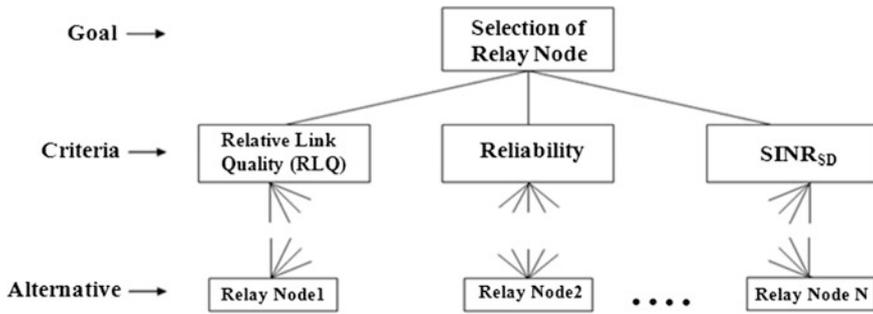


Fig. 3 AHP hierarchy for proposed relay node selection scheme

Table 1 The fundamental 1–9 scale

Intensity of importance in a numeric scale	Definition
1	Equal significance
3	Moderate significance
5	Essential or strong significance
7	Very strong significance
9	Extreme significance
2, 4, 6, 8	In between values of the two adjacent judgments

$$A = (a_{ij})_{n \times n} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{9}$$

where a_{ij} indicates the ratio of the i th and the j th decision factor, and n is the number of decision factors. In other term, every element in the matrix A is a pairwise ratio of the corresponding local weights or the matrix whose rows provide the ratio of the weights of each element with respect to all others, i.e., $a_{ij} = w_i/w_j$ where $i, j = 1, 2 \dots n$. The matrix A also maintains reciprocal property $a_{ij} = 1/a_{ji}$.

Again, the matrix A becomes a consistent matrix if every element satisfies the following equation:

$$a_{ij}a_{jk} = a_{ik}, \quad i, j, k = 1, 2, \dots, n \tag{10}$$

For the pairwise comparison-based matrix A , the priority weight vector is obtained using the theory of eigenvectors and eigenvalues. In the case of a consistent matrix A , the priority weight vector is the eigenvector of A in accordance with the maximal eigenvalue ‘ n ’, eigenvalue can be obtained based on the matrix equation $AW = nW$. For the matrix whose judgments are not fully consistent, the eigenvalue λ_{\max} is different from ‘ n ’ and the matrix equation is $AW = \lambda_{\max}W$. The elements of priority vector $W^T = \{w_1, w_2, \dots, w_n\}$ are considered as the local

Table 2 Random index (RI)

Average random consistency index (RI)										
n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

weight of the criteria. In this paper, we consider relative link quality (RLQ), reliability, and $SINR_{SD}$ are the three criteria; hence, the local weight can be expressed as $W_j^T = \{w_\alpha, w_\beta, w_\gamma\}$, where $w_\alpha, w_\beta, w_\gamma$ are the local weights of RLQ, reliability and $SINR_{SD}$ correspondingly. Due to the people’s random judgments, most of the matrices are not perfectly consistent. The judgments errors can be measured by the consistency ratio (C.R.), which is determined as the ratio of consistency index (C.I.) to random index (R.I.) defined as $C.R. = C.I./R.I.$. All R.I. values for different number of decision factors are given in Table 2. And C.I. of the AHP matrix is defined as $C.I. = (\lambda_{max} - n)/(n - 1)$.

Where

$$\lambda_{max} = \frac{1}{n} \sum_{i=1}^n \frac{(AW)_i}{W_i} \tag{11}$$

For the consistent matrix, the value of C.I. is equal to zero; otherwise C.I. should have positive value. When $C.R. \leq 0.1$, the judgment error is within the limit and weight coefficients of the local weight W_j are considered as the weight of the criteria. Adjustment of the pairwise comparisons is needed when $C.R. > 0.1$ to satisfy the consistency check. This process is continued to the downward level by level to the bottom of the hierarchy, for calculating the weights of the alternatives. In this paper, authors consider that three relay nodes can decode the signal; hence, the weights of three relay nodes under the three decision factors, i.e., RLQ, reliability, and $SINR_{SD}$ result in a 3×3 matrix, defined as w_{nij} , present below:

$$w_{nij} = \begin{matrix} & \begin{matrix} \text{RLQ} & \text{Reliability} & \text{SINR}_{SD} \end{matrix} \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \end{matrix} & \begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix} \end{matrix} \tag{12}$$

Consider $n_1, n_2,$ and n_3 represent the relay nodes. The final or global weight matrix w_{ni} is formulated as

$$w_{ni} = w_{nij} \cdot w_j \tag{13}$$

Here final or global weight of each alternative can be computed as

$$w_{ni} = \sum_{j=1}^n w_{nij} \cdot w_j \tag{14}$$

In this paper $n = 3$. Hence, the relay node with the highest weight is selected as the best relay to forward the decoded data to SD.

5 Results and Discussion

Authors describe the relay selection AHP process in cognitive radio networks using the following example as shown below;

$$A = \begin{matrix} & \begin{matrix} \text{RLQ} & \text{Reliability} & \text{SINR}_{SD} \\ (\alpha) & (\beta) & (\gamma) \end{matrix} \\ \begin{matrix} \alpha \\ \beta \\ \gamma \end{matrix} & \begin{pmatrix} 1 & 1/2 & 1/4 \\ 2 & 1 & 1/2 \\ 4 & 2 & 1 \end{pmatrix} \end{matrix} \tag{15}$$

where the three decision factors are expressed by α , β , and γ . In this paper, authors consider $SINR_{SD}$ is the most important decision factor, reliability is the second significant decision factor, and RLQ is the third significant decision factor or criterion. The relative significance of the decision factors is $\alpha:\beta:\gamma = 1:2:4$. Hence, the pairwise comparison of three possible pairs are expressed as $\alpha:\beta = 1:2$, $\beta:\gamma = 1:2$, and $\gamma:\alpha = 4:1$. The eigenvector is calculated as $W_j^T = \{0.1429, 0.2857, 0.5714\}$ based on the above-mentioned assumption, which shows the local weights of RLQ, reliability, and $SINR_{SD}$ correspondingly. The eigenvalue $\lambda_{max} = 3.0$ can be also calculated according to the Eq. 11. Therefore, the value of consistency ratio can be calculated as $C.R. = 0.0 < 0.1$ which indicates that matrix A is a perfectly consistent. Relative importance of relay nodes determines the weight matrixes of alternative relays under the three decision factors. Finally, each alternative relay's weight is calculated based on Eq. 14. Thus, the relay node with the highest weight (relay 2 with weight 0.4823) is selected as the best relay (see Fig. 4a). The amount of latency which is introduced by the decision-making process need to be kept as minimum as possible. In Fig. 4b, the simulation of average latency is shown as a function of alternative relay nodes. This depicts a significant low latency in CR networks.

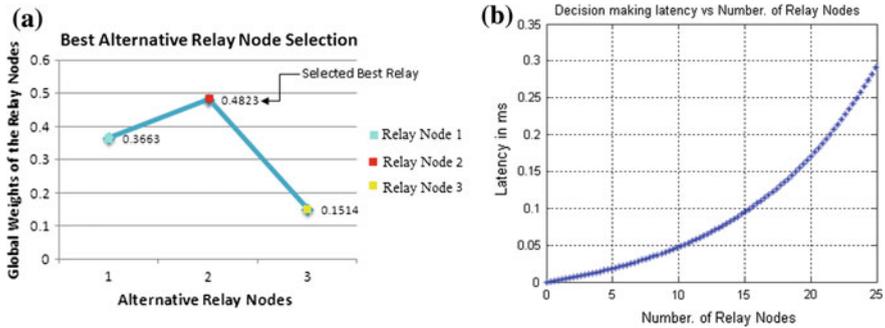


Fig. 4 a Best alternative relay node selection. b Decision-making latency

6 Conclusion

AHP-based decision-making technique for best relay selection considers three input parameters along with their individual weights as per their relative significance in the problem and thus is a practicable solution to find out the best relay for secondary data transmission. The simulation software programs for the proposed system are very simple and do not consume much time to respond.

References

1. Banerjee, J.S., Chakraborty, A.: Fundamentals of software defined radio and cooperative spectrum sensing: a step ahead of cognitive radio networks. In: Kaabouch, N., Hu, W. (eds.) Handbook of Research on Software-Defined and Cognitive Radio Technologies for Dynamic Spectrum Management, pp. 499–543. IGI Global, USA (2015)
2. Banerjee, J.S., Chakraborty, A.: Modeling of software defined radio architecture and cognitive radio, the next generation dynamic and smart spectrum access technology. In: Rehmani, M.H., Faheem, Y. (eds.) Cognitive Radio Sensor Networks: Applications, Architectures, and Challenges, pp. 127–158. IGI Global, USA (2014)
3. Banerjee, J.S., Chakraborty, A., Karmakar, K.: Architecture of cognitive radio networks. In: Meghanathan, N., Reddy, Y.B. (eds.) Cognitive Radio Technology Applications for Wireless and Mobile Ad Hoc Networks, pp. 125–152. IGI Global, USA, (2013)
4. Banerjee, J.S., Karmakar, K.: A comparative study on cognitive radio implementation issues. *Int. J. Comput. Appl.* **45**(15), 44–51 (2012)
5. Banerjee, J.S., Chakraborty, A., Chattopadhyay, A.: Fuzzy based relay selection for secondary transmission in cooperative cognitive radio networks. In: Proceedings of OPTRONIX, Springer, India (2016)
6. Zou, Y., Zhu, J., Zheng, B., Yao, Y.D.: An adaptive cooperation diversity scheme with best-relay selection in cognitive radio networks. *IEEE Trans. Signal Process.* **58**(10), 5438–5445 (2010)

7. Chakraborty, A., Banerjee, J.S Chattopadhyay, A.: Non-uniform quantized data fusion rule alleviating control channel overhead for cooperative spectrum sensing in cognitive radio networks. In: Proceedings of IACC. IEEE (2017)
8. Chakraborty, A., Banerjee, J.S.: An advance Q learning (AQL) approach for path planning and obstacle avoidance of a mobile robot. *Int. J. Intell. Mechatron. Robot.* **3**(1), 53–73 (2013)
9. Saaty, T.L.: How to make a decision: the analytic hierarchy process. *Eur. J. Oper. Res.* **48**(1), 9–26 (1990)

Mutual Information-Based Hierarchical Band Selection Approach for Hyperspectral Images

Sonia Sarmah and Sanjib Kumar Kalita

Abstract Hyperspectral images consist of hundreds of spectral bands with relatively narrow bandwidth and hence records detailed information of the objects. Due to this detailed and enormous amount of information content, the use of hyperspectral images has become very popular in various fields such as land cover monitoring, agriculture, defense, etc. However, this increased spectral dimension results in increased computational complexity. Hence, the selection of minimal subset of spectral bands to represent the actual information effectively without much degradation is a challenge in the field of hyperspectral image analysis. This paper proposes a hierarchical band selection approach by constructing a spectral partition tree-based on mutual information. Initially, each spectral band has been considered as a leaf node. To minimize the redundancy of information carried by neighboring bands, in every level, new nodes are created by merging adjacent bands or group of bands, for which mutual information has been used as the deciding criterion. Finally from each group of bands, a representative band is selected which jointly form the set of selected bands. Experiment is carried out on the AVIRIS Indian Pines dataset by designing training and testing samples containing only the selected set of bands. The experimental results of the proposed method are found to be very promising and competitive with the existing techniques.

Keywords: Hyperspectral images · Spectral partition tree · Entropy
Mutual information · Support vector machine

S. Sarmah (✉) · S.K. Kalita
Department of Computer Science, Gauhati University, Guwahati, India
e-mail: sarmahsonia07@gmail.com

S.K. Kalita
e-mail: sanjib959@rediffmail.com

1 Introduction

Hyperspectral sensors capture images with narrow and contiguous spectral bands covering spectrum not only from the visible range, but also from the ultra violet and infra red region. The resultant datasets are three-dimensional which are represented as data cubes of size $P \times Q \times N$, where P and Q are the spatial dimensions and N is the spectral dimension. Each spatial plane can be viewed as collection of two-dimensional scenes containing $P \times Q$ pixels, each taken at a specific wavelength λ and each pixel can be viewed as a vector consisting of N reflectance values. One such sensor is the airborne visible/infrared imaging spectrometer (AVIRIS) that captures images with up to 224 spectral bands ranging from 400 to 2500 nm [1]. With such high spectral dimension, much more detailed and discriminative information of the objects can be acquired which results in increased classification accuracy. However, performance of many supervised classification methods get strongly affected by increased dimensionality. Hence, spectral dimension reduction is a crucial step in hyperspectral image analysis. Though the hyperspectral images capture a wide range of spectrum, not all bands are equally important for classification purpose as each band does not contribute equally toward the discrimination of the objects. Some bands contain very little or no relevant information. Moreover, the adjacent bands often share redundant information. Thus, this is important to remove such redundant and noisy information before going to the classification phase.

In literature, many dimension reduction techniques are available. Techniques like PCA and ICA map the higher dimensional feature space in to lower dimensional feature space by carrying out linear or nonlinear transformations [2, 3]. In such transformations, original interpretation of spectral data gets compromised. Feature selection is another method for dimension reduction where only the most salient features are selected. This type of methods can broadly be divided into two categories (1) wrappers and (2) filters [4]. In wrappers, a learning algorithm is applied to examine the utility of the features of the dataset. As feature selection is highly dependent on the learning algorithm to be applied, in wrappers it is difficult to switch among the learning algorithms. Moreover, for large databases containing many features, wrappers may become intractable as each and every potential subset of features is evaluated by the learning algorithm. In filters, subset of features is selected as a preprocessing step irrespective of the learning algorithm to be applied and hence is relatively more general compared to wrappers [5, 6].

In feature selection techniques, the main objective is to select the set of features which carry maximum information about the class and yet with minimum correlation between each other. In [2, 7], authors have shown that the adjacent bands of the hyperspectral images are highly correlated and they have utilized the correlation effectively to reduce the number of bands. However in these schemes, the correlation with the class was not considered. In our proposed algorithm, to avoid redundant information carried by adjacent bands, initial band grouping phase is carried out. In this phase, bands are partitioned into K ($K \leq N$) groups by

constructing a bottom up spectral partition tree by recursively merging adjacent spectral bands with redundant information. In the second and final phase, a band is selected from each group as representative which jointly forms the filtered set of selected bands. Both the phases are discussed in Sect. 3.

2 Entropy and Mutual Information-Based Band Selection Scheme

2.1 Entropy

In information theory, entropy is the measure of information content of a random variable in terms of uncertainty. Let X be a discrete random variable with probability distributions $p(x)$, where $x \in X$, the entropy is defined by

$$H(X) = - \sum_{x \in X} p(x) \log p(x) \quad (1)$$

In literature, entropy has been used as band selection criteria for hyperspectral images [2, 3, 8, 9]. Entropy of each spectral band is calculated to measure their information content. The one with the higher entropy values are selected for classification purpose. As only the bands with highest information contents are selected, the classification accuracy is often high. However, the measured information content by entropy suffers from lack of reference or objective. The information content in some bands, though high, may not have any relevance to the target classification.

2.2 Mutual Information (MI)

Mutual information is another technique used to evaluate the effectiveness of a spectral band. Unlike entropy, in mutual information-based approaches apart from measuring the information content of the individual spectral bands, the relevance of the information with the target (reference) image is also considered [10, 11]. In information theory, MI measures the mutual dependence between two random variables. Mathematically, if X and Y are two discrete random variables with probability distributions $p(x)$ and $p(y)$, joint probability distribution $p(x, y)$, where $x \in X$ and $y \in Y$, then MI is calculated as [12],

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (2)$$

$$= H(X) - H(X|Y) \tag{3}$$

$$= H(X) + H(Y) - H(X, Y) \tag{4}$$

Here, $H(X)$ and $H(Y)$ are the entropies of X and Y , respectively. $H(X|Y)$ is the conditional entropy of X with respect to Y and $H(X, Y)$ is the joint entropy of X and Y . Equations (5, 6) represents conditional and joint entropy.

$$H(X|Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x|y) \tag{5}$$

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y) \tag{6}$$

In [11], authors have calculated MI of each band in the dataset with the corresponding reference image to measure the effectiveness of the spectral bands in terms of information content about the various classes (Fig. 1).

3 Proposed MI-based Hierarchical Band Selection Approach

In this paper, we have proposed a hierarchical band selection scheme based on mutual information. The algorithm is divided into two phases: (i) band grouping and (ii) band selection. Figure 2 shows the proposed MI-based band grouping and selection method.

3.1 Band Grouping Phase

The binary partition tree (BPT) in hyperspectral images was introduced in [12, 13] to merge regions based on spatial features. Inspired by the BPT concept, in the proposed algorithm we have constructed a spectral partition tree (SPT) by iterative bottom up merging of the spectral bands. The individual bands form the initial

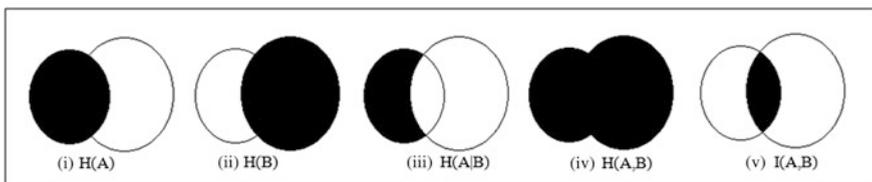


Fig. 1 Diagrammatic representation of entropy and mutual information

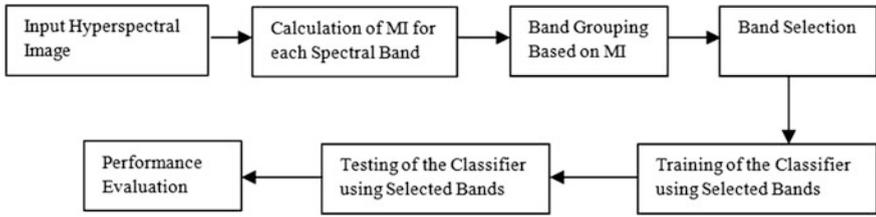
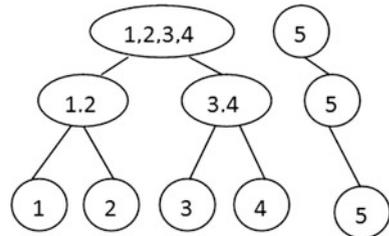


Fig. 2 Block diagram of the proposed method

Fig. 3 Example of construction and representation of the spectral binary tree



nodes, i.e., the leaf nodes of the SPT. Thereafter, in each iteration the nodes for the upper level are created by merging two adjacent nodes. The merging criterion is decided by the difference of MI values of the spectral bands with the reference image and the complementary threshold T . In the first iteration where the input is the set of N spectral bands of the image, two adjacent bands n and $n + 1$, are merged if the difference $d(n) = |\text{MI}(n) - \text{MI}(n + 1)| < T$. This condition ensures that the adjacent bands which do not carry any significant complementary information are grouped. The new node which is formed as a result of merging is represented in the next level as a set of the merged spectral bands. The nodes which cannot be merged are left as it is for the next iteration. For the rest of the iterations, any two adjacent nodes, which may contain more than one spectral band (as a result of merging in the previous iterations), are merged if $d_{\text{max}}(n) = |\max(\text{MI}(n)) - \max(\text{MI}(n + 1))| < T$ and $d_{\text{min}}(n) = |\min(\text{MI}(n)) - \min(\text{MI}(n + 1))| < T$. Here $d_{\text{max}}(n)$ is the difference between the largest MI values of the n th and $n + 1$ th node respectively. Similarly, $d_{\text{min}}(n)$ is the difference between the smallest MI values of the n th and $n + 1$ th node, respectively. The algorithm stops if two consecutive iterations produce the same set of nodes. Figure 3 depicts an example of construction and representation of spectral binary tree.

3.2 Band Selection Phase

In this phase, a representative band from each group of the K groups formed in band grouping phase is selected. From each group, the band yielding the highest MI value is selected as the group representative. Any representative band having less MI value than a desired threshold θ , is discarded.

Algorithm: Band Selection**Input: band groups formed in phase 1** $i \leftarrow 1$ $S \leftarrow \emptyset$ **while** $i \leq K$ **do**select the band B, with maximum MI value from the i th group.**if** $MI(B) < \theta$ **then do** $S \leftarrow S \cup B$ **else** $S \leftarrow S \cup S$ **end if****end while**

4 Experimental Setup and Results

4.1 Dataset Description

For our experiment, we have used the AVIRIS Indian Pines Dataset which was collected over Northwest Indiana in June 1992 by AVIRIS sensor. The dataset contains the Indian pine image consisting of 220 spectral bands and 21,025 pixels (145×145). The image is accompanied with a reference image where 10,249 pixels are labeled with a number from 1 to 16, denoting the class to which a pixel belongs to. The rest 10,776 pixels are labeled as 0 indicating that these either belong to areas which are not of interest or could not be labeled due to technical difficulties. Out of the 16 classes only nine classes, as listed in Table 1, have been considered for our experiment. Other classes are discarded as the rest have relatively limited number of labeled samples (Fig. 4).

Table 1 Numbers of samples for different classes in Indian pine image dataset

S. No.	Class	Number of samples
1	Corn-no-till	1428
2	Corn-min-till	830
3	Grass-Pasture	483
4	Grass-Trees	730
5	Hay-windrowed	478
6	Soybean-no-till	972
7	Soybean-min-till	2455
8	Soybean-clean	593
9	Woods	1265

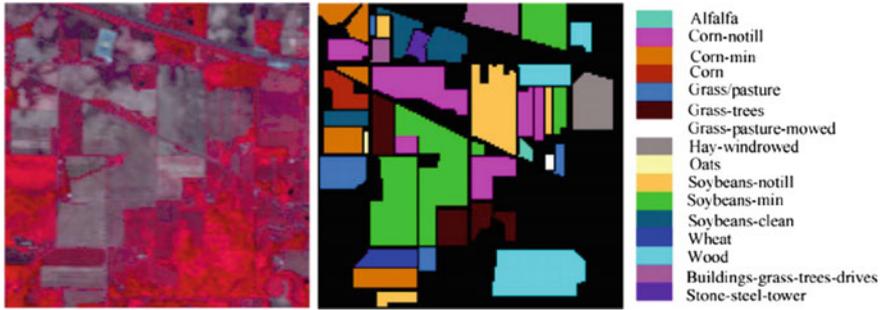


Fig. 4 False color image of the reference image for Indian pines dataset showing the various classes present

4.2 Experiment and Result

Experiment was performed by first calculating the MI of each spectral band of the Indian Pine image with the reference map. Figure 5 shows the calculated results plotted against the band numbers. It can be seen that some band have very low MI values. This is due to water absorption in those bands.

In the band grouping phase, the construction of the spectral binary tree was experimented with different thresholds. For band selection phase, the threshold θ was set as 0.30. Number of groups formed in band grouping phase and number of selected representative bands from the groups for each threshold are listed in Table 2. For classification, support vector machine was used as it is known to work well with high input space. For training and testing, 200 random samples from each class with selected bands were taken. The classifier was also tested with samples from the original Indian pines image (containing 220 bands) and the corrected image (containing 200 bands). Table 2 shows the achieved experimental results.

From the experimental results, it is observed that as we limit the number of bands, the overall accuracy (OA) of classification also decreases. However, this decrease is not that significant and results are still comparable with that when all the bands are used. As the number of bands decreases from 200 to 38, the classification accuracy decreases by only 0.23%. However, when the number of bands further gets decreased to 27 the degradation in classification accuracy is a bit higher. The most significant observation is that with all the 220 bands the achieved OA is lower

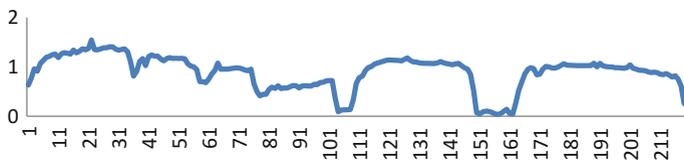


Fig. 5 Mutual information of the spectral bands of the Indian Pie image with the reference image

Table 2 Experimental results of the proposed algorithm

Threshold	No. of groups formed in band grouping phase	No. of representative bands selected	Overall accuracy (%)	Degradation in overall accuracy (%)
None	220 (using all the bands of the dataset)	220	86.55	
None	200 (using the bands of the corrected image)	200	87.06	-0.51
0.10	42	38	86.83	0.23
0.15	31	27	85.67	1.16
0.20	23	20	85.33	0.34
0.25	19	17	84.77	0.56

than that using 38 bands. This is due to the rejection of the groups of bands which have very low MI with the target image. The proposed algorithm discards the bands which are noisy and do not contain any relevant information are avoided. In the corrected image with 200 bands, the noisy bands are already removed and classification accuracy is higher.

5 Conclusion

In this paper, we have proposed a hierarchical band selection approach based on mutual information. A spectral partition tree is constructed to group the adjacent bands carrying redundant information. From each group in the band selection phase, only one band which carries maximum information about the target image is selected to eliminate redundancy. From the results, it has been seen that the bands that get selected through the proposed algorithm, though limited, are highly capable of representing original image without losing much information. Through this algorithm, the noisy and irrelevant bands gets automatically discarded in the band selection phase. When the threshold is high, very few numbers of bands get selected and the achieved overall accuracy of classification is also low. As more number of bands is added to the selected set, the overall accuracy also improves significantly. However, after some limit, this growth in classification accuracy with number of selected bands tends to be slower. The additional accuracy achieved becomes insignificant compared to the increase in spectral dimension.

References

1. AVIRIS: Airborne visible/infrared imaging spectrometer. <http://aviris.jpl.nasa.gov/>
2. Chang, C.-I., Qian, D., Sun, T.-L., Althouse, M.: A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **37**(6), 2631–2641 (1999)

3. Velez-Reyes, M., Jimenez, L.: Subset selection analysis for the reduction of hyperspectral imagery. *IEEE Int. Geosci. Remote Sens. Symp. Proceed.* **3**, 1577–1581 (1998)
4. Kohavi, R., John, G.: Wrappers for feature subset selection. *Artif. Intell.* **97**, 273–324 (1997)
5. Al-Ani, A.: A dependency-based search strategy for feature selection. *Expert Syst. Appl. Int. J.* **36**, 12392–12398 (2009)
6. Bonev, B., Escolano, F., Angel-Cazorla, M.: A novel information theory method for filter feature selection. *MICAI 2007: Advances in Artificial Intelligence*, pp. 431–440. Springer, Berlin (2007)
7. Kumar, S., Ghosh, J., Crawford, M.M.: Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **39** (2001)
8. Groves, P., Bajcsy, P.: Methodology for hyperspectral band and classification model selection. In: *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 120–128 (2003)
9. Bajcsy, P., Groves, P.: Methodology for hyperspectral band selection. *Photogram. Eng. Remote Sens. J.* **70**(7), 793–802 (2004)
10. Conesea, C., Masellia, F.: Selection of optimum bands from TM scenes through mutual information analysis. *ISPRS J. Photogram. Remote Sens.* **48**(3), 2–11 (1993)
11. Guo, B., Gunn, S., Damper, B., Nelson, J.: Adaptive band selection for hyperspectral image fusion using mutual information. In: *7th International Conference on Fusion* (2005)
12. Novovicova, J., Somol, P., Haindl, M.: Conditional mutual information based feature selection for classification task. In: *CIARP 2007*, vol. 4756, pp. 417–426, LNCS (2007)
13. Valero, S., Salembier, P., Chanussot, J.: New hyperspectral data representation using binary partition tree. In: *Proceedings of IEEE IGARSS Conference*, pp. 80–83 (2010)

Petri Net: Design and Analysis of Parallel Task Scheduling Algorithm

Sasmita Parida, Suwendu Chandan Nayak, Piyush Priyadarshi,
Prasant Kumar Pattnaik and Gaurav Ray

Abstract In real time, most of the tasks are deadline based. The deadline-based task has different parameters: arrival time, start time, execution time, and deadline. Many performance-based task scheduling algorithms are proposed by number of researchers theoretically. But due to change of implement environment, the performance varies. Petri net is a graphical and mathematical model to evaluate and analysis of the system. In Petri net, conflicts are occurred during firing. In this paper, we designed and modeled the Petri net for scheduling deadline-based task by resolving the conflicts. We also proposed a scheduling mechanism and firing rules to schedule deadline-based tasks. The designed model increases the resource utilization of a physical system in cloud computing. The performance of the proposed model is analyzed using the PIPE v4.3.0. We analyzed the reach ability graph, convertibility graph, and steady-state analysis of the model.

Keywords Task · Deadline · Scheduling · Backfilling · Petri net
Reach ability · Convertibility

S. Parida (✉) · S.C. Nayak · P. Priyadarshi · G. Ray
Department of Computer Science and Engineering, C.V. Raman College of Engineering,
Bhubaneswar, India
e-mail: sasmitamohanty5@gmail.com

S.C. Nayak
e-mail: suwendu2006@gmail.com

P. Priyadarshi
e-mail: priyadarship4@gmail.com

G. Ray
e-mail: gouravrayray@gmail.com

P.K. Pattnaik
School of Computer Engineering, KIIT University, Bhubaneswar, India
e-mail: patnaikprasant@gmail.com

1 Introduction

Since the last few years, task scheduling in cloud computing comes forward as a challenging and interesting area for researchers. A better resource utilization can be achieved by adapting new and modified task scheduling algorithms. The basic concept of task scheduling is to schedule all the tasks within the deadline with limited resources. Optimal scheduling problem is known as NP-hard problem [1]. In cloud computing, the pool of resources is distributed geographically and is available for the users on pay for use basis. Resources are allocated or released according to the requirement of user applications at any time and charged. However, resource under-provisioning usually decreases the system's performance, whereas resource over-provisioning always leads to idle resources [2]. Therefore, it is a challenge for cloud providers to provide the exact amount of cloud resources to the user with minimum cost.

The OpenNebula is one of the open source cloud platforms which provides a flexible platform for both cloud users and cloud service providers. The OpenNebula supports basically three types of leases: (1) Advance reservation, (2) Best Effort Leases, and (3) Immediate Leases [3]. The user request is called task or lease in OpenNebula. In real time, when a time constraint is associated with the best effort lease, it is called deadline sensitive leases [4, 5]. Many algorithms are used to schedule deadline sensitive leases like swapping and backfilling algorithm [4, 6]. Backfilling algorithm performs better than swapping. In case of swapping algorithm leases are non-preemptive, whereas in backfilling algorithm leases are preemptive in nature. It provides better resource utilization than swapping by allocating more number of leases. Moreover, different types of backfilling algorithms are proposed by researchers like: *EASY* Backfilling, *Conservative* Backfilling, *Slack-based* Backfilling, *Selective* Backfilling, *Relaxed* Backfilling, and *Multiple-queue* Backfilling.

The Petri net is one of the power mechanisms which provides the representation and analysis of concurrent systems [7]. The representation specifies the nature and work flow of the system where as, analysis specifies the performance of the system in real time. Resource allocation system (RAS) is mostly based on the analysis of the system, control techniques that are used to evaluate the performance [8]. The Petri net models can be used to analyze interdependencies among the tasks, criticality which is occurred during system implementation. Moreover, it can also handle substitution, conflicting resource priorities which is a common scenario. In real time the available resources are in different variants, so it is more challenging to allocate resources to the users. The above challenges can be handled by Petri net [7]. The mechanism for conflict challenges in resource allocation is proposed using AHP [9] when there is more than one similar tasks or having similar priorities.

In this work, we proposed a Petri net model for scheduling deadline-based tasks, where the tasks are scheduled in parallel. The aim of the paper is to design a model

using Petri net which can schedule tasks by allocating with required VMs and observe the performance. The rest part of the paper is organized as: The related work are discussed in Sect. 2, The problem formulation and proposed model using Petri net is available in Sects. 3 and 4 presents the result and discussion of the proposed model. The conclusion and future scope is presented in Sect. 5.

2 Related Works

Resource allocation is a challenging issue in cloud computing [10]. In a physical machine, different virtual machines (VMs) are created according to the user request. A single user may demand number of virtual machines [11]. In our previous work [5], we discussed backfilling algorithm. In this paper, we observed some disadvantages and proposed a mechanism for truthful resource allocation in cloud computing. Calheiros and Buyya [12] also proposed a mechanism for task scheduling and task replication for deadline-based task. A resource cost optimization technique is proposed by Chaisiri et al. [13] in cloud computing environment. In recent year [14], an energy aware scheduling is also proposed. In this work, the authors proposed a mechanism to reduce the energy in cloud computing by efficient task scheduling. Kalra and Singh [15] discussed different metaheuristic approach which is used for task scheduling in cloud computing. The work provides the basic review of metaheuristic approaches in cloud computing.

The task replication is proposed by Calheiros and Buyya [12]. The tasks are replicated if sufficient resources are not available within the deadline. Here the tasks are associated with a deadline. Ergu et al. [16] proposed a scheduling algorithm using the analytical hierarchy process in cloud computing. The work performs better resource utilization as compared to other scheduling algorithm for deadline-based task. Petri net is a basic model of parallel and distributed systems, designed by Carl Adam Petri in 1962 [7]. The author used it for resource allocation in the project. Moreover, Petri net is also used for server consolidation for heterogeneous computing in cloud computing environment [17]. Petri nets are also used to solve multi criteria fuzzy rules by evaluating priorities for the conflicting process. In the real time information based system state is used to solve conflicts using Petri nets [18].

Petri net is used to analyze and resolve the conflicts in a system. In this work, we try to increase resource allocation performance. The system is modeled using Petri net which allocated the required VMs. We also discussed how the different tasks are fired within their deadline with their required resources. We also observed some of the conflicts occurred during the system modeling and modified the firing principles. The detail of the proposed work is followed in the next section.

3 Proposed Model

3.1 Background of Petri net

As we discussed in Sect. 2, Petri net is a graphical and mathematical model for system to evaluate and analysis. Formally, a Petri net is a 5-tuple as [7]:

$$PN = (P, T, A, W, M)$$

where

$P = \{p_1, p_2, \dots, p_M\}$ is a finite set of places.

$T = \{t_1, t_2, \dots, t_N\}$ is a finite set of transitions.

$A \subseteq (P \times T) \cup (T \times P)$ is a set of arcs.

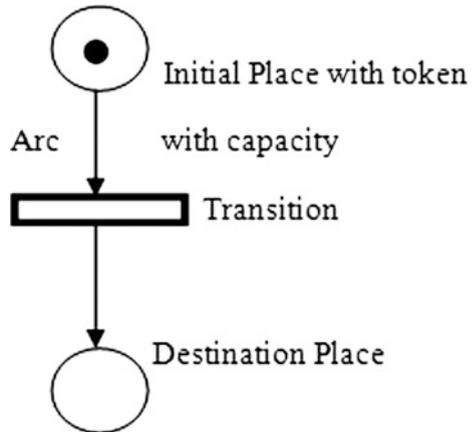
$W : A \rightarrow \{1, 2, 3, \dots\}$ is a weight function.

$M : P \rightarrow \{0, 1, 2, \dots\}$ is the set of initial markings.

where $P \cap T = \emptyset$. and $P \cup T \neq \emptyset$.

The transition (firing) rules are applicable in untimed Petri nets for simulating the dynamic behavior of a system. Figure 1 shows the different states used in Petri net. There is an initial position (outer circle) which contains a token (filled circle). An initial position may contain more than one token as per the system design requirement. There is an arc from the initial position to transition. It holds the weight or capacity. Based upon the firing rules the transition is occurred from the initial state to the destination state.

Fig. 1 States of a Petri net



3.2 Task Scheduling Model

As we discussed in Sect. 1, the aim of the work is to model the task scheduling algorithm using Petri net and to observe the performance. Here, we considered backfilling algorithm to model using Petri net. During modeling, we changed the firing rules and implemented some additional steps in Petri net. These steps are added in the algorithm. In Open Nebula, user request is represented using following parameters: lease number (Task serial number), node (Number of VMs), arrival time, Start time, duration sand deadline. So a task T can be represented as

$$T = \{n, A, N, S, E, D\}$$

In OpenNebula, user needs the number of VMs. So the challenge is that how all the leases or tasks could be scheduled within their deadline. As we discussed, the backfilling algorithm provides better scheduling of tasks along with better resource utilization. In backfilling algorithm, all the tasks are sorted according to their start time. The first task is scheduled first. Then the free nodes are evaluated and another task is selected from the queue whose start time is same with the first task where the required node is less than or equal to the free nodes. The existing backfilling is used for parallel scheduling. In this paper, we modeled it in a single physical machine. The VMs are set as weight to fire transitions. The weights are adjusted by which most of the tasks are scheduled. Moreover, we analyze the performance of proposed model with different parameters that is discussed in Sect. 4.

Step 1: *Initialize queue (Q) with n number of leases or tasks with their start time*
 Step 2: *Schedule 1st task from Q*
 Step 3: *Select a task T_i which can be scheduled parallel with first task as*
 If $T_i(S) == T_1(S)$ and
 $T_i(N) \leq N_{free}$ and meet its deadline at time slot t
 Step 4: *Schedule the next task at time slot t_i where N is maximum and S is equal.*
 And repeat Steps 3 to 4 until the Q is empty.
 Else
 Schedule next task with different S and repeat Steps 3 to 4 until Q is empty
 Step 5: *end*

The above backfilling algorithm is formulated in the Petri net with five tuple form as

$$PN = \{P, T, I, O, M\}$$

where

- P task
- T number of VMs required by P
- I input function with start time S
- O output function
- M initial marking at time t .

The transition firing is based on transition enable initiated by input function and fire rules by which tasks are scheduled within the deadline. The VMs on the physical machine goes from busy state to free state and vice-versa. The proposed algorithms based on the earliest deadline of the task, number of VMs required by the task, and the duration of the execution.

3.2.1 Proposed Steps

- Step 1 *Select the task according to its earliest deadline.*
- Step 2 *Set the required VMs as the weight of arc.*
- Step 3 *Put number of VMs required by that task in put stage.*
- Step 4 *If resource available, fire transition.*
- Step 5 *Move allocated VMs to busy state from free state.*
- Step 6 *Execute lease for predefined time.*
- Step 7 *Check the available VMs in initial state.*
- Step 8 *Fire another transition where it meets the required at the same start time*
- Step 9 *All VMs go to free state.*
- Step 10 *Repeat Steps 1 to 8 until reach to maximum deadline.*

3.2.2 Proposed Rules for Firing

Rule 1. Select transition delay (T_{dl}) as minimum execution time (E_i) of the tasks in queue.

$$T_{dl} = \min(E_i)$$

Rule 2. If $E_i > T_{dl}$, return the task to original place else move it out.

The Petri net model is varied for different set of tasks. The output is also changed basis upon the type of Petri net used. In this work, the above task scheduling algorithm is modeled using timed Petri net. Figure 2 shows the proposed model for backfilling algorithm in Petri net. In each stage of the proposed model, VM has two

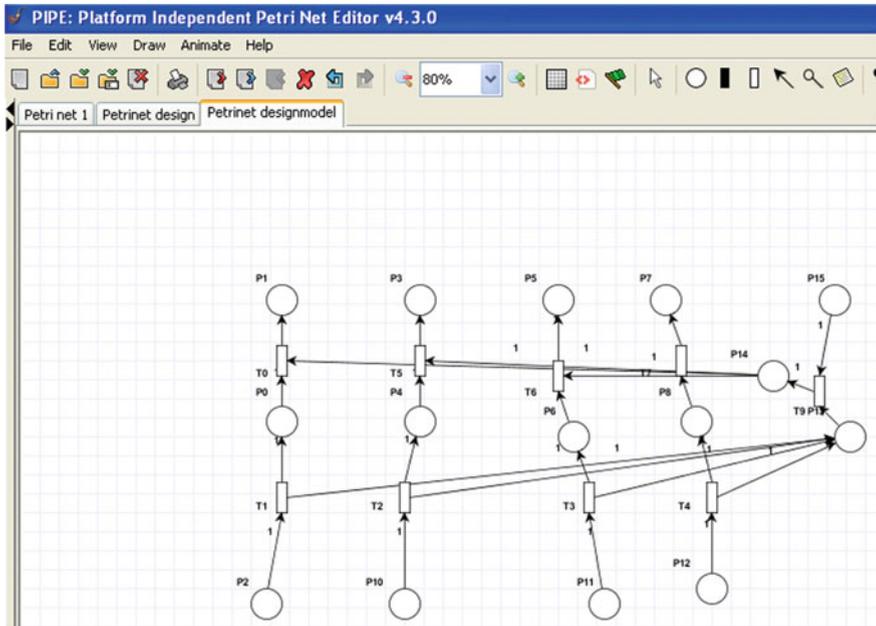


Fig. 2 Proposed Petri net model

states: busy state and free state. When the VM is allocated to any task, it goes to busy state for the time duration of the execution duration of the task. When the task is completed by the VM goes to free state. For better understanding, we considered the below illustration.

In the model, there are number of inputs and number of outputs. This is to only for designing the model that supports parallel execution of tasks. We used one control input where maximum tokens are available. The tokens are nothing but the number of VMs possible in physical machine. Moreover, there are some intermediate nodes in between input layer and output layer. The inputs, outputs, intermediate, and control node is connected through arcs having different weights. Whenever the task is executed, a number of the transitions is made called tangible states. It is an interconnected network that looks like a directed graph.

4 Simulations and Result Analysis

In this work, we designed the model using PIPE (Platform Independent Petri Net Editor) which is an open source Petri net tool. The simulation results are varied from problem to problem. To observe the performances of the proposed model, we considered the below illustration.

4.1 Illustration

Table 1 shows the tasks information. There are four tasks with maximum node requirement four. The tasks are having their start time, duration and deadline. Though $Max(N)$ is 4, so in the physical machine four VMs will be created to schedule these tasks. The tangible states of the illustration are shown in the Fig. 3. There are 16 process which are computing as p_0, p_1, \dots, p_{15} .

Table 1 Information of leases or tasks

Task no.	Nodes	Submit time (AM)	Start time (PM)	Duration	Deadline (PM)
1	2	11.10	12.00	20	12.30
2	3	11.20	12.00	40	01.00
3	2	11.30	12.00	30	01.50
4	4	11.40	01.00	20	01.50

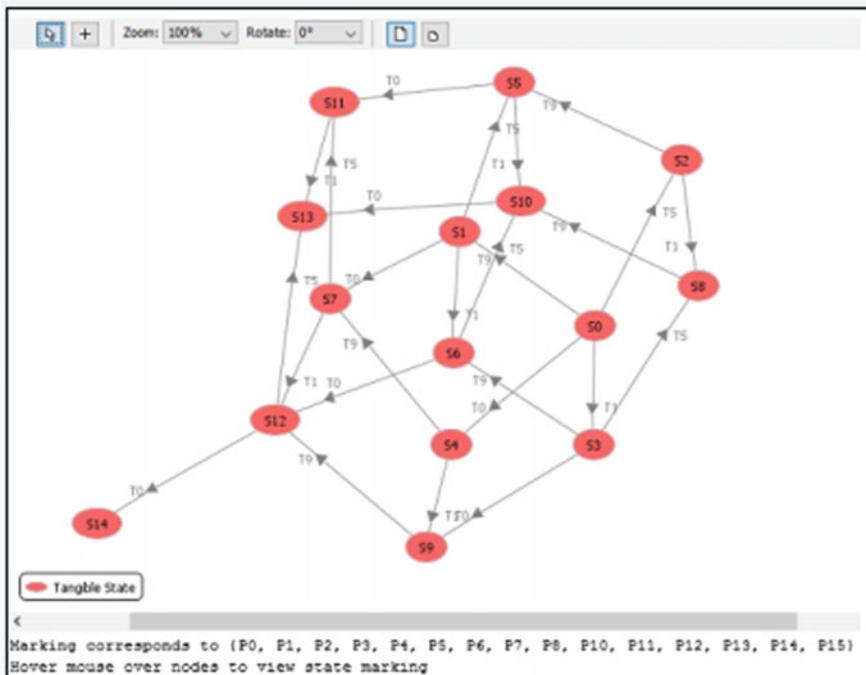


Fig. 3 Tangible states of proposed model for illustration

4.2 Performance

Moreover, the performance can be observed in many ways in PIPE. In this work, we considered the following parameters of PIPE such as: forward incidence matrix, backward incidence matrix, GSPN (general stochastic Petri net) steady-state analysis, *P* invariants, and simulation result. There are two types of matrixes used among the process and transitions, that is forward incidence matrix and backward incidence matrix [19] as shown in Fig. 4a, b.

The incidence matrix represents the communication among the process and the transitions in the designed Petri net model. Here 16 process and 10 transitions are occurred. Similarly, P-invariant is created for the model which is shown in the Fig. 5. P-invariant is used to know the model is positive or negative. If the P-invariant is positive, it means it is bounded and when it is unbounded then the P-invariant is negative. The steady-state analysis specifies the states of the process of the model in Petri net. Steady-state specifies the behavior of the system or process that is unchanging in time. The steady-state of the system is shown in the Fig. 6, which specifies that in the system two or more reversible process occurs at the same time.

(a)		(b)																	
	T0	T1	T2	T3	T4	T5	T6	T7	T9		T0	T1	T2	T3	T4	T5	T6	T7	T9
P0	0	1	0	0	0	0	0	0	0	P0	1	0	0	0	0	0	0	0	0
P1	1	0	0	0	0	0	0	0	0	P1	0	0	0	0	0	0	0	0	0
P2	0	0	0	0	0	0	0	0	0	P2	0	1	0	0	0	0	0	0	0
P3	0	0	0	0	0	1	0	0	0	P3	0	0	0	0	0	0	0	0	0
P4	0	0	1	0	0	0	0	0	0	P4	0	0	0	0	0	1	0	0	0
P5	0	0	0	0	0	0	1	0	0	P5	0	0	0	0	0	0	0	0	0
P6	0	0	0	1	0	0	0	0	0	P6	0	0	0	0	0	0	1	0	0
P7	0	0	0	0	0	0	0	1	0	P7	0	0	0	0	0	0	0	0	0
P8	0	0	0	0	1	0	0	0	0	P8	0	0	0	0	0	0	0	1	0
P10	0	0	0	0	0	0	0	0	0	P10	0	0	1	0	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	P11	0	0	0	1	0	0	0	0	0
P12	0	0	0	0	0	0	0	0	0	P12	0	0	0	0	1	0	0	0	0
P13	0	1	1	1	1	0	0	0	0	P13	0	0	0	0	0	0	0	0	1
P14	0	0	0	0	0	0	0	0	1	P14	1	0	0	0	0	1	1	1	0
P15	0	0	0	0	0	0	0	0	0	P15	0	0	0	0	0	0	0	0	1

Fig. 4 a Forwards incidence matrix. b Backwards incidence matrix

P-Invariants

P0	P1	P2	P3	P4	P5	P6	P7	P8	P10	P11	P12	P13	P14	P15
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1	1	0	0	1	0	0	0
0	1	1	1	0	1	0	1	0	1	1	1	1	1	0
0	1	0	1	0	1	0	1	0	0	0	0	0	1	1

The net is covered by positive P-Invariants, therefore it is bounded.

Fig. 5 P-invariant

GSPN Steady State Analysis Results

Set of Tangible States

	P0	P1	P2	P3	P4	P5	P6	P7	P8	P10	P11	P12	P13	P14	P15
M0	1	0	1	0	1	1	0	1	0	0	0	0	0	2	0
M1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0
M2	2	0	0	0	1	1	0	1	0	0	0	0	1	2	0
M3	0	1	1	0	1	1	0	1	0	0	0	0	0	1	0
M4	2	0	0	1	0	1	0	1	0	0	0	0	1	1	0
M5	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0
M6	1	1	0	0	1	1	0	1	0	0	0	0	1	1	0
M7	1	1	0	1	0	1	0	1	0	0	0	0	1	0	0
M8	0	2	0	0	1	1	0	1	0	0	0	0	1	0	0

Fig. 6 Steady-state analysis

4.3 Results

Moreover, the simulation result is shown in Fig. 7 of the designed Petri net model. The result depends upon the number of simulation and iteration. In Figure 7, one iteration for one simulation is shown. As we modeled, parallel scheduling algorithm number of out puts are came out. In the proposed model, $p_1, p_3, p_5,$ and p_7 are the output process. When we run the simulator, we need these tasks should be scheduled in parallel. In simulation, the number of VMs is represented by number of tokens.

In Figure 7, we can observe that only the process p_1, p_5 and p_7 contains 1, which specifies these tasks are scheduled in parallel. Similarly, the simulation also shows the confidence intervals of the process when the simulation was executed. It also shows all the processes with their average number of tokens in the simulation.

Petri net simulation results

Place	Average number of tokens	95% confidence interval (+/-)
P0	0.25	0.53676
P1	1	0.24005
P2	0.5	0.39199
P3	0.75	0.36667
P4	0.25	0.36667
P5	1	0
P6	0	0
P7	1	0
P8	0	0
P10	0	0
P11	0	0
P12	0	0
P13	0.25	0.39199
P14	0.25	0.39199
P15	0	0

Fig. 7 Simulation result

Moreover, the simulation result can be changed for second time simulation to schedule another set of tasks. The average number tokens and confidence interval will be changed according to the problem and nature of tasks.

5 Conclusions and Future Work

In this proposed work, we modeled existing backfilling algorithm using Petri net. The existing backfilling algorithm is used for scheduling deadline-based tasks in OpenNebula. The performance of the backfilling algorithm is observed by different performance parameters as shown in Figs. 4, 5, 6, and 7. We assigned a number of VMs to the tasks as per their requirements. The proposed model schedules the tasks by which better resource utilization is achieved. This work provides an idea to design task scheduling algorithm for resource in Petri net and to observe their performances. Moreover, other performance parameters of Petri net can be observed for scheduling algorithms in different scenario in future. In spite of this, more models can be designed for task scheduling and can be simulated in Petri net in future.

References

1. Alkhanak, E.N.: Cost optimization approaches for scientific workflow scheduling in cloud and grid computing: a review, classifications, and open issues. *J. Syst. Softw.* **113**, 1–26 (2016). <http://www.sciencedirect.com/science/article/pii/S0164121215002484>
2. Li, X., Cai, Z.: Elastic resource provisioning for cloud workflow applications. *IEEE Trans. Autom. Sci. Eng.* 1–16 (2015). <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7352380>
3. Sotomayor, B.: Resource leasing and the art of suspending virtual machines. In: 2009 11th IEEE International Conference on High Performance Computing and Communications, pp. 59–68 (2009). <http://www.computer.org/portal/web/csd/doi/10.1109/HPCC.2009.17>
4. Nathani, A., Chaudhary, S., Somani, G.: Policy based resource allocation in IaaS cloud. *Future Gen. Comput. Syst.* **28**(1), 94–103 (2012). <http://dx.doi.org/10.1016/j.future.2011.05.016>
5. Parida, S., Nayak, S., Tripathy, C.: Truth Full Resource Allocation Detection in Cloud Computing. WCI, ACM, New York, pp. 487–491
6. Parida, S., Nayak, S.C.: Study of deadline sensitive resource allocation scheduling policy in cloud computing. *Int. J. Comput. Sci. Mobile Comput.* **3**(12), 521–528 (2014)
7. Kumar, K., Ganesh, L.S.: Use of petri nets for resource allocation in projects. *IEEE Trans. Eng. Manage.* **45**(1), 49–56 (1998)
8. Colom, J.: The resource allocation problem in software applications: a Petri Net perspective, pp. 219–233 (2012)
9. Nayak, S.C., Tripathy, C.: Deadline sensitive lease scheduling in cloud computing environment using AHP. *J. King Saud Univ. Comput. Inform. Sci.* (2016). <http://dx.doi.org/10.1016/j.jksuci.2016.05.003>
10. Di, S., Kondo, D., Wang, C.L.: Optimization of composite cloud service processing with virtual machines. *IEEE Trans. Comput.* **64**(6), 1755–1768 (2015)
11. Sotomayor, B.: Capacity leasing in cloud systems using the OpenNebula engine. *Most*, 2008, pp. 1–5 (2008). <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Capacity+Leasing+in+Cloud+Systems+using+the+OpenNebula+Engine#0>
12. Calheiros, R.N., Buyya, R.: Meeting deadlines of scientific workflows in public clouds with tasks replication. *IEEE Trans. Parallel Distrib. Syst.* **25**(7), 1787–1796 (2014)
13. Chaisiri, S., Member, S.: Optimization of resource provisioning cost in cloud computing. *IEEE Trans. Serv. Comput.* **5**(2), 164–177 (2012)
14. Duan, H.: Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems. *Future Gen. Comput. Syst.* (2016). <http://linkinghub.elsevier.com/retrieve/pii/S0167739X16300292>
15. Kalra, M., Singh, S.: A review of metaheuristic scheduling techniques in cloud computing. *Egypt. Inform. J.* **16**(3), 275–295 (2015). <http://linkinghub.elsevier.com/retrieve/pii/S1110866515000353>
16. Ergu, D.: The analytic hierarchy process: task scheduling a resource allocation in cloud computing environment. *J. Supercomputing*, 835–848 (2011)
17. Al-azzoni, I.: Server consolidation for heterogeneous computer clusters using colored petri nets and CPN tools. *J. King Saud Univ. Comput. Inform. Sci.* **27**(4), 376–385 (2015). <http://dx.doi.org/10.1016/j.jksuci.2015.02.001>
18. Kato, E.R., Morandin, O., Sgavioli, M.: A conflict solution manufacturing system modeling using fuzzy coloured Petri Net. In: 2010 IEEE International Conference on Systems Man and Cybernetics, pp. 3983–3988 (2010)
19. Amer-Yahia, C., Zerhouni, N., Moudni, A.E., Ferney, M.: Some subclasses of petri nets and the analysis of their structural properties: a new approach. *IEEE Trans. Syst. Man Cybern.* **29**(2), 164–172 (1999)

Opportunistic Selection of Threshold Using Double Threshold in Cognitive Radio Networks

G. Soumya and G. Verma

Abstract In cognitive radio selecting, the value of threshold is one of the important and challenging works. This selection is done mainly using two principles. First principle is constant detection rate (CDR), in which probability of detection is taken as constant. Second principle is constant false alarm rate (CFAR), in which probability of false alarm rate is taken as constant. From these two principles, CDR is preferable to give sufficient protection of licensed users. Here double threshold technique, in which two threshold levels are present, is used. Interference is more when signal-to-noise ratio is less. Implementation of double threshold method, according to the farness of secondary transmitter from primary receiver gives some better results to reduce that problem. In this paper, we proposed an approach by taking the distance of secondary transmitter from primary receiver as a parameter, double threshold and single threshold along with CDR principle.

Keywords Cognitive radio · Energy detection · CDR principle
Distance · Signal-to-noise ratio · Double threshold

1 Introduction

As the numbers of wireless users are increasing, spreading of new wireless equipments and applications and demand of the radio spectrum is increased. The accommodation of large number of wireless communication equipments with high data rate within an available frequency range is a challenging and difficult task. This leads to the problem of lack of spectrum [1]. Federal Communication Commission

G. Soumya (✉) · G. Verma
Department of Electronics and Communication Engineering,
National Institute of Technology Kurukshetra (NITKRR),
Kurukshetra 136119, Haryana, India
e-mail: gsoumy491@gmail.com

G. Verma
e-mail: gauravnitk13@gmail.com

measurements revealed that some part of the radio spectrum is not utilized [2]. This type of reports creates a situation to think about spectrum reuse. That leads to introduce the cognitive radio. In the cognitive radio, primary users and secondary users are present. Primary users are those to whom spectrum is actually allocated. Primary users are called as licensed users. Secondary users are those who are ready to utilize the spectrum when the spectrum is free from licensed users. Secondary users are called as unlicensed users [3]. Two important parameters which are used in the evaluation of the performance of cognitive radio are probability of detection P_d and probability of false alarm P_{fa} . The value of P_d should be high for the sufficient protection for primary users. The value of P_{fa} should below maximum utilization of licensed channel [2]. The spectrum sensing is an important issue. Energy detection scheme is preferable among the available techniques for spectrum sensing. Because for energy detection scheme, no need to have prior information of primary users and it is simple as it contains less complexity in computations [1].

In the cognitive radio system, threshold selection is done using mainly two principles: principle of constant detection rate (CDR) and principle of constant false alarm rate (CFAR) [1]. In this paper to give sufficient level of protection for primary users, CDR principle is preferred. In CDR, cognitive radio kept P_d as constant and then targets to decrease the value of P_{fa} . The authors in [4, 5] used the CDR principle with one aspect, that is, increasing the achievable throughput. In low SNR region, the results are not good for energy detection scheme because of noise uncertainty [2]. To overcome this problem and to increase the reliability, double threshold detection method is introduced especially in low SNR region. In double threshold method, there are two threshold levels present.

Related to cognitive radio, already so much research have done by considering distance of primary receiver from secondary transmitter is an important entity. In [5], the authors explained a method to find detection threshold by considering distance of primary receiver from secondary transmitter is main variable. In [6], the authors used distance information to regulate the transmission power then unwanted interference automatically reduced to some extent. In [7], the authors considered distance of primary receiver from the secondary transmitter as an important parameter, and proposed probability of interference to increase the utilization of licensed channel.

In this paper, distance of secondary transmitter from primary receiver is considered as main variable. For that, one critical distance should be calculated from [1]. When the distance is less than the critical distance, then primary user communication will get affected [8]. So to reduce the problem to some extent, double threshold is used here. If distance is greater than the critical distance, then single threshold is used.

The rest of this paper contains: Sect. 2 presents overview of energy detection. Section 3 contains overview of double threshold. Section 4 contains proposed model. Section 5 contains simulation and results finally conclusion is presented in Sect. 6.

2 Overview of Energy Detection

The symbols H_0, H_1 represent the hypothesis which is used to indicate whether the primary user is present or absent in a frequency band. These are represented as [9]:

$$H_0 : y[n] = u[n] \tag{1}$$

$$H_1 : y[n] = s[n] + u[n] \tag{2}$$

where $s[n]$ is primary user signal and $u[n]$ is noise signal. $s[n]$ and $u[n]$ are taken as independent and identically distributed. The mean values of both signals are zero. The variances are σ_p^2 and σ_u^2 , respectively. Here $y[n]$ is complete signal gained by the sensing node. The following equation shows test statistic of energy detection [10]:

$$\xi[n] = \frac{1}{N} \sum_{n=1}^N |y[n]|^2 \tag{3}$$

where N is number of samples counted at the node of sensing and formulated as τf_s . Here τ is the sensing time and T is the frame duration and f_s represents sampling frequency. If detection threshold is considered as λ then P_d and P_{fa} are shown as [1]:

$$P_d = P_r(\xi[n] > \lambda | H_1) = \int_{\lambda}^{\infty} p_1(x) dx \tag{4}$$

$$P_{fa} = P_r(\xi[n] > \lambda | H_0) = \int_{\lambda}^{\infty} p_0(x) dx \tag{5}$$

Here $p_0(x)$ and $p_1(x)$ are the probability density functions of $\xi[n]$ under hypothesis of H_0 and H_1 , respectively. If $s[n]$ and $u[n]$ both are Circularly Symmetric Complex Gaussian (CSCG), then P_d and P_{fa} are written as [5, 9]:

$$P_d = Q\left(\left(\frac{\lambda}{\sigma_u^2} - \text{SNR}_p - 1\right) \sqrt{\frac{N}{2\text{SNR}_p + 1}}\right) \tag{6}$$

$$P_{fa} = Q\left(\frac{\lambda}{\sigma_u^2} - 1\right) \sqrt{N} \tag{7}$$

Here SNR_p denotes the PU signal SNR is obtained at the sensing node given by [1]:

$$\text{SNR}_p = \frac{P_p}{\text{noise power}} = \frac{\sigma_p^2}{\sigma_u^2} \tag{8}$$

where P_p represents the power gained by the sensing node from primary signal. From Eqs. (6) and (7) it is known that the values of $N, \text{SNR}_p, \sigma_u^2, P_d$ and P_{fa} mainly

depends on λ . The selection of λ properly to give better performance is a critical task.

CDR principle:

In CDR, P_d value is given, i.e., $\overline{P_d}$. The values of λ and P_{fa} are computed as

$$\lambda_d = \sigma_u^2 \left((Q^{-1}(\overline{P_d}) \cdot \sqrt{\frac{2\text{SNR}_p + 1}{N}}) + \text{SNR}_p + 1 \right) \quad (9)$$

$$P_{fa,CDR} = Q \left(\left(\frac{\lambda_d}{\sigma_u^2} - 1 \right) \sqrt{N} \right) \quad (10)$$

The transmitted power received at sensing node of primary user is given by [1]:

$$P_{rps} = P_p \cdot K_{ps} \cdot \left(\frac{d_0}{d_{ps}} \right)^r \quad (11)$$

Here P_p is primary user transmitted power, d_{ps} is distance of sensing node from primary user, d_0 is chosen distance considered as reference. The K_{ps} is a constant. It depends on the value of path loss of free space and antenna characteristics; r is value of path loss exponent. Here K_{ps} and d_0^r are constants. By substituting SNR_p in Eq. (9), we have the following equation

$$\lambda_d = \sigma_u^2 \left(\left(Q^{-1}(\overline{P_d}) \cdot \sqrt{\frac{1}{N}} + \frac{1}{\sigma_u^2} \cdot c \cdot P_p \cdot \frac{1}{d_{ps}^r} + 1 \right) \sqrt{N} \right) \quad (12)$$

Here c is formulated as $K_{ps} \cdot d_0^r$. While substituting in Eq. (9), $2\text{SNR}_p + 1 \cong 1$ is assumed to avoid complexity in calculation.

Critical distance:

A network model is considered with one primary transmitter which acts as base station with the range of service is R [1]. When primary receiver is located on boundary of cell which is closer to secondary transmitter, the receiver SINR is [1]:

$$\text{SINR} = \frac{P_{rpp}}{P_{rsp} + \text{noise}} \quad (13)$$

Here P_{rpp} is received power of primary user at its receiver end and P_{rsp} is the secondary signal power obtained at its primary receiver end while missed detection. By substituting P_{rpp} and P_{rsp} , we get the following equation [1]

$$\text{SINR} = \frac{\beta \cdot P_p \cdot \frac{1}{R^r}}{\alpha \cdot P_s \cdot \frac{1}{d^r} + \text{noise}} \quad (14)$$

where d is the distance between primary receiver and secondary transmitter at worst condition, P_s denotes the secondary user transmission power, r indicates path loss exponent and α is a constant. Here the unknown term β is constant.

Under missed detection condition, required desired message can be translated at the primary user if $SINR > SINR_{th}$ only. The minimum SINR, which is required for the primary user to successfully decode the desired information, is represented by $SINR_{th}$. In Eq. (14), all the values of α, β, P_p, P_s are known values. It is observed that SINR depends mainly on the distance of primary receiver from secondary transmitter which is shown as d . Here $SINR_{th}$ should be considered as any marginal value [1]. Then the critical distance can be written as [1]

$$d_c = \exp \left(\frac{1}{r} \cdot \ln \frac{\alpha \cdot P_s \cdot SINR_{th}}{\beta \cdot P_p \cdot \frac{1}{R^r} - \text{noise} \cdot SINR_{th}} \right) \tag{15}$$

3 The Overview of Double Threshold

In double threshold, two thresholds are present. If observed energy lies above upper value, then primary user presence is considered. If it lies below the lower value, then primary user absence is considered. The region which presents between both levels is known as uncertainty region. Here CR unable finalizes any decision. In these situations, sensing operation is repeated. Double threshold energy detection method improves the reliability, but the number of sensing repetitions is more. Thresholds in double threshold λ_1 and λ_2 are calculated as:

$$\lambda_1 = (1 - \rho) \tag{16}$$

$$\lambda_2 = (1 + \rho) \tag{17}$$

where ρ , the uncertainty parameter and this is used in calculation of thresholds.

Throughput:

In every time frame, it is assumed that τ is the sensing time then data transmission time will be $T - \tau$ [3]. Let us assume that probabilities of hypotheses are $P(H_0)$ and $P(H_1)$, respectively. Let SNR_s indicates the SNR of the secondary point to point link and SNR_p the primary user’s SNR at the secondary receiver. Achievable throughput of the secondary network is obtained from [3]:

$$\varphi = \frac{T - \tau}{T} [C_0(1 - P_f)P(H_0) + C_1(1 - P_d)P(H_1)] \tag{18}$$

where C_0 and C_1 are the capacities of the secondary user in the absence and presence of primary user, respectively. Those are formulated as:

$$C_0 = \log_2(1 + \text{SNR}_s) \quad (19)$$

$$C_1 = \log_2\left(1 + \frac{\text{SNR}_s}{1 + \text{SNR}_p}\right) \quad (20)$$

4 The Proposed Model

In this paper, we are using both single threshold and double threshold. Here distance is unknown parameter. And one critical distance is calculated using Eq. (15). Based on that, consider some range of distance and observe the through put. When the distance is less than the critical distance, the interference is more. So in that region, double threshold is used to reduce that problem. Here CDR principle is only used to sufficiently protect the primary users for the total distance range. While implementing double threshold using CDR principle, following parameters are:

$$P_{dd} = Q\left(\left(\frac{\lambda_{dd}}{\sigma_u^2} - \text{SNR}_p - 1\right)\sqrt{\frac{N}{2\text{SNR}_p + 1}}\right) \quad (21)$$

$$\lambda_{dd} = (1 - \rho)\lambda_d \quad (22)$$

$$P_{fa} = Q\left(\left(\frac{\lambda_{dd}}{\sigma_u^2} - 1\right)\sqrt{N}\right) \quad (23)$$

5 Simulations and Analysis

As discussed in Sect. 1, $s[n]$ and $u[n]$ are taken as CSCG. The parameters are considered as frame duration $T = 100$ ms, target frequency band = 60 MHz probability of hypothesis H_0 , i.e., $P(H_0) = 0.8$, transmitted power of primary user = 1 W, threshold signal-to-noise ratio = 8 dB, and service range = 450 m. In this paper, we used CDR principle. Probability of detection is taken as 0.99. Critical distance is calculated using Eq. (21). Up to critical distance, we used double threshold concept and after that critical distance single threshold is used. Figure 1 shows how output is varying according to the distance for double threshold and single threshold. Figure 2 shows the throughput variation along the distance for various path loss exponents like 4.3, 4.4 and 4.5.

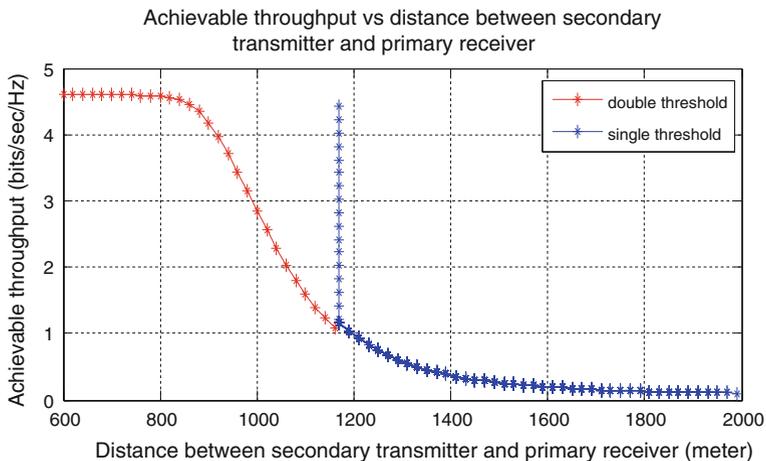


Fig. 1 Achievable throughput of CR for proposed model

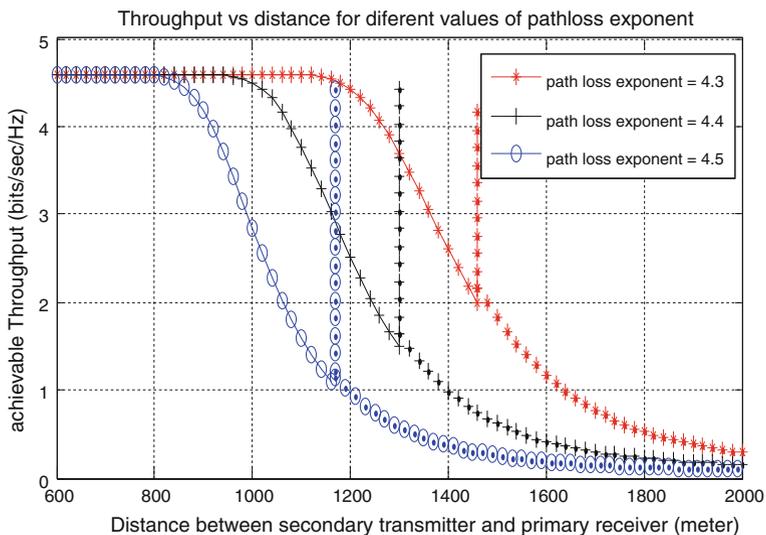


Fig. 2 Throughput of CR for different values of path loss exponent

6 Conclusion

In this paper, we first analyzed about the energy detection and CDR principle is selected as the aim to provide sufficient protection for the primary user. Double threshold overview is observed. Based on distance, double threshold is

implemented where the interference is more. At distance which is less than the critical distance, the interference is more. So in this paper, double threshold is implemented in that region only. In future in cognitive radio applications if this double threshold is introduced, then interference reduced and good results also expected.

References

1. Verma, G., Sahu, O.P.: Intelligent selection of threshold in cognitive radio system. *Telecommun. Syst.* Springer, 1–10 (2016)
2. Verma, G., Sahu, O.P.: Interference aware optimization of throughput in cognitive radio system, pp. 1–6. *Defence J.*, India (2015)
3. Verma, G., Sahu, O.P.: Opportunistic selection of threshold in cognitive radio networks. *Wirel. Pers. Commun.* Springer, 1–16 (2016)
4. Liang, Y.C., Zeng, Y., Peh, C.Y., Hoang, A.T.: Sensing throughput tradeoff for cognitive radio networks. *IEEE Trans. Wirele. Commun.* **7**(4), 1326–1337 (2008)
5. Pandit, S., Singh, G.: Throughput maximization with reduced data loss rate in the cognitive radio network. *Telecommun. Syst.* 209–215 (2013)
6. Choi, H.H., Jang, K., Cheong, Y.: Adaptive sensing threshold control based on transmission power in cognitive radio systems. In: *IEEE Conference Cognitive Radio Oriented Wireless Networks and Communication (CrownCom)*, pp. 1–6 (2008)
7. Hamdi, K., Zang, W., Lataief, K.B.: Power control in cognitive radio systems based on spectrum sensing side information. In: *International conference on communications (ICC)*, pp. 5161–5165 (2007)
8. Liu, K.H., Hseih, H.Y.: On using interference-aware spectrum sensing for dynamic spectrum access in cognitive radio networks. *IEEE Trans. Mob. Comput.* **12**(3), 461–474 (2013)
9. Stotas, S., Nallanathan, A.: Optimal sensing time and power allocation in multiband cognitive radio networks. *IEEE Trans. Commun.* **59**(1), 226–235 (2011)
10. Digham, F., Alouini, M.S., Marvin, K.: On the energy detection of unknown signal over fading channels. *IEEE Trans. Commun.* **55**(1), 21–24 (2007)

X-Band Patch Array Antenna

Nitya Nirmal, Ayushi Agarwal, Arun Kumar Singh,
Rabindranath Bera and Bansibadan Maji

Abstract The following contains a detailed study as well as design of the hybrid patch array antennas and their application in transmission of signals unaltered. The different dimensions of design including substrate material, width, distance between two patch and other elements have been done through various theoretical analysis and techniques. Practical approach for triangular patches according to the distribution of current is summarized below. The mathematical calculation related to the array is emphasized and critical factors are also considered while determining the development of this antenna.

Keywords Rectangular patch · Triangular patch · Return loss · Gain · X-band

1 Introduction

The purpose of this paper is to serve analytic and practical approach to the signal reception in the X-band within the bandwidth of 8–11 GHz. The comprehensive study of the mechanism involved in serving this process and to engage the prop-

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

N. Nirmal
e-mail: nityanirmal4@gmail.com

A. Agarwal
e-mail: ayushiag529@gmail.com

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

B. Maji
Department of Electronics and Communication Engineering,
National Institute of Technology, Durgapur, Durgapur, West Bengal, India
e-mail: bmajiecenit@yahoo.com

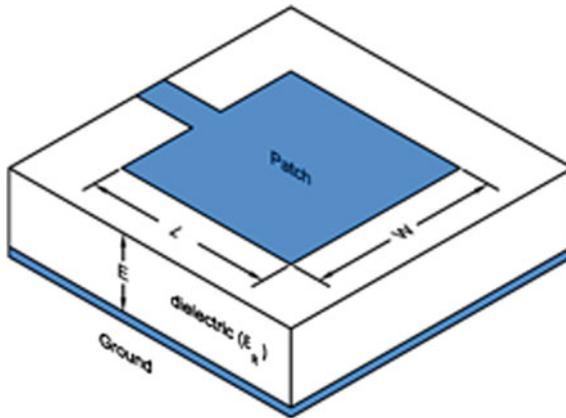


Fig. 1 Microstrip patch antenna schematic

erties of the antenna is thereby explained. Wireless communication is the developing and innovative upgradation in technology and has a wide spectrum of applications [1]. The acknowledgment of the patch array antenna is in demand for the applications in specific resonating frequencies and to have a higher gain and bandwidth response [2, 3]. The software design completion is done through HFSS, i.e., high frequency structural simulator. The live examples of expertise comes in the reception through DTH where the beam width is focused toward the satellite for greater signal strength and quality [4]. The ideology surrounds the effective delivery of the signal and lie under the desired output.

The mathematical modeling of patch antenna is done as per [5, 6]. All the approach followed for the design was based on the standard derivation of dimensions for rectangular patch antenna (Fig. 1).

2 Antenna Design

Figure 2 shows an array antenna of 1×3 which is designed in X-band. The current distribution is shown in the figure. The return loss was decreased to a significant value and the gain was also around 7 dB (Figs. 3 and 4).

In the 1×3 design, we find that there is an increase in gain, i.e., 7.5 dB and with a return loss of -45 dB. Though the current distribution was uneven but higher impedance matching and lower return loss, the efficiency was increased as compared to simple rectangular designs (Fig. 5).

The design shows a hybrid type of patch array where rectangular patch is designed together with triangular patch. Since rectangular patch has better current distribution and triangular patch has better efficiency, the design will offer a system with overall better performance (Figs. 6 and 7).

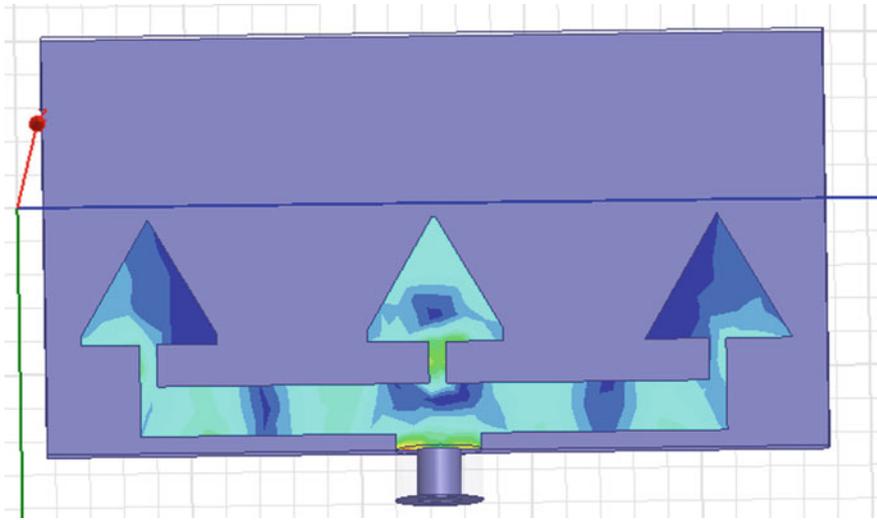


Fig. 2 1 × 3 microstrip patch triangular array antenna

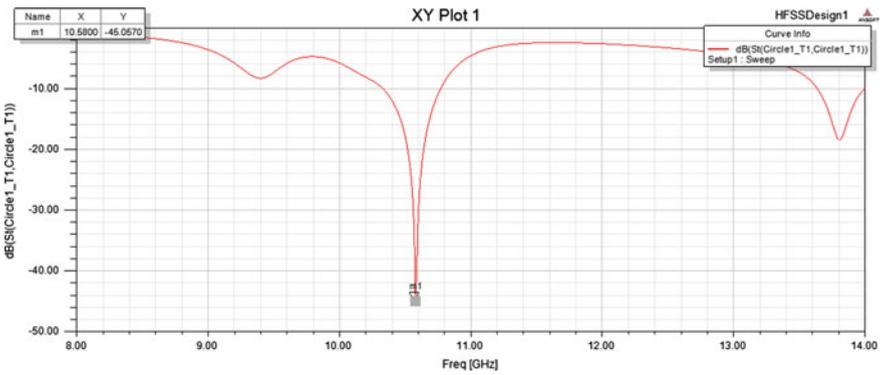


Fig. 3 Return loss versus frequency of X-band array antenna

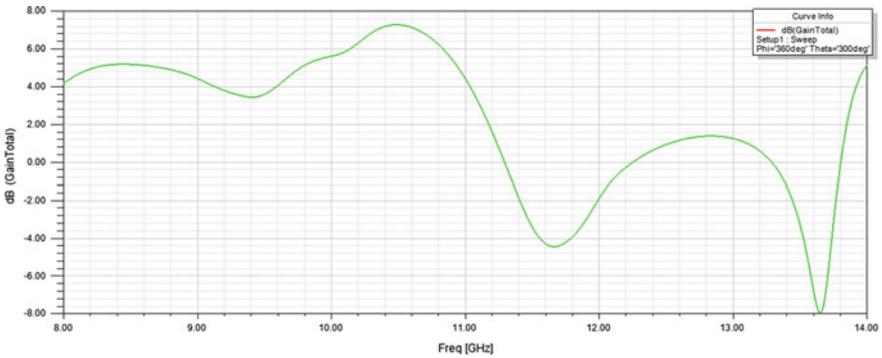


Fig. 4 Gain versus frequency of X-band array antenna



Fig. 5 1 × 2 hybrid patch array antenna

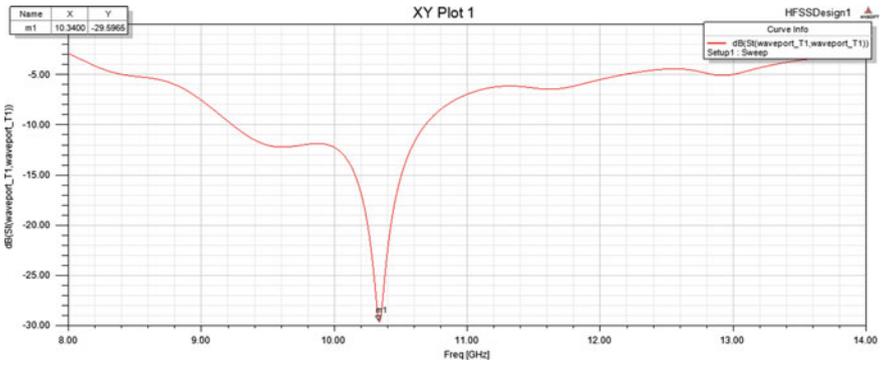


Fig. 6 Return loss versus frequency of hybrid array antenna

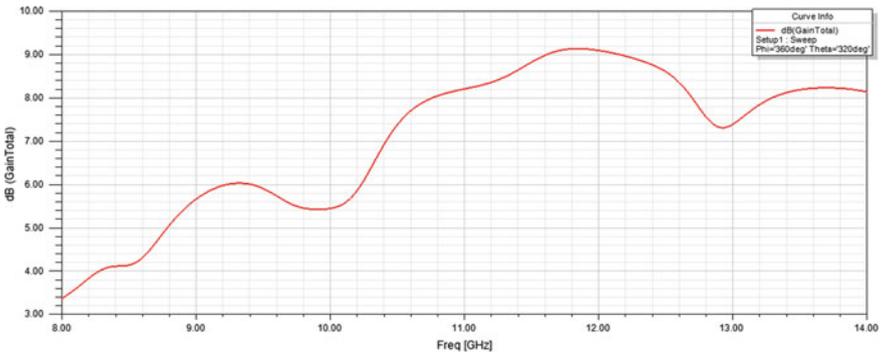


Fig. 7 Gain versus frequency of X-band array antenna

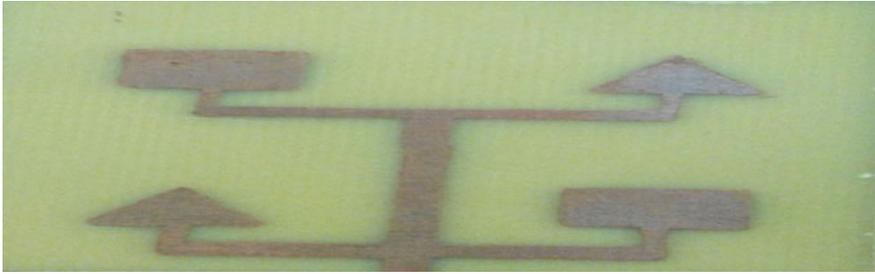


Fig. 8 2 × 2 hybrid patch array antenna

In the above design of hybrid antenna, the return loss is -29.5965 dB at 10.4 GHz and the gain is 10.3400 dB. As compared to triangular patch array of 1×3 , the gain is increased by 3 dB. The advantage of hybrid antenna is thus achieved in this design.

Since the previous design of 1×2 hybrid antenna yielded a better result, the author has presented a 2×2 array of hybrid type. Figure 8 shows the fabricated design of 2×2 hybrid array antenna.

Figures 9 and 10 shows return loss and gain of the X-band hybrid antenna. Interestingly, the design of 2×2 hybrid antenna is multiband antenna which is resonating at three different frequencies in the X-band. The return loss is around -20 dB which is in acceptable range. The gain was found to be around 9 dB.

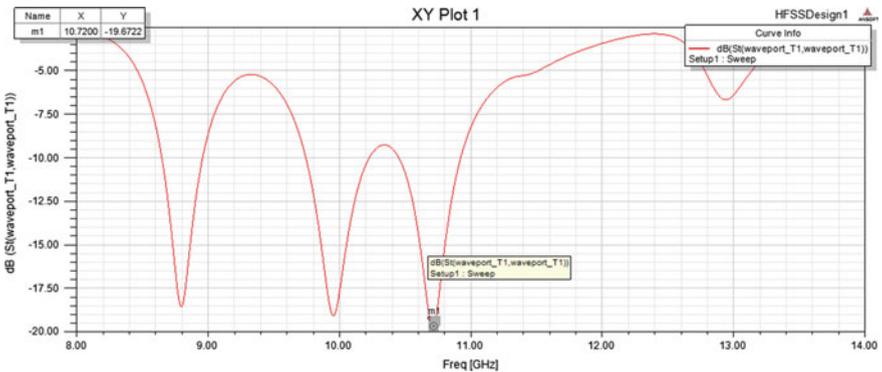


Fig. 9 Return loss versus frequency of 2×2 hybrid array antenna

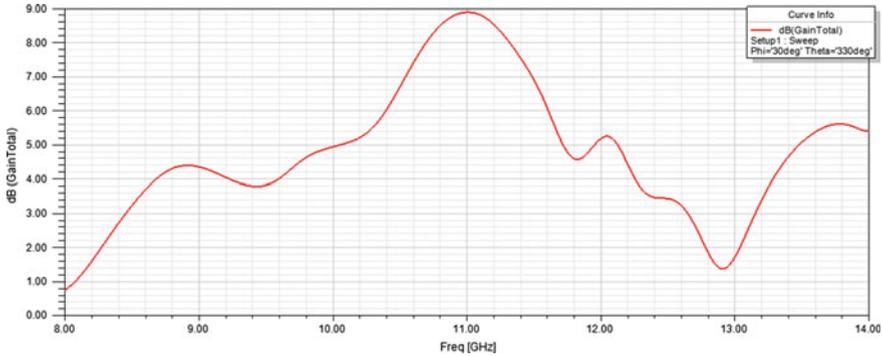


Fig. 10 Gain versus frequency of 2×2 hybrid array antenna

3 Summary/Conclusion

The X-band comprises of frequencies ranging from 8 to 12 GHz which itself resembles its wide applications, from microwave transmission to DTH reception. It is widely used for transmission purpose. The gain and return loss could be modified as per its design parameters. The hardware design of this was basically done in the conventional way which has larger exposure to human error. For improved outcomes, the more scientific approach of laser cutting and exposure to UV radiation can be performed. The future applications could be DTH reception on wheels where no compromise with the signal quality and signal strength is done.

References

1. Yadav, A., Chauhan, B., Jain, A.: Microstrip symmetrical E-shape patch antenna for the wireless communication systems **2**, 241–244 (2012)
2. Liu, S.F., Shi, X.W.: Study on the impedance-matching technique for high-temperature super conducting microstrip antennas **77**, 281–284 (2007)
3. Arya, A., Sharma, S., Yadav, P., Dindha, R.: Performance comparison between rectangular & circular patch antenna array with EBG structure **7** (2012)
4. Nashaat, D., Elsadek, A.H., Abdallah, A.E., Iskander, F.M., El Hennawy, M.H.: Ultrawide bandwidth 2×2 microstrip patch array antenna using electromagnetic band-gap structure **59**, 1528–1533 (2011)
5. Dhar, S., Patra, K., Ghatak, R., Gupta, B., Poddar, R.D.: A dielectric resonator-loaded minkowski fractal-shaped slot loop heptaband antenna **63**, 1527–1529 (2015)
6. Roy, A.A., Mom, M.J., Igwe, G.A.: Enhancing the bandwidth of a microstrip patch antenna using slots shaped patch **2**, 23–30 (2013)
7. Kushwaha, S.R., Srivastava, D.K., Saini, J.P.: A multi-slotted wide microstrip patch antenna for dual frequency **3(2)**, 3523–3525 (2012)
8. Majumder, A.: Rectangular microstrip patch antenna using coaxial probe feeding technique to operate in S-band **4**, 1206–1210 (2013)

9. Dalli, A., Zenkouar, L., Bri, S.: Comparison of circular sector and rectangular patch antenna arrays in C-band **4**, 457–467 (2012)
10. Huque, M.T.I., Hosain, M.K., Islam, M.S., Chowdhury, M.A.: Design and performance analysis of microstrip array antennas with optimum parameters for X-band applications **2**, 82–87 (2011)
11. Rishishwar, D., Shrivastava, L.: Rectangular microstrip patch antenna with FSS and slotted patch to enhance bandwidth at 2.4 GHz for WLAN applications **2**, 59–62 (2014)
12. Kaur, G., Singla, G., Kaur, S.: Design of wideband micro strip patch antenna using defected ground structure for wireless applications **3**, 762–765 (2013)
13. Dharsandiya, N., Parmar, D.I.: Optimization of antenna design for gain enhancement using array **4**, 1038–1043 (2014)

A Compact Pentagonal Textile Microstrip Antenna for Wide Band Application

Rachana Kushwaha, Vinod Kumar Singh, Nikhil Kumar Singh,
Anurag Saxena and Deepak Sharma

Abstract In this article, a compact-slotted pentagonal-shaped patch antenna for multiband applications with improved bandwidth is presented. The presented antenna is designed on a textile substrate with dielectric constant of 4.4. The compact size of antenna is $50 \times 50 \times 1$ mm. This presented antenna has been simulated using CST tool. The presented antenna resonates at frequency 4 GHz with the reflection coefficient of -31.84 . It covers a bandwidth of 6.45 GHz extending from 2.85 to 9.30 GHz. It has maximum directivity of 2.932 dBi.

Keywords Partial ground · Wide band · Wireless application · Patch antenna
CST tool

1 Introduction

Currently, the fast development of modern communication systems is required for transportable devices for some important features which includes easy designing, light weight, small in size, compatible with microwave, millimeter wave integrated circuits, less production cost, and easy fabrication of microstrip antennas. The microstrip antenna has numerous useful properties which includes minute size, low-cost of the fabrication, light weight, ease of installation, but the main limitations of printed antennas remains their narrow bandwidth features which limits the range of frequency over which the antenna can work effectively. Microstrip antenna

R. Kushwaha · D. Sharma
Electrical Engineering Department, Arya Institute of Engineering
& Technology, Kookas, Jaipur, India
e-mail: rachanakushwaha.12@gmail.com

V.K. Singh (✉) · N.K. Singh · A. Saxena
Department of Electrical Engineering, S R Group of Institutions, Jhansi, UP, India
e-mail: singhvinod34@gmail.com

N.K. Singh
e-mail: nitinm18@gmail.com

comprises three most important parts which is substrate, patch, and ground. One side is radiating patch, ground plane on the other side, and a dielectric substrate sandwiched between it. The conducting patch is placed on the dielectric substrate which is used as a radiating element. On the other side of dielectric substrate, there is a conducting layer used as a ground part [1–6]. Microstrip antenna plays very significant role in wireless communication system. Microstrip antennas are used in high performance aircrafts, radar, missiles, and other spacecraft. It has many advantages such as its light weight, simple structure, ease of integration, and less cost. Microstrip antenna requires very less space for installation as these are simple and small in size. The only space required is the space for the feed line which is placed behind the ground plane. Microstrip antennas are low profile, simple, conformable to planar and non-planar surfaces, inexpensive to manufacture using modern printed circuit technology [7–14].

There are varieties of methods for enhancing the bandwidth of textile microstrip antenna such as broadening the substrate thickness, utilizing low dielectric material, using numerous feeding techniques, and impedance matching. But the bandwidth and the thickness of the antenna are the both contradictory properties, that is, improvement of bandwidth will increase the size and thickness of presented antenna [15–22].

2 Antenna Design Configuration

The presented antenna works for three resonant frequencies which is shown Fig. 1. If there is more than one resonant part available with each operating at its own resonant frequency, then the overlapping of multiple resonances lead to broadband

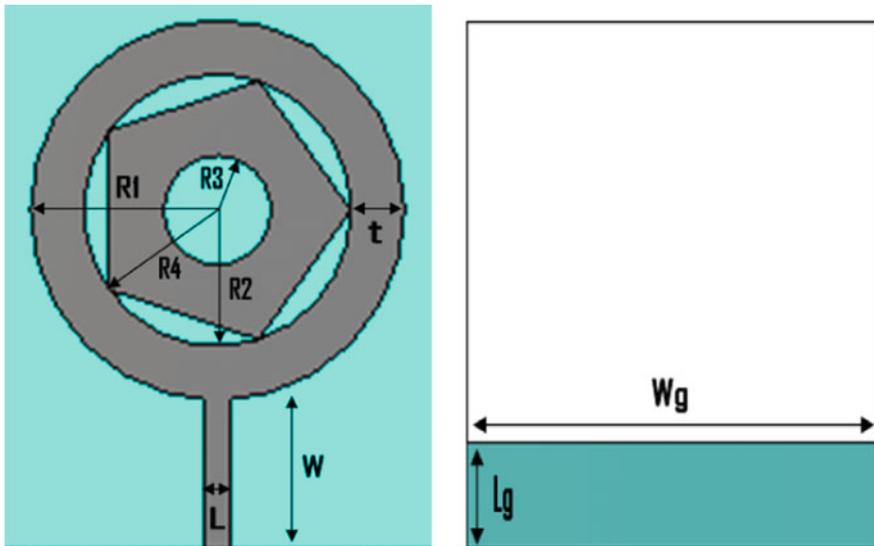


Fig. 1 Configuration of the proposed pentagonal textile microstrip antenna

Table 1 Dimensions of proposed textile antenna

Design parameters	Value
Permittivity of material (ϵ_r)	1.70
The loss tangent ($\tan\delta$)	0.025
Thickness of substrate (h)	1.0 mm
Outer circle patch radius ($R1$)	14 mm
Inner circle patch radius ($R2$)	10 mm
Inner circle slot radius ($R3$)	4 mm
Pentagonal patch radius ($R4$)	5 mm
Microstrip feed line ($L \times W$)	2×15
Substrate dimension ($L_s \times W_s$)	50×50
Partial ground plane ($L_g \times W_g$)	50×10

applications. The basis of the proposed antenna was an annular ring with outer patch radius 14 mm and inner patch radius of 10 mm. Then pentagonal shape was obtained by cutting a pentagon shape slot from the circular patch, and finally the proposed design was obtained by cutting a circle in pentagon (Table 1).

3 Optimization of Proposed Microstrip Antenna

The CST software has been used to design and optimize the results created by the antenna. Initially, circular patch was used to achieve the desired bandwidth, then the circular patch was modified to pentagonal-shaped patch and finally the proposed

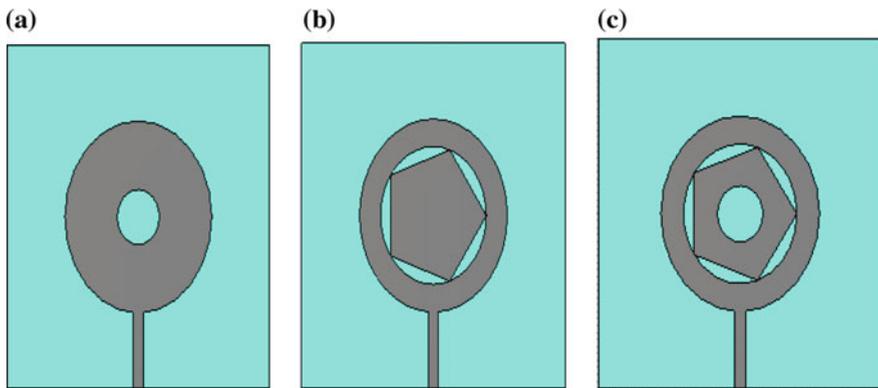
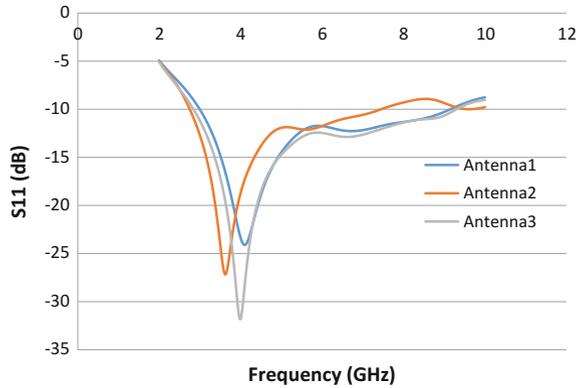


Fig. 2 Geometry of the proposed antennas **a** antenna1 **b** antenna2 **c** proposed antenna3

Fig. 3 Simulated return loss versus frequency for the proposed textile microstrip antenna1, antenna2 and antenna3



shaped patch was achieved to optimize the results. Figure 2 shows the geometry of antenna1, antenna2, and antenna3. The return loss for antenna1, antenna2, and antenna3 has been denoted by blue, orange, and gray colored curves, respectively, in Fig. 3.

4 Radiation Pattern

Figure 4 depicts simulated 2-D radiation pattern of proposed microstrip antenna at 4 GHz which describes the main lobe direction = 168° and angular width (3 dB) = 73° at $\varphi = 90^\circ$, and it gives a main lobe magnitude = 2.6 dBi and main lobe direction = 180° at $\varphi = 90^\circ$. Figure 5 shows simulated 3-D radiation pattern of proposed microstrip antenna at 4 GHz, The proposed antenna gives good radiation efficiency of about -1.807 dB and directivity of 2.932 dBi as shown in Fig. 5.

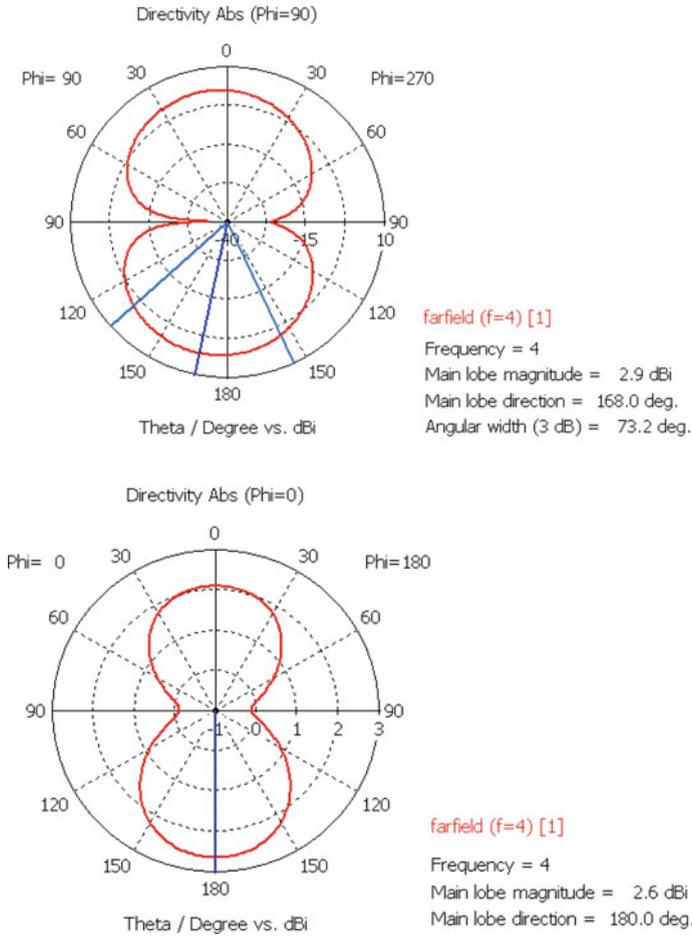


Fig. 4 Simulated 2D radiation pattern of proposed textile microstrip antenna at 4 GHz

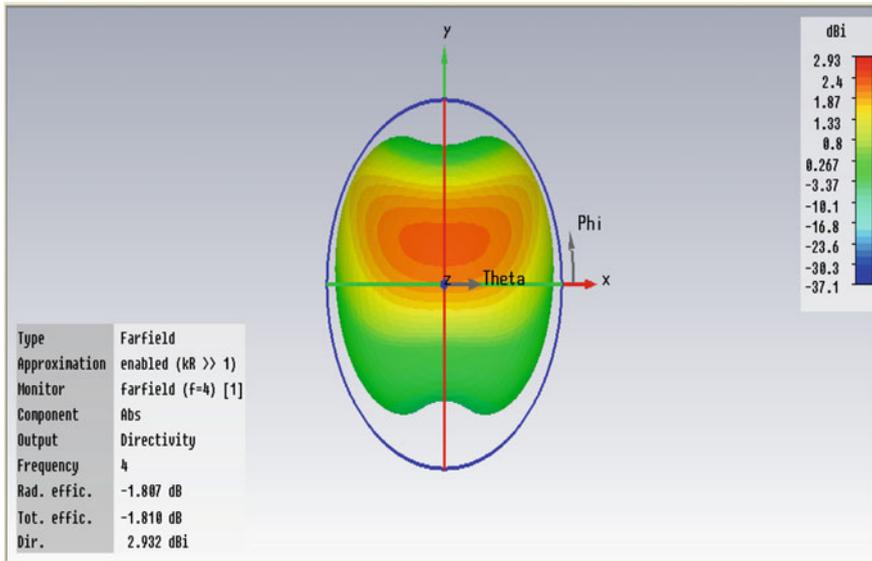


Fig. 5 Simulated 3D radiation pattern of proposed textile microstrip antenna at 4 GHz

5 Conclusion

A pentagonal-shaped microstrip antenna is estimated for broadband application. The overall dimension ($50 \times 50 \times 1$ mm) of presented antenna is very compact, thus suitable for installation on heavy satellites. The results are obtained using CST software. The antenna has maximum gain of about 2.932 dB. The antenna has a stable radiation pattern over the entire 6.45 GHz bandwidth that extends from 2.85 to 9.30 GHz, thus the presented antenna is appropriate for broadband application.

References

1. Ram Krishna, R.V.S., Kumar, R.: Design of temple shape slot antenna for ultra wideband applications. *Prog. Electromagn. Res. B* **47**, 405–421 (2013)
2. Balanis, C.A.: *Antenna Theory: Analysis and Design*. Wiley, New York (2004)
3. Wong, K.-L.: *Compact and Broadband Microstrip Antennas*. Wiley, New York. ISBN: 0-471-41717-3 (Hardback) (2002)
4. Singh, V.K., Ali, Z., Ayub, S., Singh, A.K.: A wide band compact microstrip antenna for GPS/DCS/PCS/WLAN applications. In: *Intelligent Computing, Networking, and Informatics*, vol. 243, pp. 1107–1113. Springer, Berlin (2014). ISBN: 978-81-322-1664-3
5. Srivastava, R., Singh, V.K., Ayub, S.: Comparative analysis and bandwidth enhancement with direct coupled C slotted microstrip antenna for dual wide band applications. In: *Frontiers of Intelligent Computing: Theory and Applications*, vol. 328, pp. 449–455. Springer, Berlin (2015). ISBN: 978-3-319-12011-9

6. Singh, V., Ali, Z., Ayub, S., Singh, A.: Bandwidth optimization of compact microstrip antenna for PCS/DCS/bluetooth application. *Central Eur. J. Eng.* **4**(3), 281–286 (2014). Springer
7. Singh, N., Singh, A., Singh, V.: Design & performance of wearable ultra wide band textile antenna for medical applications. *Microw. Opt. Technol. Lett.* **57**(7), 1553–1557 (2015)
8. Srivastava, S., Singh, V., Ali, Z., Singh, A.: Duo triangle shaped microstrip patch antenna analysis for WiMAX lower band application. In: *International Conference on Computational Intelligence: Modelling Techniques and Applications (CIMTA-2013)*, vol. 10, pp. 554–563. Procedia Technology Elsevier (2013)
9. Singh, N.K., Singh, V.K., Naresh, B.: Textile antenna for microwave wireless power transmission. In: *International Conference on Computational Modelling and Security (CMS 2016)*, *Procedia Computer Science*, vol. 85, pp. 856–861 (2016)
10. Gupta, N., Singh, V.K., Ali, Z., Ahirwar, J.: Stacked textile antenna for multi band application using foam substrate. In: *International Conference on Computational Modelling and Security (CMS 2016)*, *Procedia Computer Science*, vol. 85, pp. 871–877 (2016)
11. Rawat, A.K., Singh, V.K., Ayub, S.: Compact wide band microstrip antenna for GPS/WLAN/WiMax applications. *Int. J. Emerg. Trends Eng. Dev.* **7**(2), 140–145 (2012)
12. Baudh, R.K., Kumar, R., Singh, V.K.: Arrow shape microstrip patch antenna for WiMax application. *J. Environ. Sci. Comput. Sci. Eng. Technol.* **3**(1), 269–274 (2013)
13. Loni, J., Ayub, S., Singh, V.K.: Performance analysis of Microstrip patch antenna by varying slot size for UMTS application. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 01–05 (2014)
14. Srivastava, R., Ayub, S., Singh, V.K.: Dual band rectangular and circular slot loaded microstrip antenna for WLAN/GPS/WiMax applications. In: *IEEE Conference on Communication Systems and Network Technologies (CSNT-2014)*, pp. 45–48 (2014)
15. Dhupkariya, S., Singh, V.K.: Textile antenna for C-band satellite communication application. *J. Telecommun. Switching Syst. Netw.* **2**(2), 20–25 (2015)
16. Loni, J., Singh, V.K.: Development of bandwidth enhanced microstrip patch antenna for UMTS application. *J. Microw. Eng. Technol.* **2**(1), 01–07 (2015)
17. Singh, V.K., Naresh, B.: Multi resonant microstrip antenna with partial ground for radar application. *J. Telecommun. Switching Syst. Netw.* **2**(1), 01–05 (2015)
18. Singh, M., Singh, V.K., Naresh, B.: Rectangular slot loaded circular patch antenna for WLAN application. *J. Telecommun. Switching Syst. Netw.* **2**(1), 07–10 (2015)
19. Ali, Z., Singh, V.K., Kumar, A., Shahanaz, A.: E shaped microstrip antenna on rogers substrate for WLAN applications, pp. 342–345. In: *Proceedings of IEEE* (2011)
20. Din, N.M., Chakrabarty, C.K., Bin Ismail, A., Devi, K.K.A., Chen, W.Y.: Design of RF energy harvesting system for energizing low power devices. *Prog. Electromagnet. Res.* **132**, 49–69 (2012)
21. Mai, A.R., Osman, M.K.A., Rahim, M.Azfar, Samsuri, N.A., Zubir, F., Kamardin, K.: Design, implementation and performance of ultra-wideband textile antenna. *Prog. Electromagnet. Res. B* **27**, 307–325 (2011)
22. Singh, V.K., Singh, N.K.: Compact circular slotted microstrip antenna for wireless communication systems. *J. Microw. Eng. Technol.* **1**(1), 07–14 (2015)

Similarity Measurement of Gene Using Arc Tan Function in Gene Ontology

K. Vengatesan, S.B. Mahajan, P. Sanjeevikumar, R.M. Kulkarni and Sana Moin

Abstract The Gene Ontology (GO) is a technique that measures second-hand the semantic relationship of genes using an exponential and arc tan-based technique. The proposed work develops an enhanced shortest path-based measure using GO semantic relationship. The existing work presented an undeviating corridor based on an amalgam appraise of ontological match. Using GO graph and in sequence pleased are factors that improve the effectiveness of transfer functions of semantic connection of gene ontology. The semantic technique evaluates weighted paths for GO similarity measure. Other similarity measures use the shortest path in their calculation but require the specificity of a concept in hybrid measure performance of correlated genes.

Keywords Gene ontology · Semantic similarity · Shortest path
Arc tan function · Relational similarity

K. Vengatesan (✉)

Department of Computer Engineering, Sri Satya Sai University of Technology
and Medical Science, Sehore, India
e-mail: vengicse2005@gmail.com

S.B. Mahajan · R.M. Kulkarni

Department of Electrical and Electronics Engineering,
Marathwada Institute of Technology, Aurangabad, India
e-mail: sagar25.mahajan@gmail.com

R.M. Kulkarni

e-mail: kulkarni.rishi123@gmail.com

P. Sanjeevikumar

Department of Electrical and Electronics Engineering, University of Johannesburg,
Auckland Park, Johannesburg, South Africa
e-mail: sanjeevi_12@yahoo.co.in

S. Moin

Department of Computer Science and Engineering, Manipal Institute of Technology,
Manipal, India
e-mail: sana.moin.star@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

A. Kalam et al. (eds.), *Advances in Electronics, Communication
and Computing*, Lecture Notes in Electrical Engineering 443,
https://doi.org/10.1007/978-981-10-4765-7_83

801

1 Introduction

Computational ontology denotes officially designing the formation of a system, i.e., the appropriate entities and relations that appear since its observation, and which are helpful to reasons. The gene ontology (GO) is cooperation between sculpt creature information to explain gene harvest commencing all organisms using a reliable and quantifiable language [1]. GO produces group of plainly definite, ordered vocabularies that portray natural processes, molecular operations, and cellular mechanism of gene harvest in together a central processing unit and human-understandable way. The method explains solution aspects of GO, which, when unnoticed, know how to basis unacceptable fallout, and deal with how these pitfalls can be uninvolved [2]. The GO are organized vocabulary and category that encircle diverse area of molecular and cellular environmental science and are munificently effortlessly reached for group of people exploit in the marginal note of genes, gene foodstuffs and sequences. Most of the information on human being appearance and genome explanation sets touch the GO and make a contribution their marginal note sets for the GO supply [3]. The GO record incorporates the vocabularies and contributed observations and gives whole right to use this in turn unusual structures. The performed outlines of sequence deviation and descent of mitochondrial DNA (mtDNA) alleles in the coconut crab, *Birgus Zatro*, treatment self-control enzyme carry out. Populations from the Pacific Ocean had mtDNAs with boost multiplicity, and a Poisson allowance of progression had a variety of, i.e., the inherited autograph of swiftly mounting residents [4]. In which the build up a microscopic bodily reproduction of premature development, where phenotype—creature life expectation—is straight connected to genotype—the faithfulness of its proteins in their resident conformations which be capable of unwavering exactly in the sculpt [5]. Simulating sculpt on a mainframe and scrutinizing the “Big Bang” circumstances repeatedly, exponential inhabitants development follows almost straight away as industrious sequence–structure mixture (precursors of sure proteins) are dogged [6]. The transformation among *Streptococcus pneumonia* beneficiary strains and benefactor DNA into related sprains and kind follows the explained exponential connection. It explains that the Hex disparity restore structure fakes a momentous barrier to recombination more the whole choice of progression divergence (0.6–27%) studied [7]. The enduring non-Hex-mediated blockade to recombination can be afforded by assortment of mechanism. The problem discusses the potential further methods of sexual separation, in show of previous conclusion starting *Streptococcus*, *Escherichia*, and *Bacillus* [8]. An exponentially narrowed form is initiated addicted to multimode interference (MMI) strategy. Compared through a parabolic ally conical arrangement, which has been effectively handled in MMI, devices, this organization can more decrease the length of these devices. The performances of the 1×2 MMI coupler among exponentially tapered form, like the visual broadcast, the wavelength reply and the manufacture acceptance, are explored by the 2-D limited variation beam dissemination system [9]. Gene regulatory networks have a significant task in every process of life, adding cell

separation, metabolism, the cell cycle, and signal transduction. Studying the dynamics of these networks can discard light on the devices of infections that occur when these cellular processes are deregulated. Perfect predictions of the performance of regulatory networks will also speed up biotechnological projects, i.e., calculations are faster and cheaper than the practical experiments [10]. The vibrant behaviors are anticipated into depiction gene network incorporating regular biochemical motifs and mingle modern technique for model heritable networks [11].

2 Proposed Model of Gene Ontology

Gene networks signify abstract models for difficult interaction approaches between genes and proteins. Thus, while arithmetic structures can, e.g., predict appearance levels of a miniature pair of genes in few well-studied pathways, the normal genetic interactions remains in high level are unknown. As a result, various assumptions concerning genetic interactions handled in unsupervised network reconstruction methods may not generalize to new interaction types or organisms. Figure 1 represents the architecture diagram of exponential and arc tan-based shortest path measure based on the weighted paths, biological gene ontology network, and hybrid measure and diffusion kernels. From which biological gene ontology network follows three steps such as protein functions, genes, and false positive or false negatives.

As more interactions between genes are investigated, ally determined and/or authenticated, statistical learning methods can be handled to both conclude the nature of such interactions and also to handle this knowledge to calculate new and unobserved interactions. But supervised approach of network renovation frequently focuses on a particular type of genetic interaction. The exponential and arc tan-based shortest path measures for semantic similarity in gene ontology are divided into three phases.

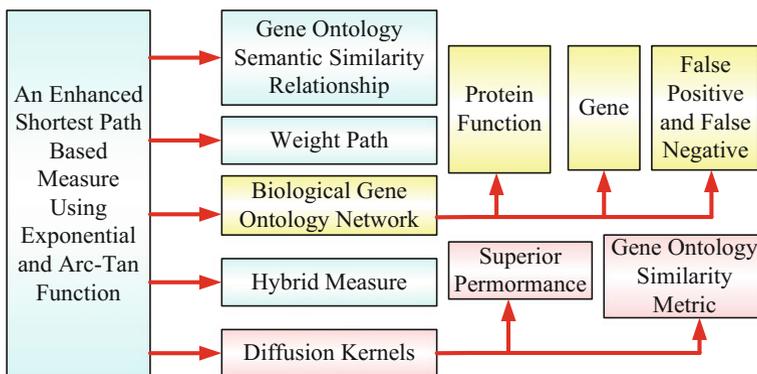


Fig. 1 Architecture diagram of exponential and arc tan-based shortest path measure

2.1 Gene Ontology—Relational Similarity

Gene ontology is based on hierarchical relations in which consider the matrix of gene pair shrewd similarities surmise true hierarchical arrangement narrate to manifold higher level terms. These works find maximal relational weight induced through progressive threshold of a correspondence template this process evaluates the ability to reconstruct GO biological process ontology from similarity matrix. This scheme constructs relational weight with a specified echelon in the ladder by optimizing robustness utility from numerous germ nodes. The robustness function includes a constraint tuned to find the several levels of chain of command.

2.2 Gene Exponential Relational

Gene exponential relation analyzes the relative changes in gene expression from real-time quantitative exponential gene ranges. Gene exponential presents the derivation, assumptions, and applications. It determines the absolute relation of various exponents. Term mapping of gene exponential relation is evaluated using a score function similarity of the group of genes assign to expressions (inherent similarity). Another one is resemblance of the ontology hierarchy surrounding every term. Every set of terms united and receives an arrangement attain ranging commencing 0–1. 1 represents uniqueness in equally essential and relational resemblance. Multimode exponential relation is done for supervised prediction of gene networks. This scheme coupled with a penalized estimation for improved gene relational prediction. Facilitates the analysis of gene networks with multiple edges provides a systematic values for incorporating: initially multiple sources of biological data and second one is diverse attributes regarding function and location of genes. Finally the structure of networks is observed.

2.3 Arc Tan-Based Shortest Path

Shortest path between two gene ontology similarity relational weights is made proportional to constant angle of arc tan. The covariates of the arc tan constant are gene ontology network attributes, added (i) degrees of nodes and the meagerly of the observed network; (ii) data on gene activity in the cell, adding steady state, or time course gene expression levels; and (iii) functional genes attributes, adding genes are transcription factors, and their pathway membership. Weighted estimation is arrived for arc tan shortest path for gene ontology pair similarity. Gene ontology relational network is built by thresholding the absolute values of pair wise correlations between genes at a higher similarity weight range.

The shortest path between nodes are measured using Dijkstra’s algorithm that follows the time complexity $O(\log(N * E))$. The similarity between genes is calculated using Eq. (1) and the dissimilarity between the genes are calculated using Eq. (2). Finally, the overall similarity value between t_1 and t_2 are written using Eq. (3).

$$\text{Sim}(t_1, t_2) = \frac{\arctan(\text{sim}(t_1, t_2))}{\frac{\pi}{2}} \tag{1}$$

$$\text{dis}(t_1, t_2) = 1 - \frac{\arctan(\text{dis}(t_1, t_2))}{\frac{\pi}{2}} \tag{2}$$

$$\text{SPBHM}(t_1, t_2) = \frac{\left(\frac{\arctan(\text{sim}(t_1, t_2))}{\frac{\pi}{2}} + 1 - \frac{\arctan(\text{dis}(t_1, t_2))}{\frac{\pi}{2}}\right)}{2} \tag{3}$$

3 Performance Metric

This work quantifies the performance of exponential and arc tan-based shortest path measure for semantic similarity in gene ontology. The analysis is done through the Java platform. This scheme improves the effectiveness of transfer functions of semantic similarity of gene ontology. The performance measures of the proposed work are analyzed with the following metrics such as exponential relational rate, arc tan shortest path range, and similarity ratio.

3.1 Exponential Relational Rate

This form develops and expands the framework of relational exponential for graph model for supervised prediction of genetic interactions. This scheme is coupled with a penalized estimation framework to improve the prediction performance. Figure 2 demonstrates exponential relational rate. X axis represents the gene size values, whereas Y axis denotes exponential relational rate using both the concept of semantic similarity in gene ontology. When gene size increased, the exponential relational rate increased accordingly. The shortest path measure for semantic similarity (SPMSS) achieves high performance of 5–10% when compared to the existing system (HMGO). In Table 1 exponential rates of HMGO and SPMSS is compared.

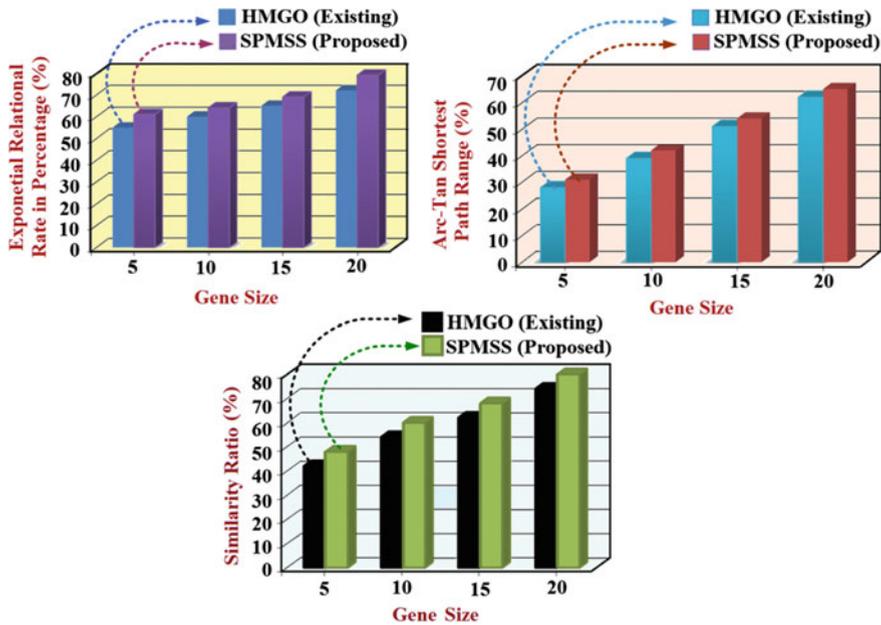


Fig. 2 Graph of gene size versus exponential relational rate (*top left*), gene size versus arc tan shortest path range (*top right*), gene size versus similarity ratio (*bottom*)

Table 1 Tabularize form of gene size versus exponential relational rate, gene size versus arc tan shortest path range, gene size versus similarity ratio for HMGO (existing) and SPSS (proposed)

Gene size	Exponential relational rate		Arc tan shortest path		Similarity ratio (%)	
	HMGO (existing)	SPSS (proposed)	HMGO (existing)	SPSS (proposed)	HMGO (existing)	SPSS (proposed)
5	55	61	28	31	42	48
10	60	64	39	42	54	60
15	65	69	51	54	62	68
20	72	79	62	65	74	80

3.2 Arc Tan Shortest Path

The process of arc tan is to regularize the remoteness observed by summing the weights of the phrases resting on the undeviating path. Consequently, the sum of the weights on the shortest corridor is reliable through the supposed reserve and can be utilized as its evaluation. A novel type of algorithm is utilized to discover the shortest path on the novel edge-weighted graph. Figure 2 demonstrates arc tan shortest path range. X axis represents the gene size values, whereas Y axis denotes arc tan shortest path range using both the concepts of semantic similarity in gene

ontology. When gene size increased, the arc tan-based shortest path range increased accordingly. The shortest path measure for semantic similarity (SPMSS) achieves the high performance of 10–15% when compared to the existing system (HMGO). In Table 1 arc tan shortest path range for HMGO and SPMSS is compared.

3.3 Similarity Ratio

The objective of the legalization is to discover whether the anticipated semantic likeness is in stroke through the correspondence based on the appearance data. The connection based on the gene expression data is computed utilizing the Pearson correlation and is submitted to as the appearance similarity. Normally, a high correlation designates a better performance. Figure 2 demonstrates similarity ratio. *X* axis represents the gene size values, whereas *Y* axis denotes similarity ratio using both the concept of semantic similarity in gene ontology. When gene size increased, the similarity increased accordingly. The shortest path measure for semantic similarity (SPMSS) achieves the high performance of 10–15% when compared to the existing system (HMGO). In Table 1 similarity ratio for HMGO and SPMSS is compared.

4 Conclusion

The enhanced shortest path is a novel technique to measure the semantic similarity via exponential and arc tan function. The proposed scheme improves the effectiveness of transfer functions and semantic similarity of gene ontology. In calculation, the weights allocated to the underside are more constant with universal elucidation than the earlier edge-based methods. The performance measures of the proposed work are analyzed and compare with different metrics such as exponential relational rate, arc tan shortest path range, and similarity ratio. In future work, we noticed that different types of approach can also be used to improve the prediction quality for other gene network attributes, such as protein–protein network and the metabolic network.

References

1. Haris, A.: The gene ontology (GO) database & informatics resource. *Nucleic Acids Res.* **32** (2004)
2. Young, M., Wakefield, M., Smyth, G., Oshlack, A.: Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* (2010)

3. Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F., Dowson, C.: Barriers to genetic exchange between bacterial species: streptococcus pneumonia transformation. *J. Bacteriol.* 1016–1023 (2000)
4. Albert, R.: Boolean modeling of genetic regulatory networks. *Lect. Notes Phys.* **650**, 459–481 (2004)
5. Paul, S., Baxter, D., Byrne, J.: Modeling transcriptional control in gene networks-methods, recent results, and future directions. *Bull. Math. Biol.* **62**(2), 247–292 (2000)
6. Xu, T., Du, L., Zhou, Y.: Evaluation of GO-based functional similarity measures using *S. cerevisiae* protein interaction and expression profile data. *BMC Bioinf.* **9** (2008)
7. Barrett, T., Troup, D., Wilhite, S., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tom, M., Edgar, R.: NCBI GEO: mining tens of millions of expression profiles-database and tools update. *Nucleic Acids Res.* **35** (2006)
8. Jiang, J., Conrath, D.: Semantic similarity based on corpus statistics and lexical taxonomy. In: *International Conference on Research in Computational Linguistics (ROCLING)* (1997)
9. Schlicker, A., Domingues, F., Rahnenfuhrer, J., Lengauer, T.: A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinf.* **7** (2006)
10. Vengatesan, K., Selvarajan, S., Pragadeeswaran, S.: The performance analysis of microarray data using occurrence clustering. *Int. J. Math. Sci. Eng.* **3**, 69–75 (2014)
11. Yu, G., Li, F., Qin, Y., Xiaochen, B., Wu, S.W.: GO SemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinf. Appl. Note* **26**, 976–978 (2010)